Utrecht University

Masters Thesis

Topics in Bayesian nonparametrics

Author:
Stefan Franssen, Bsc

Supervisors:
Prof. Dr. A. Van der Vaart
Dr. C. Spitoni
Second Reader:
Dr. K. Dajani

A thesis submitted in fulfillment of the requirements
for the degree of Master of Science

August 14, 2018

Abstract

In this thesis we consider a class of priors, Stick-breaking processes, and compute their posterior distribution. For this posterior we prove consistency results. Furthermore we prove consistency results based on mixtures from stick-breaking processes.

Acknowledgements

I would like to acknowledge Sophie Huiberts for proofreading a part of this thesis. I would like to thank my supervisors for fruitful discussions and supervision of this project.

Contents

Chapter 1

Introduction

1.1 Flow of thesis

In this thesis we start by giving a short introduction to both Bayesian and nonparametric statistics. Here we also introduce the Dirichlet process and species sampling processes. After that, we introduce the stick-breaking processes, which is a class of distributions which is a bit more restrictive than the species sampling processes. The main object of study in this thesis will be the stick-breaking processes. We start by studying simple properties of these processes. After we have proven these simple properties we go on to compute a posterior for the stick-breaking processes. Finally, using this description of the posterior we can prove theorems about when the posterior is consistent. We will both do this in the model where we have observations coming from random measures which admit a stick-breaking representation and where we use the stick-breaking representation in mixtures to build larger models.

1.2 Notation used

- For sequences $(a_n)_{n \in \mathbb{N}}, (b_n)_{n \in \mathbb{N}}$ we say that a_n is asymptotic to b_n, denoted by $a_n \sim b_n$, if $\lim_{n \to \infty} \frac{a_n}{b_n} = 1$.

- If D is a distribution, $X \sim D$ means that X has probability measure D, i.e. $\mathbb{P}(X \in A) = D(A)$. $X_1, X_2, \cdots \overset{\text{iid}}{\sim} D$ means that X_i is an independent sample from D and $X_i \overset{\text{ind}}{\sim} D_i$ means that the distribution of X_i is D_i and all the X_i are independent.

- If $(X, \mathcal{X})$ is a measurable space, then $\mathcal{M}(X)$ denotes the space of all the probability measures on $(X, \mathcal{X})$, unless stated otherwise $\mathcal{M}(X)$ is endowed with the topology of weak convergence.

- For random variables Y, Z the notation $X|Y \sim Z$ means that the conditional distribution of X given Y is Z.

- $\mathrm{Beta}(a, b) = \int_0^1 v^{a-1}(1-v)^{b-1} \, \mathrm{d}\, v$.

- $\mathrm{Be}(a, b)$ is the Beta a, b distribution given by

$$\mathrm{Be}(a, b)(A) = \frac{\int_A v^{a-1}(1-v)^{b-1} \, \mathrm{d}\, v}{\mathrm{Beta}(a, b)}.$$

- If X is a random variable, then $\mathbb{E}[X]$ is the expected value of X. If F is a σ-algebra, then $\mathbb{E}[X|F]$ is the conditional expectation of X given F. If Y is a random variable, $\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)]$ where $\sigma(Y)$ is the σ-algebra generated by the random variable Y. $\mathbb{E}_{X \sim P}[f(X)]$ is the expected value of $f(X)$ given that X is distributed according to P. The obvious extensions to conditional expectations and conditional distributions hold.

- First moment and second moment of a random variable X are $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$ respectively. In general k-th moment of a random variable is a name for $\mathbb{E}[X^k]$.

- If X is a set, $X^{\mathbb{N}}$ is the space of all sequences in X, equivalently, $X^{\mathbb{N}}$ is the space of all functions from $\mathbb{N}$ to X.

- $\coprod$ is the coproduct, i.e. disjoint union in case of sets.

- If $X_1, \cdots, X_n$ is a sample, $\hat{X}_1, \cdots, \hat{X}_m$ are the m distinct observations in the sample $X_1, \cdots, X_n$ in order of first occurrence, and $N_n = (N_{1,n}, \cdots, N_{K_n,n})$ is the vector of K_n elements, where $K_n = m$ and $N_{j,n}$ is the number of times $\hat{X}_j$ appeared in the sample $X_1, \cdots, X_n$.

- If $x = (x_1, \cdots, x_n)$ is a vector, then $x_{-i} = (x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$ the same vector with the i-th coordinate left out.

- We define for every natural number n the set $[n] = \{1, \cdots, n\}$.

- The n simplex is the set of all $n+1$ tuples of nonnegative real numbers adding up to 1: $\mathbb{S}_n = \{(x_1, \cdots, x_{n+1}) : x_i \in \mathbb{R}_{\geq 0}, \sum_{i=1}^{n+1} x_i = 1\}$. We denote $\mathbb{S}_\infty$ the set of sequences of nonnegative real numbers summing to one.

Chapter 2

Prior knowledge

2.1 Introduction to Bayesian theory

We begin by introducing Bayesian theory and how it differs from the *frequentist* approach to statistics. In basic frequentist statistics, we study observations from a process with certain parameters, and we want to estimate these parameters. However, we do not consider these parameters as being random variables themselves. In *Bayesian* statistics, we equip these parameters with a distribution, which would denote our uncertainty of the process generating the random variables, and then, given the new observations, we want to *update* our knowledge of the process. This leads to the *posterior*. The data X given θ follows a distribution P_θ, and θ comes from a prior Π.

A statistical model starts with a parameter space Θ, and for every parameter $\theta \in \Theta$ a distribution P_θ. Often we assume more structure on the parameter space Θ such as a topology or even a metric. These induce a σ-algebra on Θ, the Borel σ-algebra.

Definition 2.1.1. *Let X be a random variable. Let Θ be a measurable space. A **(proper) prior** Π on Θ is a probability measure, such that the distribution of X given θ is P_θ, where the distribution of θ is Π, i.e. $\theta \sim \Pi$, then $X|\theta \sim P_\theta$.*

In *parametric* statistics we work with Θ some subset of $\mathbb{R}^n$. In *nonparametric* statistics we relax this assumption and allow our parameter space to be much bigger. In parametric statistics, one has the parameter space Θ, small, in some sense. Usually, we use $\Theta \subset \mathbb{R}^n$ with the euclidean topology,

and there should be some "nice" relation between $\theta \in \Theta$ and P_θ. This is usually encoded in smoothness conditions in theorems. The limitation of these kinds of statistics is that the models have to be small, and we cannot apply the standard methods to large models. For example, say you know that the observations come from a density, and you want to find the density which maximizes the likelihood of these observations. Putting larger and larger spikes on the points we can make the likelihood arbitrarily large, so there will not be a density which maximizes the likelihood. This means that the main tool from parametric statistics, the MLE, will not work. The major downside of using small models is that if the true distribution is outside the small model, we will not be able to control how well our estimates are. One, therefore, wishes to use models which can capture all distributions, and show they always "work".

Definition 2.1.2. *The **posterior** $\Pi(\cdot|X)$ is the conditional probability distribution of θ given X, in the model where $X|\theta \sim P_\theta$ and Π is the prior on θ.*

Note that the posterior distribution is unique Π almost surely. This is often not a problem, however, for theorems about the behavior of the posterior we cannot always conclude that every posterior for this prior will behave as we want.

To compute the posterior is in many parametric cases easy due to *Bayes formula*. If you have a *dominated* collection of measures P_θ, it is possible to select densities p_θ relative to some σ-finite dominating measure μ such that the map $(x, \theta) \mapsto p_\theta(x)$ is *jointly* measurable. Then a *version* of the posterior distribution $\Pi(\cdot|X)$ is given by

$$\Pi(B|X) = \frac{\int_B p_\theta(X) \, \mathrm{d}\,\Pi(\theta)}{\int p_\theta(X) \, \mathrm{d}\,\Pi(\theta)}$$

Roughly, what we want from a good statistical model is that it will find the true distribution in some sense. Notice that just stating that eventually, with probability one we will find the true distribution will not work, because sample variance would change our estimate. So what we want to do is state that we can get arbitrarily close, in finite time, with probability tending to one as one gets more data. This also has the advantage that while the truth might not be a part of your parameter space, you can still talk about the parameters which are close to your true distribution.

In order to be able to speak about *consistency*, we need some topological notion of closeness. So we introduce the setting which always holds when talking about consistency and contraction rates. For every $n \in \mathbb{N}$, let $X^{(n)}$ be an observation in the sample space $(\mathcal{X}^n, \mathfrak{X}^n)$ with distribution P_θ^n indexed by a parameter θ belonging to a first countable topological space Θ. Given a prior Π on the Borel sets of Θ, we can act like the observations $X^{(n)}$ came from Π and form the posterior $\Pi_n(\cdot|X^{(n)})$.

Definition 2.1.3. *The posterior distribution* $\Pi_n(\cdot|X^{(n)})$ *is said to be **(weakly) consistent** at $\theta_0 \in \Theta$ if, for all open neighborhoods U of θ_0,* $\Pi_n(U^c|X^{(n)}) \to 0$ *in* $\mathbb{P}_{\theta_0}^n$ *probability, as $n \to \infty$. The posterior is said to be **strongly consistent** at $\theta_0 \in \Theta$ if this convergence is in the almost-sure sense.*

2.2 The Dirichlet process and discrete random structures.

In order to define the Dirichlet process, we need to define the Dirichlet distribution first. Let λ^k denote the Lebesgue measure on $\mathbb{R}^n$. We can view the n simplex as a subspace of $\mathbb{R}^n$ by sending $(x_1, \cdots, x_{n+1}) \in \mathbb{S}_n$ to $(x_1, \cdots, x_n)$. Since $\sum_{i=1}^{n+1} x_i = 1$ this parametrizes the n simplex.

Definition 2.2.1. *The **Dirichlet distribution** with parameters $n, \alpha_1, \cdots, \alpha_n$ is a probability distribution on $n-1$ simplex given by*

$$A \mapsto \frac{\Gamma(\sum_{i=1}^n \alpha_i)}{\prod_{i=1}^n \Gamma(\alpha_i)} \int_A \prod_{i=1}^n x_i^{\alpha_i} \, d\lambda^{n-1}(x_{-n}).$$

We denote this distribution by $Dir(n, \alpha_1, \cdots, \alpha_n)$.

Because any random variable with a Dirichlet distribution lives on the $K-1$ simplex, the coordinates are nonzero and add up to one, so it can function as a probability measure. This will be used to construct random measures as in the next definition. In Bayesian nonparametrics there is a process which is a main method of defining priors, namely the *Dirichlet process*. We cite the definition from [1, ch. 4]

Definition 2.2.2. *A random measure P on $(\mathcal{X}, \mathfrak{X})$ is said to possess a **Dirichlet process** distribution $DP(M\alpha)$ with base measure $M\alpha$, for $M > 0$*

*and α a probability measure, if for every finite measurable partition $A_1, \cdots, A_k$
of $\mathcal{X}$,*

$$(P(A_1), \cdots, P(A_k)) \sim Dir(k; M\alpha(A_1), \cdots, M\alpha(A_k)).$$

A priori it is not clear that the Dirichlet process exists. However, we will see a theorem from Sethuraman which gives a representation showing that such a process exists, namely, a process of this form has a *stick-breaking representation* which can be used to generalize the Dirichlet process. In my thesis, this is what we will be doing, and we try to derive the same kind of consistency results for the more general class of distributions we will be considering.

The following theorem by Sethuraman [1, theorem 4.12] gives an explicit representation of the Dirichlet process.

Theorem 2.2.3. *If $\theta_1, \cdots \overset{iid}{\sim} \alpha$ and $V_1, \cdots \overset{iid}{\sim} Be(1, M)$ are independent random variables and $W_j = V_j \prod_{l=1}^{j-1}(1 - V_l)$, then $\sum_{j=1}^{\infty} W_j \delta_{\theta_j} \sim DP(M\alpha)$.*

From the stick-breaking representation theorem, Theorem 2.2.3, by Sethuraman, we see that if $P \sim \mathrm{DP}(M\alpha)$ then P is almost surely discrete. We can study discrete random structures in a more general setting. This is first done by studying exchangeable partitions, and then using the techniques and concepts from the theory of discrete random structures to build a general class of distributions which can used as priors.

If we sample from a random discrete measure, and then look at the tied observations of this sample, we naturally get random exchangeable partitions of finite sets. Conversely, if we specify a distribution on partitions in a clever way, we can use this to create samples.

Recall that a partition $\{A_1, \cdots, A_k\}$ of the finite set $[n]$ is a decomposition into disjoint subsets of $[n]$ whose union is $[n]$. The cardinalities $n_i = |A_i|$ of the sets in a partition of $[n]$ are said to form a partition of n: an unordered set $\{n_1, \cdots, n_k\}$ of natural numbers such that $n = \sum_{i=1}^{k} n_i$. Note that here we do remember how many times a specific value occured in this set, so that $\{2, 2\}$ is not the same as $\{2\}$. The sets in a partition are considered unordered, but if we list the sets in a specific order, then the cardinalities match that order. An ordered partition $(n_1, \cdots, n_k)$ of n is called a composition of n, and the set of all compositions of n is denoted by $\mathcal{C}_n$. The particular order by the sizes of the smallest element in every A_i is called the order of appearance. A random partition of $[n]$ is a random element defined on some probability

space taking values in the set of all partitions of $[n]$. Its induced distribution is a probability measure on the set of all partitions of $[n]$.

Definition 2.2.4 (exchangeable partition). *A random partition $\mathcal{P}_n$ of $[n]$ is called exchangeable if its distribution is invariant under the action of any permutation of $\sigma : [n] \to [n]$. Equivalently, a random partition $\mathcal{P}_n$ is called exchangeable if there exists a symmetric function $p : \mathcal{C}_n \to [0, 1]$ such that, for every partition $\{A_1, \cdots, A_k\}$ of $[n]$,*

$$\mathbb{P}(\mathcal{P}_n = \{A_1, \cdots, A_k\}) = P(|A_1|, \cdots, |A_k|).$$

The function p is called the exchangeable partition probability function (EPPF) of $\mathcal{P}_n$.

We want to extend this definition to partition structures, which are a way to link partitions across n. The goal is to capture the behavior of looking at the partition defined from the first n observations. So if we have a partition coming from $X_1, \cdots, X_n$ and we look at what partitions we can get from $X_1, \cdots, X_{n+1}$, we are very restricted in what can happen, namely, only one element of the partition can change, and this one will receive the extra element $n + 1$. This can be repeated if we want to include or leave out more element.

Definition 2.2.5 ((infinite) exchangeable partition). *An **infinite exchangeable random partition** is a sequence $(\mathcal{P}_n)_{n \in \mathbb{N}}$ of exchangeable random partitions of $[n]$ that are consistent in the sense that $\mathcal{P}_{n-1}$ is equal to the partition obtained from $\mathcal{P}_n$ by leaving out the element N, almost surely, for every N. The function $p : \cup_{N=1}^{\infty} \mathcal{C}_n \to [0, 1]$ whose restriction $\mathcal{C}_n$ is equal to the exchangeable partition probability function of $\mathcal{P}_n$ is called the **exchangeable partition probability function (EPPF)** of $(\mathcal{P}_n)_{n \in \mathbb{N}}$.*

If $n = (n_1, \cdots, n_k)$ is a vector, we denote n^{j+} to be the vector $n^{j+} = (n_1, \cdots, n_{j-1}, n_j + 1, n_{j+1}, \cdots, n_k)$ for $j \leq k$ and $(n_1, \cdots, n_k, 1)$ if $j = k + 1$. If $n \in \mathcal{C}_n$, then n^{j+} is an element in $\mathcal{C}_{n+1}$. By being slightly loose with specification, we can give a working definition of the predictive probability function.

Definition 2.2.6. *The predictive probability function is the function $p = (p_1, p_2, \cdots)$ with $p_j : \cup_n \mathcal{C}_n \to [0, 1]$ with*

$$p_j(n) = \frac{p(n^{j+})}{p(n)}, \qquad j = 1, \cdots, k + 1,$$

for every vector n with k distinct elements.

We note that the processes we will study will (often) be a special case of species sampling models. For discussion and contrast, we will refer to these processes as well.

We first start by defining a special kind of topological space, which has a lot of nice properties. This space makes a lot of definitions behave better, however we will not refer to these kinds of details in this thesis. It is however needed to state the definition of a species sampling model.

Definition 2.2.7 (Polish space)**.** *A topological space X is called a* **Polish** *space if it is a complete separable metric space relative to some metric that generates the topology.*

With the definition of Polish spaces, it becomes possible to state what a species sampling process is.

Definition 2.2.8 (Species sampling model)**.** *A* **Species sampling model** *is a pair consisting of a sequence (X_i) of random variables and a random measure P such that $X_1, X_2, \cdots | P \overset{iid}{\sim} P$. and P takes the form*

$$P = \sum_{j=1}^{\infty} W_j \delta_{\theta_j} + (1 - \sum_{j=1}^{\infty} W_j)G,$$

for $\theta_1, \theta_2, \cdots \overset{iid}{\sim} G$ with G an atomless probability distribution on a Polish space $\mathcal{X}$, and an independent random subprobability vector (W_j). The random distribution P in a SSM is called a **Species sampling process***. If $\sum_{j=1}^{\infty} W_j = 1$, the species sampling process is called proper.*

Lemma 2.2.9 (Lemma 3.4 of [1])**.** *Let $V_i \overset{ind}{\sim} \mathcal{D}_j$ be a sequence of random variables. Suppose that $W_j = V_j \prod_{i=1}^{j-1}(1-V_i)$. Then $W = (W_1, W_2, \cdots) \in \mathbb{S}_{\infty}$ if and only if $\mathbb{E}\left[\prod_{l=1}^{j}(1 - V)\right] \to 0$ as $j \to \infty$. For independent random variables $V_1, V_2, \cdots$ this condition is equivalent to $\sum_{l=1}^{\infty} \log \mathbb{E}[1 - V_l] = -\infty$. In particular, for iid variables $V_1, V_2, \cdots$ it suffices that $\mathbb{P}(V_1 > 0) > 0$. If for every $k \in \mathbb{N}$ the support of $(V_1, \cdots, V_k)$ is $[0, 1]^k$ then the support of W is the whole space $\mathbb{S}_{\infty}$.*

We can justify the name of the predictive probability functions with the following lemma [1, Lemma 14.11].

Lemma 2.2.10. *The predictive distributions of variables $X_1, X_2, \cdots$ in the species sampling model take the form $X_1 \sim G$ and, for $n \geq 1$,*

$$X_{n+1}|X_1, \cdots, X_n \sim \sum_{j=1}^{K_n} p_j(N_n)\delta_{\tilde{X}_j} + p_{K_n+1}G,$$

where the functions $p_j : \cup_n \mathcal{C}_n \to [0,1]$ are the predictive probability functions of the infinite exchangeable random partition generated by $X_1, X_2, \cdots$. Here K_n denotes the number of distinct observations of $X_1, \cdots, X_n$, $\tilde{X}_j$ is the j-th distinct observation and N_n is the vector such that $N_{j,n}$ is the number of times the j-th distinct observation occurs in $X_1, \cdots, X_n$.

A concept that will appear often in the theory of discrete random structures are the size biased permutations. The intuition behind this is that the locations θ_i all are independent and have the same distribution. This means that if we pick any permutation of the weights, denoted by W^σ, the distribution of

$$F' = \sum_{j=1}^{\infty} W_j^\sigma \delta_{\theta_j} + (1 - \sum_{j=1}^{\infty} W_j^\sigma)G$$

is equal to the distribution of

$$F = \sum_{j=1}^{\infty} W_j \delta_{\theta_j} + (1 - \sum_{j=1}^{\infty} W_j)G.$$

So we cannot distinguish the order of W. It turns out that the size-biased permutations are often the right thing to look at.

An element of the infinite simplex $W \in \mathbb{S}_\infty$ naturally defines a probability distribution on the natural numbers, namely $\mathbb{P}_w(I = i) = w_i$. In the concept of proper species sampling processes, W is an element of the infinite simplex, so every such random weight vector W defines such a distribution. Now if we look conditional on W, the probabiity that $I = i$ is exactly the probability that we draw θ_i. Now because we do not know W, we cannot distinguish between W and any permutation acting on W, so we look at the object which has an invariant distribution under taking (random) permutations. It turns out the right concept is size-biased permutations.

Definition 2.2.11. *The **size-biased permutation** of a probability distribution $W = (w_j)$ on $\mathbb{N}$ is the random vector $(\tilde{w}_1, \tilde{w}_2, \cdots)$, for $\tilde{w}_j = w_{\tilde{I}_j}$, and $\tilde{I}_1, \tilde{I}_2, \cdots$ the distinct values in an i.i.d. sequence with $\mathbb{P}(I = i|W) = w_i$.*

Observe that taking the size biased permutation of a random vector in size-biased order gives a vector with the same distribution as before.

From Lemma 2.2.10 we can deduce that you can actually specify a species sampling process by actually specifying one of the following three pairs of objects:

- G and the distribution of (W_j),

- G and the exchangeable partition probability function,

- G and the prediction probability function.

To express the exchangeable probability function in terms of the distribution of W we have the following formula

Lemma 2.2.12 (theorem 14.14 [1])**.**

$$p(n_1, \cdots, n_k) = \mathbb{E}\left[\prod_{j=1}^{k} \tilde{W}_j^{n_j-1} \prod_{j=2}^{k}(1 - \sum_{i\leq j} \tilde{W}_i)\right],$$

where $\tilde{W}$ again is the size-biased permutation of W.

Lemma 2.2.13 (Kingman's formula for EPPF, ex 14.1 [1])**.** *The EPPF of a proper species sampling model can be written in the form*

$$p(n_1, \cdots, n_k) = \sum_{1\leq i_1 \neq \cdots \neq i_k < \infty} \mathbb{E}\left[\prod_{j=1}^{k} W_{i_j}^{n_j}\right].$$

There is a kind of generalization of the Dirichlet process which is called the *Pitman-Yor process* [3]. This exploits results [2] by Pitman which allows you to give an explicit computation for the posterior. For details see [1, chapter 14].

The Pitman-Yor process is a special kind of species sampling process. If $0 \leq \sigma < 1$ and base measure G then this admits a stick-breaking representation. It will be an example of a stick-breaking process which we will define later. See [1, theorem 14.33] and [1, theorem 14.25]

Theorem 2.2.14. *Let $V_j \stackrel{ind}{\sim} Be(1-\sigma, M+j\sigma)$ and set $\tilde{W}_j = V_j \prod_{l=1}^{j-1}(1-V_l)$. Let $\theta_1, \cdots \stackrel{iid}{\sim} G$. Suppose that $P = \sum_{i=1}^{\infty} W_i \delta_{\theta_i}$. Then the distribution of P is given by the Pitman-Yor process, $P \sim PY(\sigma, M, G)$.*

We will contrast the new results in this thesis by the corresponding results on Species sampling models, the Dirichlet process and the Pitman-Yor. In the definition of stick-breaking processes we will not include the assumption that the base measure G is atomless. However, for many theorems, we do need this assumption.

Finally some object of interest is the support of a distribution.

Lemma 2.2.15 (Support). *The **support** of a probability measure on the Borel sets of a Polish space is the smallest closed set of probability one. Equivalently, it is the set of all elements of the space for which every open neighborhood has positive probability.*

We will sometimes conflate this definition with a more vague concept of support, namely for discrete distributions the points of positive probability.

Chapter 3

The stick-breaking process

In this chapter, we introduce the stick-breaking process and derive some first properties. The Dirichlet process is the process we want to generalize. In order to do so, we first look at a specific representation which some processes have, namely of a *stick-breaking* process. The Dirichlet process has a stick-breaking representation [1, Theorem 4.12]. We will first introduce stick-breaking as a way to give a probability distribution on the infinite *simplex*. Then we can use these weights to create the main object of study.

3.1   Introducing the stick-breaking process

We begin with a short discussion on how to define the process we want to study. We want to find a process that generates probability measures, so the goal is to figure out what requirements actually are needed to make this happen.

Definition 3.1.1. *Let $(X, \mathcal{B})$ be a measurable space. Let α be a probability distribution on $(X, \mathcal{B})$. Let $\mathcal{D}$ be a distribution on $[0, 1]$. Given **stick-breaking weights** $V_i \overset{ind}{\sim} \mathcal{D}$, we define **weights***

$$W_i = V_i \prod_{j=1}^{i-1}(1 - V_j)$$

If we let $\theta_i \overset{ind}{\sim} \alpha$, we can combine this into

$$F = \sum_{i=1}^{\infty} \delta_{\theta_i} W_i.$$

13

We will now take a look at a couple of requirements we need to get the properties we want. Then we can see how to create the precise definition.

Lemma 3.1.2. *Suppose that F follows a distribution which admits the previous representation, then every instance of F is a measure.*

Proof. $F(\emptyset) = \sum_{i=1}^{\infty} \delta_{\theta_i}(\emptyset) W_i$. Because θ_i are random variables on X, we know that $\delta_{\theta_i}(\emptyset) = 0$ for all θ_i. Hence this sum is zero. For non-negativity, we remark that the weights $W_i \geq 0$ and $\delta_{\theta_i}(A) \geq 0$, so $F(A)$ is a sum of non-negative real numbers, hence non-negative.

For additivity, we observe that, if we have disjoint A_j, then θ_i is in at most one such A_j, so we get that

$$F(\cup_j A_j) = \sum_{i=1}^{\infty} \delta_{\theta_i}(\cup_j A_j) W_i = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \delta_{\theta_i}(A_j) W_i.$$

Now we can use absolute summability of this sequence and monotone convergence theorem to swap the order of summation, which gives

$$F(\cup_j A_j) = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \delta_{\theta_i}(A_j) W_i = \sum_{j=1}^{\infty} F(A_j).$$

$\square$

In view of Lemma 2.2.9, if $\mathbb{P}(V_1 > 0) > 0$, this almost surely defines a probability measure, since the measure of the total space is $\sum_{i=1}^{\infty} W_i$, which is almost surely 1. So from now on, we assume this. This leads to the central definition we work with:

So now we know what extra requirements we need to make such random measures into probability measures. This gives rise to the main two definitions.

Definition 3.1.3. *We say that F has a **stick-breaking representation** with **base measure** α and **stick-breaking weight measure** $\mathcal{D}$ iff, for $\theta_1, \theta_2, \cdots \overset{iid}{\sim} \alpha, V_1, \cdots \overset{iid}{\sim} \mathcal{D}, W_i = V_i \prod_{j=1}^{i-1}(1 - V_j)$, such that $\mathbb{P}(V_1 > 0) > 0$, we have that F can be written as follows:*

$$F = \sum_{i=1}^{\infty} W_i \delta_{\theta_i}.$$

We write $F \sim SBP(\alpha, \mathcal{D})$.

Definition 3.1.4. *We say that F has a **stick-breaking representation** with **different relative weight distributions** with **base measure** α and **stick-breaking weight measures** $\mathcal{D}_j$ iff, for $\theta_1, \theta_2, \cdots \overset{iid}{\sim} \alpha$, $V_j \overset{ind}{\sim} \mathcal{D}_j$, $W_i = V_i \prod_{j=1}^{i-1}(1 - V_i)$, such that $\sum_{l=1}^{\infty} \log \mathbb{E}\left[1 - V_l\right] = -\infty$, we have that F can be written as follows:*

$$F = \sum_{i=1}^{\infty} W_i \delta_{\theta_i}.$$

We write $F \sim SBP(\alpha, (\mathcal{D}_j))$.

The condition on the sum of logarithms of $\mathbb{E}\left[1 - V_l\right]$ is to ensure that $\sum_{i=1}^{\infty} W_i = 1$ almost surely.

Note that $\delta_{\theta_i}(A)$ is one with probability $\alpha(A)$, so for a fixed set A, we can see the distribution of $F(A)$ as a random weighted sum of independent Bernoulli $\alpha(A)$ variables, where the weights are given by the random vector $W = (W_1, W_2, \cdots)$.

Also, note that P is almost surely a discrete random measure.

Suppose that for every n the sum of the first n weights $\sum_{i=1}^{n} W_i$ is almost surely smaller than 1. Then we can transform back to the stick-breaking points by

$$V_i = \frac{W_i}{1 - \sum_{j=1}^{i-1} W_j}$$

Lemma 3.1.5. $\sum_{i=1}^{n} W_i = 1 - \prod_{i=1}^{n}(1 - V_i)$

Proof. For $N = 1$ it is clear, and with induction $W_{n+1} + 1 - \prod_{i=1}^{n}(1 - V_i) = 1 - (1 - V_{n+1}) \prod_{i=1}^{n}(1 - V_i)$. $\square$

Further, one readily sees that if you have a stick-breaking process with an atomless base measure, then this defines a proper Species sampling model. Obviously, not all Species sampling models are stick-breaking processes, however, there are also examples of Species sampling processes which have been studied in practice which are not stick-breaking process priors. One such family of examples is the family of Gibbs processes. Some of the Gibbs processes are stick-breaking processes such as the Pitman-Yor process, but in general, they have dependent distribution on the relative stick-breaking weights. For an example, consider the Normalized Inverse-Gaussian processes. They admit a stick-breaking presentation with dependent distributions.

We start with a small but useful lemma on the sum of second moments of stick-breaking process (with identical weight distributions).

Lemma 3.1.6. *Suppose $V_1, \cdots \overset{iid}{\sim} \mathcal{D}$. Then The stick-breaking weights W_i have $\sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right] = \frac{\mathbb{E}\left[V_1^2\right]}{1 - \mathbb{E}\left[(1-V_1)^2\right]}$*

Proof. We start by expanding the definition of the W_i in terms of V_i:

$$\sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right] = \sum_{i=1}^{\infty} \mathbb{E}\left[V_i^2 \prod_{j=1}^{i-1}(1-V_j)^2\right].$$

Then using the independence of the V_i we get

$$\sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right] = \sum_{i=1}^{\infty} \mathbb{E}\left[V_i^2\right] \prod_{j=1}^{i-1} \mathbb{E}\left[(1-V_j)^2\right].$$

Now we use that all the $V_i \sim \mathcal{D}$, so then we get the following equality:

$$\sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right] = \sum_{i=1}^{\infty} \mathbb{E}\left[V_1^2\right] \prod_{j=1}^{i-1} \mathbb{E}\left[(1-V_1)^2\right].$$

Now we can move the expectation of $\mathbb{E}\left[V_1^2\right]$ outside to get that the previous term also equals

$$\mathbb{E}\left[V_1^2\right] \sum_{i=1}^{\infty} \prod_{j=1}^{i-1} \mathbb{E}\left[(1-V_1)^2\right].$$

Now we can count how many times we a term $\mathbb{E}\left[(1-V_1)^2\right]$ in the product, this yields that

$$\sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right] = \mathbb{E}\left[V_1^2\right] \sum_{i=1}^{\infty} \mathbb{E}\left[(1-V_1)^2\right]^{i-1}.$$

Using that $\mathbb{E}\left[(1-V_1)^2\right] \leq 1$ and the known limit of a geometric series we finally get:

$$\sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right] = \frac{\mathbb{E}\left[V_i^2\right]}{1 - \mathbb{E}\left[(1-V_1)^2\right]}.$$

$\square$

Chapter 4

Simple properties

In this chapter, we look at some simple properties of the stick-breaking distributions. In particular, we look at the mean and variance of integrals with respect to these measures. The results in this chapter are used to compute the posterior distribution in the next chapter.

4.1 Mean and variance of the stick-breaking process

We can compute the mean and (co)variance of the random measure as follows

Proposition 4.1.1. *Let $F \sim \Pi$, with Π being a stick-breaking process with base measure α and stick-breaking weights distributed according to D_i. Let A, B be measurable subsets of X. Then*

$$\mathbb{E}\left[F(A)\right] = \alpha(A),$$

$$Var(F(A)) = \alpha(A)\alpha(A^c) \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right],$$

$$Cov(F(A), F(B)) = (\alpha(A \cap B) - \alpha(A)\alpha(B)) \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right].$$

This proposition has a natural specialization to the case where all the stick-breaking weights have the same distribution. This yields

Corollary 4.1.2. *Let $F \sim \Pi$, with Π being a stick-breaking process with base measure α and stick-breaking weights distributed according to D. Let A, B be measurable subsets of X. Then*

$$\mathbb{E}\left[F(A)\right] = \alpha(A),$$

$$Var(F(A)) = \frac{\alpha(A)\alpha(A^c)\mathbb{E}\left[V_1^2\right]}{1 - \mathbb{E}\left[(1 - V_1)^2\right]},$$

$$Cov(F(A), F(B)) = \frac{(\alpha(A \cap B) - \alpha(A)\alpha(B))\,\mathbb{E}\left[V_1^2\right]}{1 - \mathbb{E}\left[(1 - V_1)^2\right]}.$$

Proof. We apply the previous proposition together with Lemma 3.1.6 which states that

$$\sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right] = \frac{\mathbb{E}\left[V_1^2\right]}{1 - \mathbb{E}\left[(1 - V_1)^2\right]}.$$

$\square$

The proof of the Proposition 4.1.1 is quite a long computation and is broken down into several parts for clarity.

Proof. **Computing the mean** Let A be a measurable set. Then

$$\mathbb{E}F(A) = \mathbb{E}\sum_{i=1}^{\infty} W_i \delta_{\theta_i}(A)$$

$$\stackrel{\text{MCT}}{=} \sum_{i=1}^{\infty} \mathbb{E}W_i \delta_{\theta_i}(A)$$

$$\stackrel{\text{iid}}{=} \sum_{i=1}^{\infty} \mathbb{E}W_i \mathbb{E}\delta_{\theta_i}(A)$$

$$= \sum_{i=1}^{\infty} \mathbb{E}W_i \alpha(A)$$

$$= \alpha(A) \sum_{i=1}^{\infty} \mathbb{E}W_i$$

$$\stackrel{\text{MCT}}{=} \alpha(A)\mathbb{E}\sum_{i=1}^{\infty} W_i$$

$$= \alpha(A).$$

Here the last step holds since $\sum_{i=1}^{\infty} W_i$ is almost surely 1.

Computing the variance Let A be a measurable set. Then

$$
\begin{aligned}
\mathrm{Var}\left(F(A)\right) &= \mathbb{E}\left[F(A)^2\right] - \mathbb{E}\left[F(A)\right]^2 \\
&= \mathbb{E}\left[F(A)^2\right] - \alpha(A)^2 \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{\infty} W_i \delta_{\theta_i}(A)\right)^2\right] - \alpha(A)^2 \\
&= \mathbb{E}\left[\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} W_i W_j \delta_{\theta_i}(A)\delta_{\theta_j}(A)\right] - \alpha(A)^2 \\
&= \mathbb{E}\left[\sum_{i=1}^{\infty}\sum_{j\neq i} W_i W_j \delta_{\theta_i}(A)\delta_{\theta_j}(A)\right] + \mathbb{E}\left[\sum_{i=1}^{\infty} W_i^2 \delta_{\theta_i}(A)^2\right] - \alpha(A)^2.
\end{aligned}
$$

We look at each of the components of this expression.

Rewriting $\mathbb{E}\left[\sum_{i=1}^{\infty}\sum_{j\neq i} W_i W_j \delta_{\theta_i}(A)\delta_{\theta_j}(A)\right]$ We start with

$$
\mathbb{E}\left[\sum_{i=1}^{\infty}\sum_{j\neq i} W_i W_j \delta_{\theta_i}(A)\delta_{\theta_j}(A)\right].
$$

We apply the monotone convergence theorem to move the expectation inside the sum, so we get

$$
\mathbb{E}\left[\sum_{i=1}^{\infty}\sum_{j\neq i} W_i W_j \delta_{\theta_i}(A)\delta_{\theta_j}(A)\right] = \sum_{i=1}^{\infty}\sum_{j\neq i} \mathbb{E}\left[W_i W_j \delta_{\theta_i}(A)\delta_{\theta_j}(A)\right].
$$

We use that θ_i is independent of all the θ_j with $j \neq i$ and W_j for all j, so we can rewrite this expression into

$$
\sum_{i=1}^{\infty}\sum_{j\neq i} \mathbb{E}\left[W_i W_j \delta_{\theta_i}(A)\delta_{\theta_j}(A)\right] = \sum_{i=1}^{\infty}\sum_{j\neq i} \mathbb{E}\left[W_i W_j\right]\mathbb{E}\left[\delta_{\theta_i}(A)\right]\mathbb{E}\left[\delta_{\theta_j}(A)\right].
$$

We know the expectation of $\delta_{\theta_i}(A)$, namely $\alpha(A)$, so we get

$$
\sum_{i=1}^{\infty}\sum_{j\neq i} \mathbb{E}\left[W_i W_j\right]\mathbb{E}\left[\delta_{\theta_i}(A)\right]\mathbb{E}\left[\delta_{\theta_j}(A)\right] = \alpha(A)^2 \sum_{i=1}^{\infty}\sum_{j\neq i} \mathbb{E}\left[W_i W_j\right].
$$

We use the monotone convergence theorem to move the innermost sum inside the expectation, this yields

$$\alpha(A)^2 \sum_{i=1}^{\infty} \sum_{j \neq i} \mathbb{E}\left[W_i W_j\right] = \alpha(A)^2 \sum_{i=1}^{\infty} \mathbb{E}\left[\sum_{j \neq i} W_i W_j\right].$$

We can also move the W_i outside the inner summation, to get

$$\alpha(A)^2 \sum_{i=1}^{\infty} \mathbb{E}\left[\sum_{j \neq i} W_i W_j\right] = \alpha(A)^2 \sum_{i=1}^{\infty} \mathbb{E}\left[W_i \sum_{j \neq i} W_j\right].$$

Then we know that $\sum_{j=1}^{\infty} W_j$ is almost surely one, so $\sum_{j \neq i} W_j = 1 - W_i$ almost surely. This gives

$$\alpha(A)^2 \sum_{i=1}^{\infty} \mathbb{E}\left[W_i \sum_{j \neq i} W_j\right] = \alpha(A)^2 \sum_{i=1}^{\infty} \mathbb{E}\left[W_i(1 - W_i)\right].$$

We can simplify this into

$$\alpha(A)^2 \sum_{i=1}^{\infty} \mathbb{E}\left[W_i(1 - W_i)\right] = \alpha(A)^2 \sum_{i=1}^{\infty} \mathbb{E}\left[W_i\right] - \alpha(A)^2 \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right].$$

Now we can use monotone convergence theorem again to take the first sum inside the expectation, and use that $\sum_{i=1}^{\infty} W_i$ is equal to 1 almost surely. This yields

$$\alpha(A)^2 \sum_{i=1}^{\infty} \mathbb{E}\left[W_i\right] - \alpha(A)^2 \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right] = \alpha(A)^2 \left(1 - \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right]\right).$$

Rewriting $\mathbb{E}\left[\sum_{i=1}^{\infty} W_i^2 \delta_{\theta_i}(A)^2\right]$ We start with

$$\mathbb{E}\left[\sum_{i=1}^{\infty} W_i^2 \delta_{\theta_i}(A)^2\right].$$

We observe that $\delta_{\theta_i}(A)$ is 1 if $\theta_i \in A$ and 0 otherwise, so $\delta_{\theta_i}^2 = \delta_{\theta_i}$.

$$\mathbb{E}\left[\sum_{i=1}^{\infty} W_i^2 \delta_{\theta_i}(A)^2\right] = \mathbb{E}\left[\sum_{i=1}^{\infty} W_i^2 \delta_{\theta_i}(A)\right].$$

We apply monotone convergence theorem to move the sum outside the expectation

$$\mathbb{E}\left[\sum_{i=1}^{\infty} W_i^2 \delta_{\theta_i}(A)\right] = \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2 \delta_{\theta_i}(A)\right].$$

We use the independence of W_i and θ_i to get

$$\sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2 \delta_{\theta_i}(A)\right] = \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2 \mathbb{E}\delta_{\theta_i}(A)\right].$$

Then because $\mathbb{E}\left[\delta_{\theta_i}(A)\right] = \alpha(A)$ we get

$$\sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right] \mathbb{E}\left[\delta_{\theta_i}(A)\right] = \alpha(A) \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right].$$

Combining the results We had

$$\mathbb{E}\left[F(A)^2\right] = \mathbb{E}\left[\sum_{i=1}^{\infty} \sum_{j \neq i} W_i W_j \delta_{\theta_i}(A) \delta_{\theta_j}(A)\right] + \mathbb{E}\left[\sum_{i=1}^{\infty} W_i^2 \delta_{\theta_i}(A)^2\right] - \alpha(A)^2.$$

We fill in

$$\mathbb{E}\left[\sum_{i=1}^{\infty} \sum_{j \neq i} W_i W_j \delta_{\theta_i}(A) \delta_{\theta_j}(A)\right] = \alpha(A)^2 \left(1 - \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right]\right).$$

And

$$\mathbb{E}\left[\sum_{i=1}^{\infty} W_i^2 \delta_{\theta_i}(A)^2\right] = \alpha(A) \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right].$$

This gives

$$\mathbb{E}\left[F(A)^2\right] = \alpha(A)^2 \left(1 - \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right]\right) + \alpha(A) \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right] - \alpha(A)^2.$$

Then we can take out a factor $\alpha(A)$ to get

$$\mathbb{E}\left[F(A)^2\right] = \alpha(A) \left(-\alpha(A) \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right] + \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right]\right).$$

This simplifies into

$$\mathbb{E}F(A)^2 = \alpha(A)\left((1 - \alpha(A))\sum_{i=1}^{\infty}\mathbb{E}\left[W_i^2\right]\right).$$

Now we use that $1 - \alpha(A) = \alpha(A^c)$, which gives

$$\mathbb{E}F(A)^2 = \alpha(A)\alpha(A^c)\sum_{i=1}^{\infty}\mathbb{E}\left[W_i^2\right].$$

Covariance

Let A, B be two measurable sets. Then we want to know the covariance of $F(A)$ and $F(B)$. We expand the definition for covariance, which leads to

$$\text{Cov}(F(A), F(B)) = \mathbb{E}\left[(F(A) - \mathbb{E}\left[F(A)\right])(F(B) - \mathbb{E}\left[F(B)\right])\right].$$

We fill in the known value for $\mathbb{E}\left[F(A)\right]$ and $\mathbb{E}\left[F(B)\right]$, and this leads that the previous term also equals

$$\mathbb{E}\left[(F(A) - \alpha(A))(F(B) - \alpha(B))\right].$$

We can now simplify the expression by taking known values outside and repeating the known result on expectations. This leads to

$$\text{Cov}(F(A), F(B)) = \mathbb{E}\left[F(A)F(B)\right] - \alpha(A)\alpha(B)$$

We expand the definition of $F(A)$ and $F(B)$, therefore

$$\text{Cov}(F(A), F(B)) = \mathbb{E}\left[\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\delta_{\theta_i}(A)\delta_{\theta_j}(B)W_iW_j\right] - \alpha(A)\alpha(B).$$

Using the monotone convergence theorem we can move the expectation inside the infinite sums, thus

$$\text{Cov}(F(A), F(B)) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\mathbb{E}\left[\delta_{\theta_i}(A)\delta_{\theta_j}(B)W_iW_j\right] - \alpha(A)\alpha(B).$$

We split the sums into summing over $j \neq i$ and $j = i$, because this means we sum terms W_iW_j where $i \neq j$ and we sum W_i^2. This leads to the next

equality

$$\text{Cov}(F(A), F(B)) = \sum_{i=1}^{\infty} \sum_{j \neq i} \mathbb{E}\left[\delta_{\theta_i}(A)\delta_{\theta_j}(B)W_i W_j\right] +$$

$$\sum_{i=1}^{\infty} \mathbb{E}\left[\delta_{\theta_i}(A)\delta_{\theta_i}(B)W_i^2\right] - \alpha(A)\alpha(B).$$

We use that θ_i and θ_j and $W_i W_j$ are independent and θ_i and W_i are independent. We use that, for independent random variables, the expectation of the product is the product of the expectations. Then

$$\text{Cov}(F(A), F(B)) = \sum_{i=1}^{\infty} \sum_{j \neq i} \mathbb{E}\left[\delta_{\theta_i}(A)\right] \mathbb{E}\left[\delta_{\theta_j}(B)\right] \mathbb{E}\left[W_i W_j\right] +$$

$$\sum_{i=1}^{\infty} \mathbb{E}\left[\delta_{\theta_i}(A)\delta_{\theta_i}(B)\right] \mathbb{E}\left[W_i^2\right] - \alpha(A)\alpha(B).$$

We know that $\mathbb{E}\left[\delta_{\theta_i}(A)\right] = \alpha(A)$, similarly for $\delta_{\theta_j}(B)$. Furthermore, $\delta_{\theta_i}(A)\delta_{\theta_i}(B) = \delta_{\theta_i}(A \cap B)$. The expectation of this is $\alpha(A \cap B)$. So this leads to

$$\text{Cov}(F(A), F(B)) = \alpha(A)\alpha(B) \sum_{i=1}^{\infty} \sum_{j \neq i} \mathbb{E}\left[W_i W_j\right] +$$

$$\alpha(A \cap B) \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right] - \alpha(A)\alpha(B).$$

We can move the sum over all the j inside the expectation using the monotone convergence theorem, which leads to

$$\text{Cov}(F(A), F(B)) = \alpha(A)\alpha(B) \sum_{i=1}^{\infty} \mathbb{E}\left[W_i \sum_{j \neq i} W_j\right] +$$

$$\alpha(A \cap B) \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right] - \alpha(A)\alpha(B).$$

We use that $\sum_{j=1}^{\infty} W_j = 1$ almost surely, so that in expectation the sum $\sum_{j \neq i} W_j$ equals $1 - W_i$. Using this we get that the previous result also

equals

$$\alpha(A)\alpha(B) \sum_{i=1}^{\infty} \mathbb{E}\left[W_i(1 - W_i)\right] + \alpha(A \cap B) \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right] - \alpha(A)\alpha(B).$$

We can simplify, move the sum inside the expectation and then use that $\sum_{i=1}^{\infty} W_i = 1$ almost surely to get

$$\text{Cov}(F(A), F(B)) = \alpha(A)\alpha(B)(1 - \sum_{i=1}^{\infty} \mathbb{E}W_i^2) +$$

$$\alpha(A \cap B) \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right] - \alpha(A)\alpha(B).$$

We can cancel the $\alpha(A)\alpha(B)$ terms so this also equals

$$-\alpha(A)\alpha(B) \sum_{i=1}^{\infty} \mathbb{E}W_i^2 + \alpha(A \cap B) \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right].$$

We can simplify for one last time to get

$$(\alpha(A \cap B) - \alpha(A)\alpha(B)) \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right].$$

$\square$

Retrieving the distribution of the Dirichlet process For the Dirichlet process $V_i \sim \beta(1, M)$, and $1 - V_i$ is then $\beta(M, 1)$ distributed. Then $\mathbb{E}[V_i] = \frac{1}{M+1}$, $\mathbb{E}[1 - V_i] = \frac{M}{M+1}$. The Variances are given by

$$\text{Var}(V_i) = \frac{M}{(1 + M)^2(M + 2)} = \text{Var}(1 - V_i).$$

So now we can compute the second moments by using $\mathbb{E}[X^2] = \text{Var}(X) + \mathbb{E}[X]^2$. So

$$\mathbb{E}\left[V_1^2\right] = \frac{M}{(M + 1)^2(M + 2)} + \frac{1}{(M + 1)^2} = \frac{2m + 2}{(1 + m)^2(M + 2)}.$$

$$\mathbb{E}\left[(1 - V_1)^2\right] = \frac{M}{(M + 1)^2(M + 2)} + \frac{M^2}{(M + 1)^2} = \frac{M^2(M + 2) + M}{(1 + M)^2(M + 2)}.$$

Hence, using simple algebra and canceling of terms we get:

$$
\frac{\mathbb{E}\left[V_1^2\right]}{1-\mathbb{E}\left[(1-V_1)^2\right]} = \frac{\frac{2m+2}{(1+m)^2(M+2)}}{1-\frac{M^2(M+2)+M}{(1+M)^2(M+2)}}
$$

$$
= \frac{2M+2}{(M+1)^2(M+2)-M^2(M+2)-m}
$$

$$
= \frac{2M+2}{M+2+2M(M+2)+M^2(M+2)-M^2(M+2)-M}
$$

$$
= \frac{2M+2}{2+2M(M+2)}
$$

$$
= \frac{2M+2}{2M^2+4M+2}
$$

$$
= \frac{M+1}{M^2+2M+1}
$$

$$
= \frac{M+1}{(M+1)^2}
$$

$$
= \frac{1}{M+1}.
$$

So indeed we get

$$
\mathrm{Var}(FA) = \frac{\alpha(A)\alpha(A^c)}{1+M}
$$

If we have $F \sim \mathrm{DP}(M\alpha)$.

The statement of proposition 4.3 from [1] and its proof hold almost verbatim, we only need to change the reference from proposition [1, proposition 4.2] to the previous computations and update the terms for the variance. For completeness, we include the statement and the proof here. These lemmas concern the integration of functions with respect to measures drawn from a stick-breaking process prior.

Lemma 4.1.3. *If P is distributed according to a stick-breaking process with base measure α and stick-breaking weight distributions $\mathcal{D}_j$, then for any measurable functions ϕ, ψ for which the expression on the right-hand side is meaningful,*

$$
\mathbb{E}(P\phi) = \int \phi \, d\alpha,
$$

$$
\mathrm{Var}(\phi) = \int \left(\phi - \int \phi \, d\alpha\right)^2 d\alpha \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right],
$$

$$Cov(\phi, \psi) = (\int \phi\psi \, d\alpha - \int \phi \, d\alpha \int \psi \, d\alpha) \sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right].$$

Proof. We just apply the standard machinery for extending results about integration from indicator function to general integrable functions, and the results from Lemma 4.1.3. $\square$

Lemma 4.1.4. *If P is distributed according to a stick-breaking process with base measure α and stick-breaking weight distribution $\mathcal{D}$, then for any measurable functions ϕ, ψ for which the expression on the right-hand side is meaningful,*

$$\mathbb{E}(P\phi) = \int \phi \, d\alpha,$$

$$Var(\phi) = \frac{\int (\phi - \int \phi \, d\alpha)^2 \, d\alpha \mathbb{E}V_1^2}{1 - \mathbb{E}\left[(1-V_1)^2\right]},$$

$$Cov(\phi, \psi) = \frac{(\int \phi\psi \, d\alpha - \int \phi \, d\alpha \int \psi \, d\alpha)\mathbb{E}\left[V_1^2\right]}{1 - \mathbb{E}\left[(1-V_1)^2\right]}.$$

Proof. The extension of the result from the previous proposition to this proposition is just using Lemma 3.1.6, namely that $\sum_{i=1}^{\infty} \mathbb{E}\left[W_i^2\right] = \frac{\mathbb{E}\left[V_1^2\right]}{1-\mathbb{E}[(1-V_1)^2]}$. $\square$

4.2   Support of stick-breaking process.

We now write a theorem on the support of stick-breaking processes. We refer to [1, Lemma 3.6].

Lemma 4.2.1 (Support of stick-breaking process)**.** *If $(W_1, W_2, \cdots)$ are stick-breaking weights based on stick lengths $V_i \overset{iid}{\sim} \mathcal{D}$ for a fully supported measure $\mathcal{D}$ on $[0,1]$, independent of $\theta_i \sim \alpha$ where α has full support $\mathcal{X}$, then the stick-breaking process $SBP(\alpha, \mathcal{D})$ has full support $\mathcal{M}(X)$.*

We can expand this theorem a bit further to the case where the base measure is supported on a smaller set.

Lemma 4.2.2. *The weak support of a stick-breaking process Π with base measure α is $\mathcal{H} = \{P \in \mathcal{M}(X) : supp(P) \subset supp(\alpha)\}$.*

The proof works the same, just restricting all the sets we consider to the support of α, and observing that $\mathcal{H}$ is a closed set of probability 1. Also note that by incorporating the observation of these two lemmas together with [1, Lemma 3.5] we get

Lemma 4.2.3. *Suppose the stick-breaking distribution weights V_j are drawn from $\mathcal{D}_j$, every $\mathcal{D}_j$ is fully supported on $[0,1]$ and the base measure α is supported on A, then the stick-breaking process is fully supported on $\mathcal{H} = \{P \in \mathcal{M} : supp(P) \subset supp(\alpha)\}$.*

4.3   Exchangeable partition probability functions

Lemma 4.3.1. *The exchangeable partition probability function corresponding to a stick-breaking process with atomless base measure α and relative stick-breaking weights $\mathcal{D}$ is given by*

$$p(n_1, \cdots, n_k) = \sum_{\sigma \in S^k} \prod_{j=1}^{k} \frac{\mathbb{E}\left[V^{n_{\sigma(j)}}(1-V)^{\sum_{\sigma(i)>j} n_i}\right]}{1 - \mathbb{E}\left[(1-V)^{\sum_{\sigma(i)\geq j} n_i}\right]},$$

Where $V \sim \mathcal{D}$.

Proof. By the assumption that α is atomless, the stick-breaking process actually becomes a proper species sampling process. By Lemma 2.2.13 we know that for proper species sampling processes

$$p(n_1, \cdots, n_k) = \sum_{1 \leq i_1 \neq \cdots \neq i_k < \infty} \mathbb{E}\left[\prod_{j=1}^{k} W_{i_j}^{n_j}\right].$$

Here $W_j = V_j \prod_{i=1}^{j-1}(1-V_i)$ and $V_1, V_2, \cdots \overset{\text{iid}}{\sim} \mathcal{D}$. Hence we can rewrite this into

$$\sum_{1 \leq i_1 \neq \cdots \neq i_k < \infty} \mathbb{E}\left[\prod_{j=1}^{k} V_{i_j}^{n_j} \prod_{l=1}^{i_j-1}(1-V_l)^{n_j}\right].$$

Now observe that if we sum over $1 \leq i_1 \neq \cdots \neq i_k < \infty$, this is the same as summing over all permutations of $[k]$, and then summing over all assignments

$i_1 < i_2 < i_3 < \cdots$. This gives

$$\sum_{\sigma \in S^k} \sum_{1 \leq i_1 < \cdots < i_k < \infty} \mathbb{E}\left[\prod_{j=1}^{k} V_{i_j}^{n_{i_{\sigma l}}} \prod_{l=1}^{j-1}(1 - V_l)^{n_{i_{\sigma(l)}}}\right].$$

Now we know the powers p, q of $V_l^p(1 - V_l)^q$, namely $p = n_{\sigma_j}$ if there is an i_j such that $l = i_{\sigma(j)}$ and zero otherwise, and q is the sum of the counts n_j such that $l < i_{\sigma(j)}$, i.e. $\sum_{l:l < i_{\sigma(j)}} n_l$. This allows us to factorize the product, and then because all the $V_l \sim \mathcal{D}$ we can sum the geometric series appearing. This gives

$$p(n_1, \cdots, n_k) = \sum_{\sigma \in S^k} \prod_{j=1}^{k} \frac{\mathbb{E}\left[V^{n_{\sigma(j)}}(1 - V)^{\sum_{\sigma(i)>j} n_i}\right]}{1 - \mathbb{E}\left[(1 - V)^{\sum_{\sigma(i) \geq j} n_i}\right]}$$

$\square$

Chapter 5

The posterior distribution

5.1 The posterior of a species sampling process

In this section, we will see how to find a description of the posterior and some outline on how we want to proceed to try and give a nicer form for the posterior. We start by observing that a stick-breaking process is a special case of a *species sampling process* [1, Chapter 14], for which there is already an established theory on how to find a description for the posterior. This uses the *size-biased permutation* of the weights. However, size-biased permutations are complicated objects to give closed-form expressions for, so the goal is to also find a description which is easier to work with. [1, Theorem 14.18] states

Theorem 5.1.1. *The posterior distribution of F in the model with observations $X_1, \cdots, X_n | F \overset{iid}{\sim} F$ with F following a proper species sampling prior with the weight sequences $(W_j) = (\tilde{W}_j)$ in size-biased order is the distribution of*

$$\sum_{j=1}^{K_n} \hat{W}_j \delta_{\tilde{X}_j} + \sum_{j=K_n+1}^{\infty} \hat{W}_j \delta_{\hat{X}_j},$$

where $\tilde{X}_1, \cdots, \tilde{X}_{K_n}$ are the distinct values of $X_1, \cdots, X_n$ in order of appearance, $N_{j,n}$ the number of times $\tilde{X}_j$ appeared in the sample, the variables $\hat{X}_{K_n+1}, \cdots \overset{iid}{\sim} \alpha$ and $\hat{W} = (\hat{W}_j)$ is an independent vector with distribution

given by

$$\mathbb{E}\left[f(\hat{W})|X_1,\cdots,X_n\right] = \frac{\mathbb{E}\left[f(\tilde{W})\prod_{j=1}^{K_n}\tilde{W}_j^{N_{j,n}-1}\prod_{j=2}^{K_n}\left(1-\sum_{i<j}\tilde{W}_i\right)\right]}{p(N_n)},$$

where p is the EPPF of the species sampling model, and the expectation on the right-hand side is under the prior distribution of $(\tilde{W}_i)$.

Recall that the EPPF is by [1, theorem 14.14], same statement in Lemma 2.2.12 equal to

$$p(n_1,\cdots,n_k) = \mathbb{E}\left[\prod_{j=1}^{k}\tilde{W}_j^{n_j-1}\prod_{j=2}^{k}(1-\sum_{i\leq j}\tilde{W}_i)\right],$$

where $\tilde{W}$ again is the size-biased permutation of W.

This Theorem gives us a very implicit description of the posterior, however, we can also find an explicit posterior, where we do not have the problem of having to compute the size-biased permutation. This form of the posterior will enable us to get stronger statements on the behavior of the posterior, see the chapter on consistency.

5.2   Introduction

So the goal is to find a posterior distribution. It is enough to study the canonical process. In order to do this, we need to define all the objects. Let $(\mathcal{X},\mathfrak{X},\alpha)$ be a probability space. Define $\Theta = (\mathcal{X}\times[0,1])^{\mathbb{N}}$(i.e. the countable infinite product). This means $\Theta = \{(\theta_i,V_i)_{n\in\mathbb{N}}\}$ the space of all sequences. One then can view θ_i and V_i as maps from Θ to $\mathcal{X}$ and $[0,1]$ respectively, i.e. if $\omega = (x_1,v_1,x_2,v_2,\cdots)\in\Theta$, then $\theta_i(\omega)=x_i$ a $V_i(\omega)=v_i$.

We define a distribution Π on Θ via $\theta_i \overset{iid}{\sim} \alpha$ and $V_i \overset{iid}{\sim} \mathcal{D}$, where $\mathcal{D}$ is a probability measure on $[0,1]$ with density function f.

Π will be our prior, and we have a parametrization of the probability measures on $\mathcal{X}$ by $P_\theta(A) = \sum_{i=1}^{\infty}\delta_{\theta_i(\omega)}(A)V_i(\omega)\prod_{j=1}^{i-1}(1-V_j(\omega))$. For every θ,V we can form a probability distribution $P_{\theta,V}$ defined by $A\mapsto \sum_{i=1}^{\infty}\delta_{\theta_i}(A)V_i\prod_{j=1}^{j-1}(1-V_j)$.

We model $X|\theta,V \sim P_{\theta,V}$. Now in order to create the canonical probability on $\mathcal{X}\times\Theta$ we define the following:

$$\mathbb{P}(\theta\in B, X\in A) = \int_B P_\theta(A)\,\mathrm{d}\,\Pi(\theta).$$

We know an expression for P_θ, so we can jot this down:

$$\mathbb{P}(\theta \in B, X \in A) = \int_B \sum_{i=1}^{\infty} \delta_{\theta_i(\theta)}(A) V_i(\theta) \prod_{j=1}^{i-1}(1 - V_j(\theta)) \, \mathrm{d}\,\Pi(\theta).$$

We will start with deriving the posterior in the setting where we have just one observation. This allows us to start easy and makes the general proof easier to follow.

5.3 Derivation of posterior

So Θ is a complicated probability space, and we can best study this via cylindrical sets, so we look at sets $B = \{\theta_1 \in C_1, V_1 \in B_1, \cdots, \theta_n \in C_n, V_n \in B_n\}$. It is enough to study those, since these generate the σ-algebra of the infinite product. Thus we look at those.

So we start with the first lemma, which gives an expression for the probability that θ is in B and X is in A. This is what is what is needed to verify the claim that something is the posterior.

Lemma 5.3.1.

$$\mathbb{P}(\theta \in B, X \in A) = \alpha(A) \prod_{j=1}^{n} \alpha(C_j) \prod_{j=1}^{n} \int_{B_j} (1 - v)f(v) \, dv +$$

$$\sum_{i=1}^{n} \left(\prod_{k \neq i} \alpha(C_k) \right) \alpha(C_i \cap A) \left(\prod_{j=1}^{i-1} \int_{B_j} (1 - v)f(v) \, dv \right) \int_{B_i} vf(v) \, dv$$

$$\left(\prod_{j=i+1}^{n} \int_{B_j} f(v) \, dv \right)$$

Roughly speaking, one can see this as conditioning on V_1 up to V_n, then seeing what is the probability that X came from θ_1 up to θ_n and with what probability. You can actually prove this in this way, but then we need to modify the probability space so that we get a $\bar{\Theta} = \Theta \times \mathbb{N}$, introduce a random variable I so that $\mathbb{P}(I = n|\theta_1, V_1, \cdots) = V_n \prod_{j=1}^{n-1}(1 - V_j)$, and set $X|I = \theta_I$. Then you get that $X|P \sim P$ and can talk rigorously about X coming from θ_i or not. You should interpret I as the the random variable which remembers which θ_i was chosen when drawing the random variable X from P.

Proof.

$\mathbb{P}(\theta \in C, V \in B, X \in A) = \mathbb{P}(\theta_1 \in C_1, V_1 \in B_1, \cdots, \theta_n \in C_n, V_n \in B_n, X \in A).$

This is also equal to

$$\int_{C_1} \int_{B_1} \cdots \int_{C_n} \int_{B_n} \int_{\prod_{j=n+1}^{\infty}(\mathcal{X} \times [0,1])} \sum_{i=1}^{\infty} \delta_{\theta_i}(A) V_i \prod_{j=1}^{i-1}(1 - V_j)\, \mathrm{d}\bar{\theta}\, \mathrm{d}V_n \cdots \mathrm{d}V_1\, \mathrm{d}\theta_1$$

where $\bar{\theta}$ is integration over all the coordinates after the first n (or $2n$ depending how you count). We can split this sum into the first n terms, and the tail:

$$\sum_{i=1}^{\infty} \delta_{\theta_i}(A) V_i \prod_{j=1}^{i-1}(1 - V_j) = \sum_{i=1}^{n} \delta_{\theta_i}(A) V_i \prod_{j=1}^{i-1}(1 - V_j) +$$

$$\prod_{k=1}^{n}(1 - V_k) \sum_{i=n+1}^{\infty} \delta_{\theta_i}(A) V_i \prod_{j=n+1}^{i-1}(1 - V_j).$$

Then if we compute the innermost integral, we get

$$\int_{\prod_{j=n+1}^{\infty}(\mathcal{X} \times [0,1])} \sum_{i=1}^{n} \delta_{\theta_i}(A) V_i \prod_{j=1}^{i-1}(1 - V_j) +$$

$$\left(\prod_{k=1}^{n}(1 - V_k)\right) \sum_{i=n+1}^{\infty} \delta_{\theta_i}(A) V_i \prod_{j=n+1}^{i-1}(1 - V_j)\, \mathrm{d}\bar{\theta}.$$

We can use linearity to get that this also equals

$$\int_{\prod_{j=n+1}^{\infty}(\mathcal{X} \times [0,1])} \sum_{i=1}^{n} \delta_{\theta_i}(A) V_i \prod_{j=1}^{i-1}(1 - V_j)\, \mathrm{d}\bar{\theta}$$

$$+ \int_{\prod_{j=n+1}^{\infty}(\mathcal{X} \times [0,1])} \left(\prod_{k=1}^{n}(1 - V_k)\right) \sum_{i=n+1}^{\infty} \delta_{\theta_i}(A) V_i \prod_{j=n+1}^{i-1}(1 - V_j)\, \mathrm{d}\bar{\theta}.$$

And then independence to evaluate the first integral and simplify the second

$$\sum_{i=1}^{n} \delta_{\theta_i}(A) V_i \prod_{j=1}^{i-1}(1 - V_j) +$$

$$\left(\prod_{k=1}^{n}(1 - V_k)\right) \int_{\prod_{j=n+1}^{\infty}(\mathcal{X} \times [0,1])} \sum_{i=n+1}^{\infty} \delta_{\theta_i}(A) V_i \prod_{j=n+1}^{i-1}(1 - V_j)\, \mathrm{d}\bar{\theta}$$

Proposition 4.1.1 says that the integral in this expression is $\alpha(A)$, so we get

$$\sum_{i=1}^{n} \delta_{\theta_i}(A) V_i \prod_{j=1}^{i-1}(1-V_j) + \alpha(A) \prod_{k=1}^{n}(1-V_j).$$

So we can substitute this back into the original equation and compute those integrals. Using linearity we split this, and compute every term separately. First we compute

$$\int_{C_1}\int_{B_1} \cdots \int_{C_n}\int_{B_n} \alpha(A) \prod_{k=1}^{n}(1-V_j)\,\mathrm{d}\,\mathcal{D}(v_n)\,\mathrm{d}\,\alpha(\theta_n)\cdots\mathrm{d}\,\mathcal{D}(v_1)\,\mathrm{d}\,\alpha(\theta_1).$$

We use that $\mathcal{D}$ has a density, and independence to factor this integral, which leads the solution

$$\alpha(A) \prod_{j=1}^{n} \alpha(C_j) \prod_{j=1}^{n} \int_{B_j}(1-v)f(v)\,\mathrm{d}\,v$$

Then we tackle the integrals of the first kind, now let $i \in \{1, \cdots, n\}$, we want to compute

$$\int_{C_1}\int_{B_1} \cdots \int_{C_n}\int_{B_n} \delta_{\theta_i}(A) V_i \prod_{j=1}^{i-1}(1-V_j)\,\mathrm{d}\,\mathcal{D}(v_n)\,\mathrm{d}\,\alpha(\theta_n)\cdots\mathrm{d}\,\mathcal{D}(v_1)\,\mathrm{d}\,\alpha(\theta_1).$$

We again use independence and the fact that $\mathcal{D}$ has a density to get

$$\left(\prod_{k\neq i}\alpha(C_k)\right)\alpha(C_i\cap A)\left(\prod_{j=1}^{i-1}\int_{B_j}(1-v)f(v)\,\mathrm{d}\,v\right)\int_{B_i}vf(v)\,\mathrm{d}\,v\left(\prod_{j=i+1}^{n}\int_{B_j}f(v)\,\mathrm{d}\,v\right)$$

Combining all the terms of the sums yields

$$\mathbb{P}(\theta \in B, X \in A) = \alpha(A) \prod_{j=1}^{n} \alpha(C_j) \prod_{j=1}^{n} \int_{B_j}(1-v)f(v)\,\mathrm{d}\,v+$$

$$\sum_{i=1}^{n}\left(\prod_{k\neq i}\alpha(C_k)\right)\alpha(C_i\cap A)\left(\prod_{j=1}^{i-1}\int_{B_j}(1-v)f(v)\,\mathrm{d}\,v\right)\int_{B_i}vf(v)\,\mathrm{d}\,v$$

$$\left(\prod_{j=i+1}^{n}\int_{B_j}f(v)\,\mathrm{d}\,v\right)$$

$\square$

This finishes the first half of the preparation. Remember that the goal was to find a posterior probability $P(B|x)$. For $P(B|x)$ to be a posterior distribution the following must hold: $\mathbb{E}\left[P(E|X)\mathbb{1}_A(X)\right] = \mathbb{P}(\omega \in E, X \in A)$. We formulate a small lemma which allows easier computation of this expected value.

Lemma 5.3.2.

$$\mathbb{E}\left[P(E|X)\mathbb{1}_A(X)\right] = \int_A P(\omega \in E|x)\, d\alpha(x)$$

Proof. In order to compute this, we observe the following

$$\mathbb{E}\left[P(E|X)\mathbb{1}_A(X)\right] = \mathbb{E}\left[\mathbb{E}\left[P(E|X)\mathbb{1}_A(X)|\sigma(X)\right]\right]$$
$$\stackrel{\text{using}X|P\sim P}{=} \mathbb{E}\left[P(P(E|X)\mathbb{1}_A(X))\right]$$
$$\stackrel{Lemma\ 4.1.3}{=} \int_A P(E|x)\,\mathrm{d}\,\alpha(x)$$

$\square$

Here the lemma used is the expected value of integration with respect to Dirichlet Process/Stick-breaking process measure.

From this we can quite directly deduce the form the posterior has, namely

Lemma 5.3.3. *A version of the posterior distribution is*

$$P(\{\theta_1 \in C_1, V_1 \in B_1, \cdots, \theta_n \in C_n, V_n \in B_n\}|x) =$$

$$\prod_{i=1}^n \alpha(C_i) \prod_{i=1}^n \int_{B_i} (1-v)f(v)\, dv +$$

$$\sum_{i=1}^n \left(\prod_{k=1,k\neq i}^n \alpha(C_k))\mathbb{1}_{C_i}(x) \prod_{j=1}^{i-1} \int_{B_j} (1-v)f(v)\, dv \int_{B_i} vf(v)\, dv \prod_{j=i+1}^n \int_{B_i}^n f(v)\, dv \right)$$

Proof. You just use previous lemma to verify if

$$\mathbb{E}\left[P(B|X)\mathbb{1}_A(X)\right] = \mathbb{P}(\theta \in B, X \in A).$$

Thus we should compute $\int_A P(B|x)\,\mathrm{d}\,\alpha$. For this we observe that we can use linearity to take the summation outside, integration of terms with indicator function $\mathbb{1}_{C_i}(x)$ just replaces this term with $\alpha(C_i\cap A)$ and the summand without any indicator function gets a factor $\alpha(A)$. This results in the expression we derived for $\mathbb{P}(\theta_1 \in C_1, V_1 \in B_1, \cdots, \theta_n \in C_n, V_n \in B_n)$. $\square$

5.4 Generalizing to n observations

Let α be the base measure, and $\mathcal{D}$ the weight measure. We want to compute the posterior for n iid observations in the model $P \sim \mathrm{SBP}(\alpha, \mathcal{D})$ and $X_i | P \overset{\text{iid}}{\sim} P$. In order to do some we introduce some notation which makes this a lot easier.

We set $\Theta = (\mathcal{X} \times [0,1])^{\mathbb{N}} \times \mathbb{N}^n$.

5.4.1 Computing the probability

We again start with a lemma which allows us to find the probability that X is in A, θ is in B and V is in C. Again this is what we need to verify the claim that something is the posterior.

Theorem 5.4.1. *The probability* $\mathbb{P}(X \in A, \theta \in B, V \in C)$ *equals*

$$\sum_{(i_1, \cdots, i_n) \in \mathbb{N}^n} \left(\prod_{j \notin S} \alpha(C_j) \right) \left(\prod_{j \in S} \alpha\left(C_j \cap \bigcap_{k \in T_j} A_k\right) \right) \prod_{k=1}^{\infty} \int_{V_k \in B_k} (1-v)^{\#k} v^{\#'k} \, d\mathcal{D}(v)$$

where

- $S = S((i_1, \cdots, i_n)) = \{i_1, \cdots, i_n\}$. *Note* $|S|$ *need not be* n.

- *For all* $j \in S$: $T_j = \{k : i_k = j\}$ *all the different* k *such that* i_k *is equal to an element of* S, *so equal to one of the different values* $i_1, \cdots, i_n$ *attain.*

- $\#j = \sum_{k=1}^{n} \mathbb{1}_{j < i_k}$ *the number of* i_k *such that* j *is smaller than* i_k.

- $\#'j = \sum_{k=1}^{n} \mathbb{1}_{j = i_k}$ *the number of* i_k *such that* j *is equal to* i_k.

Proof. We introduce random variables, θ_i is the i-th $\mathcal{X}$ coordinate, V_i is the i-th $[0,1]$ coordinate and I_i is the i-th $\mathbb{N}$ coordinate. We put distributions on these coordinates. θ_i are all independently distributed according to α. Independently of this, we make V_i independent sample of $\mathcal{D}$. We say I_i are, conditional on θ, V, independently distributed according to $\mathbb{P}(I_i = n | \theta, V) = V_n \prod_{j=1}^{n-1} (1 - V_j)$. Then if we put $X_i = \theta_{I_i}$, we have that $X_i | \theta, V \overset{\text{iid}}{\sim} \sum_{k=1}^{\infty} \delta_{\theta_k} V_k \prod_{j=1}^{k-1} (1 - V_j)$.

We observe that to know the distribution of $\mathbb{P}(\theta \in C, V \in B, X \in A)$ it is enough to know it on the σ-cylinders generating the infinite product, so WLOG we assume $C = \prod_{j=1}^{\infty}\{\theta_j \in C_j\}$ and $B = \prod_{j=1}^{\infty}\{V_j \in B_j\}$, and $A = \prod_{j=1}^{n}\{X_j \in A_j\}$.

Then we want to compute $\mathbb{P}(\theta \in C, V \in B, X \in A)$. What we can do is actually compute

$$\sum_{(i_1,\cdots,i_n)\in\mathbb{N}^n} \mathbb{P}(\theta \in C, V \in B, X \in A, I = (i_1,\cdots,i_n)).$$

We can do this by conditioning on V, I, since for fixed $(i_1, \cdots, i_n)$ we know that the conditional probability $\mathbb{P}(\theta \in C, V \in B, X \in A, I = (i_1, \cdots, i_n))$ is equal to

$$\int_B \mathbb{P}(\theta \in C, X \in A | I = (i_1,\cdots,i_n), V = v)\mathbb{P}(I = (i_1,\cdots,i_n)|V = v)\,\mathrm{d}\,\mathcal{D}^{\infty}(v).$$

Then we know $\mathbb{P}(I = (i_1, \cdots, i_n)|V = v)$, namely $\prod_{k=1}^{n}\left(v_k \prod_{j=1}^{k-1}(1 - v_j)\right)$. This allows us to simplify $\mathbb{P}(\theta \in C, V \in B, X \in A, I = (i_1, \cdots, i_n))$ into

$$\int_B \mathbb{P}(\theta \in C, X \in A | I = (i_1,\cdots,i_n), V = v)\prod_{k=1}^{n}\left(v_k \prod_{j=1}^{k-1}(1 - v_j)\right)\mathrm{d}\,\mathcal{D}^{\infty}(v).$$

Now we observe that $\mathbb{P}(\theta \in C, X \in A | I = (i_1, \cdots, i_n), V = v)$ does not depend on V, so this simplifies into $\mathbb{P}(\theta \in C, X \in A | I = (i_1, \cdots, i_n))$, and we can compute this. We observe that if $I_i = n$, then $\theta_n = X_{I_i}$, and so that if we want that $\theta_n \in C_n$ and $X_i \in A_i$, we also must have $\theta_n \in A_i$. Thus we get the intersections between all the sets C_n and A_i such that $I_i = n$. Note that there can be multiple of those. In order to handle this we introduced the following notation:

- $S = S((i_1, \cdots, i_n)) = \{i_1, \cdots, i_n\}$. Note $|S|$ need not be n.

- $\#j = \sum_{k=1}^{n} \mathbb{1}_{j < i_k}$ the number of i_k such that j is smaller than i_k.

- $\#'j = \sum_{k=1}^{n} \mathbb{1}_{j=i_k}$ the number of i_k such that j is equal to i_k.

Using this notation, and using the remarks we can write down the conditional probability $\mathbb{P}(\theta \in C, X \in A | I = (i_1, \cdots, i_n))$:

$$\left(\prod_{j\notin S}\alpha(C_j)\right)\left(\prod_{j\in S}\alpha(C_j \cap \bigcap_{k\in T_j} A_k)\right).$$

Filling this back into our equation for $\mathbb{P}(\theta \in C, V \in B, X \in A, I = (i_1, \cdots, i_n))$ we get that this equals

$$\int_B \left(\prod_{j \notin S} \alpha(C_j) \right) \left(\prod_{j \in S} \alpha(C_j \cap \bigcap_{k \in T_j} A_k) \right) \prod_{k=1}^n \left(v_k \prod_{j=1}^{k-1} (1 - v_j) \right) \mathrm{d}\, \mathcal{D}^\infty(v).$$

Note that $\left(\prod_{j \notin S} \alpha(C_j) \right) \left(\prod_{j \in S} \alpha(C_j \cap \bigcap_{k \in T_j} A_k) \right)$ indeed does not depend on v so we can take it outside the integral, which yields

$$\left(\prod_{j \notin S} \alpha(C_j) \right) \left(\prod_{j \in S} \alpha(C_j \cap \bigcap_{k \in T_j} A_k) \right) \int_B \prod_{k=1}^n \left(v_k \prod_{j=1}^{k-1} (1 - v_j) \right) \mathrm{d}\, \mathcal{D}^\infty(v).$$

We can use the numbers $\#k$ and $\#'k$ to count how many factors v_j and $(1 - v_j)$ we get. This allows us the simplify the integral together with independence of the V_i and the fact that B consists of σ-cylinders. Then $\int_B \prod_{k=1}^n \left(v_k \prod_{j=1}^{k-1}(1 - v_j) \right) \mathrm{d}\, \mathcal{D}^\infty(v)$ equals

$$\prod_{k=1}^\infty \int_{V_k \in B_k} (1 - v)^{\#k} v^{\#'k} \, \mathrm{d}\, \mathcal{D}(v).$$

We can combine this to find that the probability $\mathbb{P}(\theta \in C, V \in B, X \in A, I = (i_1, \cdots, i_n))$ equals

$$\left(\prod_{j \notin S} \alpha(C_j) \right) \left(\prod_{j \in S} \alpha(C_j \cap \bigcap_{k \in T_j} A_k) \right) \prod_{k=1}^\infty \int_{V_k \in B_k} (1 - v)^{\#k} v^{\#'k} \, \mathrm{d}\, \mathcal{D}(v).$$

So all that is left to do is to sum over all the possible values $i_1, \cdots, i_n$ can attain to get that $\mathbb{P}(\theta \in C, V \in B, X \in A)$ equals

$$\sum_{(i_1, \cdots, i_n) \in \mathbb{N}^n} \left(\prod_{j \notin S} \alpha(C_j) \right) \left(\prod_{j \in S} \alpha(C_j \cap \bigcap_{k \in T_j} A_k) \right) \prod_{k=1}^\infty \int_{V_k \in B_k} (1-v)^{\#k} v^{\#'k} \, \mathrm{d}\, \mathcal{D}(v).$$

Now we can normalize the distribution $\prod_{k=1}^\infty \int_{V_k \in B_k} (1 - v)^{\#k} v^{\#'k} \, \mathrm{d}\, \mathcal{D}(v)$ by multiplying and dividing by $\prod_{k=1}^{\max_j i_j} \mathbb{E}(1 - v)^{\#k} v^{\#'k} \, \mathrm{d}\, \mathcal{D}(v)$. This introduces a probability distribution on $\mathbb{N}^n$, namely we can define

$$\mathbb{P}'(i_1, \cdots, i_n) \propto \mathbb{E} \left[\prod_{j=1}^n V_{i_j} \prod_{k=1}^{i_j - 1} (1 - V_k) \right].$$

This is what we will need to construct a description of the posterior.   □

5.4.2   Integration lemma

This lemma allows easier computation of the expectation which we need to verify for the posterior.

Lemma 5.4.2. *Let A be any set of the form $A = \{X_1 \in A_1, \cdots, X_n \in A_n\}$.* *$X_i | P \sim P$ iid from P, $P \sim \Pi$. Then*

$$
\begin{aligned}
&\mathbb{E}\left[f(X)\mathbb{1}_A(X)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[f(X)\mathbb{1}_A(X)|\theta, V\right]\right] \\
&= \sum_{(i_1,\cdots,i_n)\in\mathbb{N}^n} \mathbb{E}\left[f(\theta_{i_1},\cdots,\theta_{i_n})\mathbb{1}_A(\theta_{i_1},\cdots,\theta_{i_n})\right]\prod_{k=1}^{\infty}\int (1-v)^{\#k}v^{\#'k}\,d\mathcal{D}(v)
\end{aligned}
$$

Where we can compute the expectation for functions f which only care for the (ordered set) of $\hat{x}_i$ of distinct values X_i, F_i the indices where we observed $\hat{x}_i$ and n_k how often it appeared:

$$
\int_{\cap_{k\in F_1} A_k} \cdots \int_{\cap_{k\in F_m} A_k} \hat{f}((x_1,n_1),\cdots,(\hat{x}_m,n_k))\,d\alpha(x_m)\cdots d\alpha(x_1)
$$

where $\hat{f}((x_1,k_1),\cdots,(x_m,k_m))$ is the function we get if we just count how many appearances of x_1 occurred into k_1, etc.

Proof. Let A be any set of the form $A = \{X_1 \in A_1, \cdots, X_n \in A_n\}$. $X_i|P \sim P$

iid from P, $P \sim \Pi$. Then

$$\mathbb{E}\left[f(X)\mathbb{1}_A(X)\right] = \mathbb{E}\left[\mathbb{E}\left[f(X)\mathbb{1}_A(X)|\theta, V\right]\right]$$

$$= \mathbb{E}\left[\sum_{(i_1,\cdots,i_n)\in\mathbb{N}^n} f(\theta_{i_1},\cdots,\theta_{i_n})\mathbb{1}_A(\theta_{i_1},\cdots,\theta_{i_n}) \prod_{k=1}^{n}\left(V_{i_k}\prod_{j=1}^{i_k-1}(1-V_j)\right)\right]$$

$$= \sum_{(i_1,\cdots,i_n)\in\mathbb{N}^n} \mathbb{E}\left[f(\theta_{i_1},\cdots,\theta_{i_n})\mathbb{1}_A(\theta_{i_1},\cdots,\theta_{i_n}) \prod_{k=1}^{n}\left(V_{i_k}\prod_{j=1}^{i_k-1}(1-V_j)\right)\right]$$

$$= \sum_{(i_1,\cdots,i_n)\in\mathbb{N}^n} \mathbb{E}\left[f(\theta_{i_1},\cdots,\theta_{i_n})\mathbb{1}_A(\theta_{i_1},\cdots,\theta_{i_n})\right]\mathbb{E}\left[\prod_{k=1}^{n}\left(V_{i_k}\prod_{j=1}^{i_k-1}(1-V_j)\right)\right]$$

$$= \sum_{(i_1,\cdots,i_n)\in\mathbb{N}^n} \mathbb{E}\left[f(\theta_{i_1},\cdots,\theta_{i_n})\mathbb{1}_A(\theta_{i_1},\cdots,\theta_{i_n})\right]\prod_{k=1}^{\infty}\int(1-v)^{\#k}v^{\#'k}\,\mathrm{d}\,\mathcal{D}(v)$$

Note that we can simplify this a bit further, because we can simplify the expected value as well into

$$\int_{\cap_{k\in T_{S_1}}A_k}\cdots\int_{\cap_{k\in T_{S_m}}A_k}\hat{f}((x_1,|T_{s_1}|),\cdots,(x_m,|T_{S_m}|))\,\mathrm{d}\,\alpha(x_m)\cdots\mathrm{d}\,\alpha(x_1)$$

where $\hat{f}((x_1,k_1),\cdots,(x_m,k_m))$ is the function we get if we just count how many appearances of x_1 occurred into k_1, etc. $\qquad\square$

5.4.3   Partitions

So we are summing over all the possible assignments of $(i_1,\cdots,i_n)$. We can as well sum over all the partitions of $[n]$, and then sum over the possible assignments. Then what we can further do is sum over all the possible different sizes of the elements of the partition.

   Roughly speaking, when you observe $x_1,\cdots,x_n$, you do not see what values $i_1,\cdots,i_n$ take, but (if α is atomless) you observe a grouping of $i_1,\cdots,i_n$ into which i_k are the same. So you get a partition of $[n]$. Technically you get to know the partition almost surely. However, it turs out it is easier if we forget a bit of information, namely, we can just forget the exact order of the data, but just group on $\hat{x}_i$, the distinct observations in order of appearance, and k_i the number of x_j that attains the value $\hat{x}_i$.

We are going to sum over all the possible partitions, and then over all possible assignments of $i_1, \cdots, i_n$ which are equal on the partition. This requires some careful thought on how we sum, because we want to only each assignment $i_1, \cdots, i_n$ once. Because we actually can observe such a partition, we are in the situation we want to be. So let us rewrite both the probability and the expected value after the integration lemma into this form.

We start by forming a lemma which allows us to rewriting the sum of the indices $\{i_1, \cdots, i_n\}$ into something allows us to integrate properly. This lemma states that there is a bijection between two sets, namely the set of all partitions of size m of $[n] = \{1, \cdots, n\}$ times $\mathbb{N}^m$ which remembers all the i_j which have the same value, and then just remembers which m distinct value they attain, and the values of $i_1, \cdots, i_n$. We form this map by just looking at the value of i_1 and then look at which i_k are equal to this, then look at the first i_l which is different from i_1 and look at the value of i_l and look for which i_k are equal to i_l, etc.

Lemma 5.4.3. *Let*

$$\phi : \mathbb{N}^n \to \coprod_{m=1}^{n} (\{partitions \ of \ size \ m \ of \ [n]\} \times \{f_1, \cdots, f_m : f_1, \cdots, f_m \ all \ distinct \ in \ \mathbb{N}\})$$

be the map given above. Then this map is a bijection.

Proof. So we need to argue well definedness, injectivity and surjectivity.

The well definedness is clear from the fact that we group on those i_k which are the same, so we indeed map to

$$\coprod_{m=1}^{n} (\{\text{partitions of size } m \text{ of } [n]\} \times \{f_1, \cdots, f_m : f_1, \cdots, f_m \text{ all distinct } \in \mathbb{N}\}).$$

Now we tackle the injectivity and surjectvity. We will create an inverse map. Suppose we are given a partition of $[n]$, say $F_1, \cdots, F_m$ and natural numbers $f_1, \cdots, f_m$. Since partitions are themselves unordered, wlog we can assume A_k contains the smallest element of $[n]$ not contained in $\cup_{j=1}^{k-1} A_j$, by renaming the sets if needed. Define $c(k)$ the natural number so that $k \in F_{c(k)}$, which exists because $k \in [n]$ and the F form a partition of $[n]$. Then set $i_k = f_{c(k)}$. If we now apply ϕ, we look at the values i_k which are equal to $i_1 = f_1$. Since all f_k are distinct, we know this means that if $i_j = i_1 = f_1$, we must have $c(j) = 1$. Repeat this argument for all distinct values of i_k and we see that we indeed map $i_1, \cdots, i_n$ to $F_1, \cdots, F_m, f_1, \cdots, f_m$. □

Note that we need these properties, because otherwise, if we would apply this transformation we would be over- or undercounting. With this lemma we can continue the derivation of the posterior. We denote $\hat{X}_i$ for the i-th distinct observation here $i = 1, \cdots, m$, and n_i how often the observation $\hat{X}_i$ occurred. We again denote $\#j$ to be the count of all I_k such that $I_k > j$ and $\#'j$ to be the number of I_k such that $I_k = j$. We denote $\tilde{\#}j = \sum_{k=1}^m n_k \mathbb{1}_{j < i_k}$ and $\tilde{\#}'j = \sum_{k=1}^m n_k \mathbb{1}_{i_k = j}$.

Theorem 5.4.4 (Posterior of a stick-breaking process based on n observations). *The posterior of a stick-breaking process with atomless base measure α and stick-breaking weights $\mathcal{D}$ for observations $x_1, \cdots, x_n$ is, using the notation above, given by the following hierarchical model: draw $I = (i_1, \cdots, i_m)$ proportional to*

$$\mathbb{E}\left[\prod_{k=1}^m V_{i_k}^{n_k} \prod_{j=1}^{i_k - 1}(1 - V_j)^{n_k} \mathbb{1}_{\text{ all values of } i_1, \cdots, i_m \text{ are distinct}}\right]$$

Conditional on I,

- *Draw θ_i for $i \neq I_1, \cdots, I_m$ from α independently.*

- *Set $\theta_{I_i} = \hat{x}_i$.*

- *Define distribution $\tilde{\mathcal{D}}_j$ to be the map $A \mapsto \frac{\int_A (1-V)^{\tilde{\#}j} V^{\tilde{\#}'j} \, d\mathcal{D}(v)}{\int (1-V)^{\tilde{\#}j} V^{\tilde{\#}'j} \, d\mathcal{D}(v)}.$*

- *Draw V_j from $\tilde{\mathcal{D}}_j$, independently.*

Proof. The proof is just combining the previous lemmas. Recall that $\mathbb{P}(\theta \in C, V \in B, X \in A)$ equals

$$\sum_{(i_1, \cdots, i_n) \in \mathbb{N}^n} \left(\prod_{j \notin S} \alpha(C_j)\right) \left(\prod_{j \in S} \alpha(C_j \cap \bigcap_{k \in T_j} A_k)\right) \prod_{k=1}^\infty \int_{V_k \in B_k} (1-v)^{\#k} v^{\#'k} \, d\mathcal{D}(v).$$

We now apply the transformation ϕ on these sums. Now we have explicit

expressions for T_j, namely $F_1, \cdots, F_m$, $S = \{i_1, \cdots, i_m\}$.

$$\sum_{m=1}^{n} \sum_{\substack{F_1, \cdots, F_m, \\ \text{partion of } [n]}} \sum_{\substack{i_1, \cdots, i_m \\ \text{distinct}}} \left(\prod_{j \notin S} \alpha(C_j) \right)$$

$$\left(\prod_{j \in [m]} \alpha(C_{i_j} \cap \bigcap_{k \in F_j} A_k) \right) \prod_{k=1}^{\infty} \int_{V_k \in B_k} (1 - v)^{\#k} v^{\#'k} \, \mathrm{d} \, \mathcal{D}(v).$$

This is what we want to end up with after applying the integration lemma. So lets apply the integration lemma to the hypothesized posterior and verify.

Lets first expand the posterior for our sets A, B, C. So for fixed observations $x_1, \cdots, x_n$ we get a partition of distinct elements, so we know which C_j our X_i must lie in. We denote by Z the normalization constant for the probability

$$\mathbb{P}(I = (i_1, \cdots, i_m) = \mathbb{E} \left[\prod_{k=1}^{m} V_{i_k}^{n_k} \prod_{j=1}^{i_k - 1} (1 - V_j)^{n_k} \mathbb{1}_{\text{ all values of } i_1, \cdots, i_m \text{ are distinct}} \right] Z.$$

For fixed $I = (i_1, \cdots, i_m)$ we can expand this easily using the independence. This yields $\mathbb{1}_{C_{i_j}}(\hat{x}_i)$ for the probability that $\hat{x}_i$ lies in C_{i_j}, the θ_i each give a factor $\alpha(C_i)$ for $i \neq i_1, \cdots, i_m$, and we get the weights from V_j, all multiplied together. We then sum this times the probability of choosing this particular I, which yields

$$\sum_{\substack{i_1, \cdots, i_m \\ \text{distinct}}} \left(\prod_{j \notin S} \alpha(C_j) \right) \left(\prod_{j \in [m]} \mathbb{1}_{C_{i_j}}(\hat{x}_i) \right) \frac{\prod_{k=1}^{\infty} \int_{V_k \in B_k} (1 - v)^{\tilde{\#}k} v^{\tilde{\#}'k} \, \mathrm{d} \, \mathcal{D}(v)}{\prod_{k=1}^{\infty} \int (1 - v)^{\tilde{\#}k} v^{\tilde{\#}'k} \, \mathrm{d} \, \mathcal{D}(v)} \mathbb{P}(I = (i_1, \cdots, i_n)).$$

Note that this probability cancels the numerator and yields a factor Z. So we can simplify this further into

$$\sum_{\substack{i_1, \cdots, i_m \\ \text{distinct}}} \left(\prod_{j \notin S} \alpha(C_j) \right) \left(\prod_{j \in [m]} \mathbb{1}_{C_{i_j}}(\hat{x}_i) \right) \prod_{k=1}^{\infty} \int_{V_k \in B_k} (1 - v)^{\tilde{\#}j} v^{\tilde{\#}'j} \, \mathrm{d} \, \mathcal{D}(v) Z.$$

We now apply the integration lemma to this, but we reformulate into partitions again, using the final remark of the integration lemma. We can formulate the posterior in the form that it takes a partition of $[n]$ and values

$\hat{x}_1, \cdots, \hat{x}_m$, because this is the transformation we apply as the first step into computing the posterior. If we apply this, this leads to

$$\mathbb{E}\left[P(B|X_1, \cdots, X_n)\mathbb{1}_{A_1, \cdots, A_n}(x_1, \cdots, x_n)\right]$$

$$= \sum_{(i_1, \cdots, i_n) \in \mathbb{N}^n} \mathbb{E}\left[f(\theta_{i_1}, \cdots, \theta_{i_n})\mathbb{1}_A(\theta_{i_1}, \cdots, \theta_{i_n})\right] \prod_{k=1}^{\infty} \int (1-v)^{\#k} v^{\#'k} \, d\,\mathcal{D}(v)$$

$$= \sum_{m=1}^{n} \sum_{\substack{F_1, \cdots, F_m \\ \text{partition of } [n]}} \sum_{\substack{j_1, \cdots, j_m \\ \text{distinct}}} \int_{\cap_{k \in T_{S_1}} A_k} \cdots \int_{\cap_{k \in T_{S_m}} A_k} \sum_{\substack{i_1, \cdots, i_m \\ \text{distinct}}}$$

$$\left(\prod_{j \notin S} \alpha(C_j)\right) \left(\prod_{j \in [m]} \mathbb{1}_{C_{i_j}}(\hat{x}_j)\right) \prod_{k=1}^{\infty} \int_{V_k \in B_k} (1-v)^{\tilde{\#}j} v^{\tilde{\#}'j} \, d\,\mathcal{D}(v) Z \, d\,\alpha(x_m) \cdots d\,\alpha(x_1)$$

$$\prod_{k=1}^{\infty} \int (1-v)^{\tilde{\#}j} v^{\tilde{\#}'j} \, d\,\mathcal{D}(v)$$

Now we can take the sum over $i_1, \cdots, i_m$, and the products $\prod_{j \notin S} \alpha(C_j)$ and $\prod_k = 1^{\infty} \int_{V_k \in B_k} (1-v)^{\tilde{\#}j} v^{\tilde{\#}'j} \, d\,\mathcal{D}(v)Z$ outside the integrals and the sum over $j_1, \cdots, j_m$. Then we can evaluate the integrals. Because we only integrate $\mathbb{1}_{C_{i,j}}(\hat{x}_j)$ over $\cap_{k \in F_j} A_k$ with respect to measure α, we are left with the product $\prod_{j=1}^{m} \left(\alpha(C_{i_j} \cap \cap_{k \in F_j} A_k)\right)$. Doing this we get

$$= \sum_{m=1}^{n} \sum_{\substack{F_1, \cdots, F_m \\ \text{partition of } [n]}} \sum_{\substack{i_1, \cdots, i_m \\ \text{distinct}}} \left(\prod_{j \notin S} \alpha(C_j)\right) \int_{V_k \in B_k} (1-v)^{\tilde{\#}j} v^{\tilde{\#}'j} \, d\,\mathcal{D}(v)Z$$

$$\sum_{\substack{j_1, \cdots, j_m \\ \text{distinct}}} \prod_{j=1}^{m} \left(\alpha(C_{i_j} \cap \bigcap_{k \in F_j} A_k)\right) \prod_{k=1}^{\infty} \int (1-v)^{\tilde{\#}j} v^{\tilde{\#}'j} \, d\,\mathcal{D}(v)$$

Note that the product does not depend on the specific value of the $j_1, \cdots, j_m$, so we can take it in front, and then we are just computing $\frac{1}{Z}$, the normaliza-

tion constant. So this cancels and we end up with

$$= \sum_{\substack{m=1}}^{n} \sum_{\substack{F_1,\cdots,F_m \\ \text{partition of } [n]}} \sum_{\substack{i_1,\cdots,i_m \\ \text{distinct}}} \left(\prod_{j \notin S} \alpha(C_j) \right)$$

$$\int_{V_k \in B_k} (1-v)^{\tilde{\#}j} v^{\tilde{\#}'j} \, \mathrm{d}\,\mathcal{D}(v) \prod_{j=1}^{m} \left(\alpha(C_{i_j} \cap \bigcap_{k \in F_j} A_k) \right)$$

If we now reorder the terms we get what we want, namely

$$= \sum_{\substack{m=1}}^{n} \sum_{\substack{F_1,\cdots,F_m \\ \text{partition of } [n]}} \sum_{\substack{i_1,\cdots,i_m \\ \text{distinct}}} \left(\prod_{j \notin S} \alpha(C_j) \right)$$

$$\left(\prod_{j=1}^{m} \alpha(C_{i_j} \cap \bigcap_{k \in F_j} A_k) \right) \int_{V_k \in B_k} (1-v)^{\tilde{\#}j} v^{\tilde{\#}'j} \, \mathrm{d}\,\mathcal{D}(v).$$

Which is what we computed for $\mathbb{P}(\theta \in B, V \in C, X \in A)$.   □

5.5   Generalization to distinct distributions

We can generalize outside the scope of stick-breaking processes, where the V_j are still independent, but now each V_j has their own distribution $\mathcal{D}_j$. Then the same proof as above still holds, just with more notation to keep track of the different distributions. This leads to the following theorem:

Theorem 5.5.1 (Posterior of a stick-breaking process based on n observations)**.** *The posterior of a stick-breaking process with atomless base measure* α *and stick-breaking weights* $V_i \sim \mathcal{D}_i$ *for observations* $x_1, \cdots, x_n$ *is, using the notation above, given by the following hierarchical model: draw* $I = (i_1, \cdots, i_m)$ *proportional to*

$$\mathbb{E} \left[\prod_{k=1}^{m} V_{i_k}^{n_k} \prod_{j=1}^{i_k-1} (1-V_j)^{n_k} \mathbb{1}_{\text{ all values of } i_1,\cdots,i_m \text{ are distinct}} \right]$$

Conditional on I,

- *Draw θ_i for $i \neq I_1, \cdots, I_m$ from α independently.*

- *Set $\theta_{I_i} = \hat{x}_i$.*

- *Define distribution $\tilde{\mathcal{D}}_j$ to be the map $A \mapsto \frac{\int_A (1-V)^{\tilde{\#}_j} V^{\tilde{\#}'_j} \, d\mathcal{D}_j(v)}{\int (1-V)^{\tilde{\#}_j} V^{\tilde{\#}'_j} \, d\mathcal{D}_j(v)}$.*

- *Draw V_j from $\tilde{\mathcal{D}}_j$, independently.*

Chapter 6

Consistency

In this chapter we study the consistency of the posterior, both in the standard model and the mixtures case.

6.1 Consistency

Roughly, what we want from a good statistical model is that it will find the true distribution in some sense. Notice that just stating that eventually, with probability one we will find the true distribution will not work, because sample variance would change our estimate. So what we want to do is state that we can get arbitrarily close, in finite time, with probability tending to one as one gets more data. This also has the advantage that while the truth might not be a part of your parameter space, you can still talk about the parameters which are close to your true distribution.

In order to be able to speak about *consistency*, we need some topological notion of closeness. So we introduce the setting which always holds when talking about consistency and contraction rates. For every $n \in \mathbb{N}$, let $X^{(n)}$ be an observation in the sample space $(\mathcal{X}^n, \mathfrak{X}^n)$ with distribution P_θ^n indexed by a parameter θ belonging to a first countable topological space Θ. Given a prior Π on the Borel sets of Θ, we can act like the observations $X^{(n)}$ came from Π and form the posterior $\Pi_n\left(\cdot|X^{(n)}\right)$.

Definition 6.1.1. *The posterior distribution $\Pi_n\left(\cdot|X^{(n)}\right)$ is said to be **(weakly) consistent** at $\theta_0 \in \Theta$ if, for all open neighborhoods U of θ_0, $\Pi_n\left(U^c|X^{(n)}\right) \to 0$ in $\mathbb{P}_{\theta_0}^n$ probability, as $n \to \infty$. The posterior is said to be **strongly consistent** at $\theta_0 \in \Theta$ if this convergence is in the almost-sure sense.*

In general, we do not expect the family to be consistent, because it is to big, however, we would like to derive easy tests to tell if it will be consistent or not. For this, we hope to find a simple test, however, as the Pitman-Yor process shows, the test will not be very easy, as already this process which has a nice description fails to be consistent in case where the true distribution is continuous. The Pitman-Yor process is an example of the stick-breaking process with different stick-breaking weight distributions. We will see that for a large class of stick-breaking processes, namely those with a continuous density bounded away from zero, we do get consistency.

The stick-breaking processes are, as said before, an example of species sampling models. For species sampling models there is a theorem which classifies which distributions they get consistency for. However, the condition is a very abstract one which is hard to make into something concrete.

6.2 Posterior consistency

6.2.1 Introduction of the topology.

Here we introduce the topology of convergence of bounded measurable functions. We then compute the posterior integral of a bounded measurable function.

Definition 6.2.1. *We say that a sequence of measures μ_n converges **pointwise on bounded measurable functions** to a measure μ if for every bounded measurable function f we have $\mu_n(f) \to \mu(f)$.*

We will show that under some conditions on the distribution of V, the posterior distribution of the stick-breaking process converges almost surely to the true distribution in this setting.

We will use a simple lemma which gives a necessary and sufficient condition for consistency. This is [1, Lemma 6.4]

Lemma 6.2.2 (Consistency by functionals). *If Ψ is a set of measurable real functions on Θ so that*

$$\theta_m \to \theta_0, \quad if \quad \psi(\theta_m) \to \psi(\theta_0) \forall \psi \in \Psi,$$

then the posterior distribution is (strongly) consistent at θ_0 if for each $\psi(\theta)$ the induced posterior is strongly consistent at $\psi(\theta_0)$. If the functions ψ

are uniformly bounded, then the latter is equivalent to the pair conditions
$\mathbb{E}\left[(\psi(\theta)|X^{(n)}\right] \to \psi(\theta_0)$ *and* $var\left(\psi(\theta)|X^{(n)}\right) \to 0$, *in probability (or almost surely).*

Now because we can restrict the class of bounded functions to the class of uniform bounded functions by just dividing by the bound on every function if needed. This still generates the topology of pointwise convergence on bounded measurable functions. Thus in order to prove the consistency of the posterior with respect to this topology, it is enough to study the mean and variance of functionals under this posterior.

6.2.2   Posterior consistency in case of Species sampling process, Dirichlet processes and Pitman-Yor process

In the general case of posterior species sampling processes, not much is known about the consistency of the posterior distribution. We will introduce a general theorem on the consistency of posterior distributions, and then one specific for the case of a Dirichlet or a Pitman-Yor process prior. The theorem about the consistency of the Species sampling processes states. See [1, Theorem 14.19] for a proof. The predictive probability functions $p_j(\mathbf{n})$ are given by

$$p_j(\mathbf{n}) = \frac{p(\mathbf{n}^{j+})}{p(\mathbf{n})},$$

where p is the exchange probability partition function. We have written down an expression for this, namely

$$p(\mathbf{n}) = \mathbb{E}\left[\prod_{j=1}^{K} \tilde{W}_j^{n_j-1} \prod_{j=2}^{K}\left(1 - \sum_{i<j} \tilde{W}_i\right)\right].$$

And the notation $\mathbf{n}^{j+}$ means that we increase the j-th component of n by 1, and if j is $k+1$, where there are k distinct indices of $\mathbf{n}$, we mean the vector n with a 1 appended to the end.

Theorem 6.2.3. *Let S be the support of the discrete part P_0^d of the probability measure $P_0 = P_0^c + P_0^d$. If P follows a species sampling process prior with $PPF(p_j)$ satisfying, for nonnegative numbers $\alpha_n = O(1)$ and numbers $\delta_n =$*

O(1), and $N_{j,n}$ counting how many times the j-th distinct observation occurs in $X_1, \cdots, X_j$,

$$\sum_{j=1:\tilde{X}_j \in S}^{K_n} |p_j(N_n) - \frac{\alpha_n N_{j,n} + \delta_n}{n}| \to 0 \qquad a.s. \ [P_0^\infty],$$

$$\sum_{i=1}^{K_n}\sum_{j=1}^{K_n} |p_i(N_n)p_j(N_n^{i+} - p_i(N_n)p_j(N_n)| \to 0 \qquad a.s. \ [P_0^\infty],$$

Then the posterior distribution of P in the model $X_1, \cdots, X_n | P \overset{iid}{\sim} P$ is strongly consistent at P_0 relative to the topology of pointwise convergence on bounded measurable functions if both $a_n \to 1$ and $p_{K_n+1}(N_n) \to 0$ almost surely. The latter two conditions are necessary if $P_0^d \neq 0$ or $G \neq \frac{P_0^c}{||P_0^c||}$, respectively; and equivalent if P_0 is discrete. Furthermore, if $\alpha_n \to \alpha$ and $P_{K_n+1}(N_n) \to \gamma$, and either $p_{K_n+2}(N_n^{K_n+1}+) \to \gamma$ or $\gamma = 0$, then the posterior distribution tends to $\alpha p_0^d + \beta P_0^c + \gamma G$, for $\beta = \frac{1-\alpha||P_0^d||-\gamma}{||P_0^c||}$.

The two states conditions state that prediction probability functions $P_j(N_n)$ and the sample frequency $\frac{N_{j,n}}{n}$ should not differ to much, and that the variance in the model should go to zero almost surely. The problem with this theorem is that it is very difficult to identify where the model is consistent, we have conditions, but they are related to the true distribution. Furthermore, explicitly computing $p_j(N_n)$ and taking the limits might be hard. In case of the Pitman-Yor process, we can get a better result, however, we will not get the everywhere consistency. See [1, thm 14.38] for a proof..

Theorem 6.2.4. *If P follows a Pitman-Yor $PY(\sigma, M, G)$ process, then the posterior distribution of P in the model $X_1, \cdots, X_n | P \overset{iid}{\sim} P$ converges under P_0 relative to the weak tpology to $P_0^d + (1 - \sigma)P_0^c + \sigma\xi G$, for $\xi = ||P_0^c||$. In particular, the posterior distribution is consistent if and only if P_0 is discrete or G is proportional to P_0^c or $\sigma = 0$.*

As you can see, in this setting we have a very explicit condition for consistency, however, it is not consistent for all distributions. This makes the Pitman-Yor process less suitable for applications, as you can only apply this if you know that the true distribution will be discrete. Note that the Dirichlet process with atomless base measure is a Pitman-Yor process with $\sigma = 0$. In the general case where we do not need to have a atomless base measure, we also have a consistency result on the posterior of the Dirichlet process.

6.2.3   Main result

The rest of this section is focusing on proving the consistency theorem and stating a more general conjecture. The first theorem has a complete proof. For the conjecture the proof is similar but we need to strengthen one lemma. We give the strengthened version of that lemma as a conjecture. If we can actually prove this strengthened result, the rest of the lemmas can be modified very easily to show that the consistency result extends to every distribution.

Theorem 6.2.5 (The stick-breaking process is consistent under regularity conditions)**.** *Let α be an atomless measure. Suppose that the stick-breaking distribution $\mathcal{D}$ admits a density f such that there exists constants $a, b > 0$ and a twice continuously differentiable function g on $[0, 1]$ such that $f(v) = v^{a-1}(1-v)^{b-1}g(v)$ for $v \in (0, 1)$. Let P_0 be any distribution such that P_0 does not both have infinitely many points of positive probability and a continuous part. Then the posterior in the model $X_1, \cdots, X_n | P \sim P$, where $P \sim SBP(\alpha, \mathcal{D})$, as given in Theorem 5.4.4 is consistent with respect to the topology of pointwise convergence on bounded measurable functions at P_0.*

Conjecture 6.2.6 (The stick-breaking process is consistent under regularity conditions)**.** *Let α be an atomless measure. Suppose that the stick-breaking distribution $\mathcal{D}$ admits a density f such that there exists constants $a, b > 0$ and a twice continuously differentiable function g on $[0, 1]$ such that $f(v) = v^{a-1}(1-v)^{b-1}g(v)$ for $v \in (0, 1)$. Let P_0 be any distribution. Then the posterior in the model $X_1, \cdots, X_n | P \sim P$, where $P \sim SBP(\alpha, \mathcal{D})$, as given in Theorem 5.4.4 is consistent with respect to the topology of pointwise convergence on bounded measurable functions at P_0.*

We will give a rough sketch of the rest of this section. In the end it turns out that the posterior distribution of the sample weights W_{I_k} is the object we want to study. We want to know its asymptotic conditional mean and variance. We start the search by introducing lemmas which allow us to compute the limits. After we introduced the tools to compute the limits, we find explicit expressions for the conditional moments. Then we analyze the limiting behavior. In order to do this, we need to split the analysis into two cases, weights corresponding to the observations from the discrete part of the true distribution and to the total weight of the observations coming from the continuous part of the true distribution. Then, using the expressions

of the weights, we can start the proof of the main result, namely that the posterior mean of $P(f)$ converges to $P_0(f)$ and the posterior variance of $P(f)$ converges to zero, for f a bounded, non negative measurable function. This then completes the proof.

6.2.4 Complementary lemmas

We will be using a few lemmas which can be directly stated without referring to the details of the distribution we will be studying.

We start with a small lemma which allows us to compute some ratios of expected values. We will apply this lemma many times during computation of the consistency.

Lemma 6.2.7 (Ratio of expected values in case of beta distribution). *Suppose $V \sim \beta(a, b)$, then $\frac{\mathbb{E}[V^x(1-V)^y]}{\mathbb{E}[(V^z(1-V)^w]} = \frac{B(x+a,y+b)}{B(z+a,w+b)}$.*

Proof. It is enough to observe that $\mathbb{E}\left[(V^x(1-V)^y\right] = \frac{1}{B(a,b)} \int_0^1 v^{x+a}(1-v)^{y+b} \, d\,v$, and then the normalization constants cancel in the ratio. $\square$

We also introduce a small lemma which allows us to conclude that we indeed converge. We have a ratio of sums, and we know for every term in the sum that the ratios converge, so want that the ratio of sums converges.

Lemma 6.2.8 (Approximation of ratios of sums, part 1). *Let $a_{n,m}$ and $b_{n,m}$ be two positive sequences for $1 \le m \le M$ such that for all $1 \le m \le M$ we know $\frac{a_{n,m}}{b_{n,m}} \to \lambda$ as $n \to \infty$, then $\frac{\sum_{m=1}^M a_{n,m}}{\sum_{m=1}^M b_{n,m}} \to \lambda$.*

Proof. Consider $c_{n,m} = \frac{a_{n,m}}{b_{n,m}}$. Then we can look at the minimum and maximum of $c_{n,m}$ over m. This yields two sequences

$$d_n = \min_m c_{n,m}$$

and

$$D_n = \max_m c_{n,m}.$$

Clearly, both d_n and D_n converge to λ. Observe that $a_{n,m} = \frac{a_{n,m}}{b_{n,m}} b_{n,m}$ and hence

$$d_n \sum_m b_{n,m} \le \sum_m \frac{a_{n,m}}{b_{n,m}} b_{n,m} \le D_n \sum_m b_{n,m}$$

implies
$$d_n \sum_m b_{n,m} \leq \sum_m a_{n,m} \leq D_n \sum_m b_{n,m}.$$
Using this we can now divide both above and below by $\sum_m b_{n,m}$ to get
$$d_n \leq \frac{\sum_m a_{n,m}}{\sum_m b_{n,m}} \leq D_n$$
which, since both d_n and D_n converge to λ implies
$$\frac{\sum_m a_{n,m}}{\sum_m b_{n,m}} \to \lambda.$$

$\square$

The previous lemma is enough when working with only finite number of discrete observations. However, if the number of distinct observations grows, we need a stronger lemma. The next lemma is a fundamental lemma which allows us to extend the computation to a much larger class of priors, at the price of a more complicated statement.

Lemma 6.2.9 (Approximation of ratios of sums). *Let S_n be a sequence of sets. Let $f_n, g_n : S_n \to \mathbb{R}_+$ be two sequences of positive functions which have finite total sum:*
$$\sum_{s \in S_n} f_n(s) \ \text{and} \ \sum_{s \in S_n} g_n(s)$$
both exists. Suppose that for every n there exists a subset T_n of S_n such that
$$\lim_{n \to \infty} \frac{\sum_{s \in T_n} g_n(s)}{\sum_{s \in S_n} g_n(s)} = 1.$$
Suppose furthermore that
$$\sup_{n, s \in S_n} \frac{f_n(s)}{g_n(s)} = L < \infty$$
and that the lower and upper bounds
$$c_n = \inf_{s \in T_n} \frac{f_n(s)}{g_n(s)} \ \text{and} \ C_n = \sup_{s \in T_n} \frac{f_n(s)}{g_n(s)}$$
both converge to a limit λ. Then
$$\lim_{n \to \infty} \frac{\sum_{s \in S_n} f_n(s)}{\sum_{s \in S_n} g_n(s)} = \lambda.$$

Proof. Define $U_n = S_n \setminus T_n$. Observe that

$$\lim_{n \to \infty} \frac{\sum_{s \in U_n} g_n(s)}{\sum_{s \in S_n} g_n(s)} = 1.$$

Then we can rewrite

$$\frac{\sum_{s \in S_n} f_n(s)}{\sum_{s \in S_n} g_n(s)} = \frac{\sum_{s \in T_n} f_n(s)}{\sum_{s \in S_n} g_n(s)} + \frac{\sum_{s \in U_n} f_n(s)}{\sum_{s \in S_n} g_n(s)}$$

$$= \frac{\sum_{s \in T_n} f_n(s)}{\sum_{s \in T_n} g_n(s)} \frac{\sum_{s \in T_n} g_n(s)}{\sum_{s \in S_n} g_n(s)} + \frac{\sum_{s \in U_n} f_n(s)}{\sum_{s \in S_n} g_n(s)} \frac{\sum_{s \in U_n} g_n(s)}{\sum_{s \in S_n} g_n(s)}$$

Now we can compute the limits for each term. First compute the limit of $\frac{\sum_{s \in T_n} f_n(s)}{\sum_{s \in T_n} g_n(s)}$. We can give both upper and lower bounds by C_n and c_n, which both converge to λ, so the limit of $\frac{\sum_{s \in T_n} f_n(s)}{\sum_{s \in T_n} g_n(s)}$ is λ. The limit of $\frac{\sum_{s \in T_n} g_n(s)}{\sum_{s \in S_n} g_n(s)}$ is per hypothesis 1. The limit of $\frac{\sum_{s \in U_n} g_n(s)}{\sum_{s \in S_n} g_n(s)}$ is 0, and $0 < \frac{\sum_{s \in U_n} f_n(s)}{\sum_{s \in S_n} g_n(s)} < L$, so that

$$\frac{\sum_{s \in T_n} f_n(s)}{\sum_{s \in T_n} g_n(s)} \frac{\sum_{s \in T_n} g_n(s)}{\sum_{s \in S_n} g_n(s)} + \frac{\sum_{s \in U_n} f_n(s)}{\sum_{s \in S_n} g_n(s)} \frac{\sum_{s \in U_n} g_n(s)}{\sum_{s \in S_n} g_n(s)} \to \lambda.$$

This shows that

$$\frac{\sum_{s \in S_n} f_n(s)}{\sum_{s \in S_n} g_n(s)} \to \lambda.$$

$\square$

In view of this lemma, we are interested in finding good sets T_n of all the sets S_n which we are summing over. Here good means that there is some form of uniform convergence of $\frac{f_n(s)}{g_n(s)}$, and we are not ignoring to much of the cases. Finding these sets will be the main game we will be playing. Here S_n will be the sets of all possible permutations, $g_n(s)$ will be the probability of a permutation under I, and $f_n(s)$ will be coming from the expected value of W_{I_k} given a permutation σ.

Lemma 6.2.10 (asymptotic ratio of number of distinct observations from discrete support)**.** *Let K_n be the number of distinct observations of the discrete support from a distribution P_0. Then*

$$\lim_{n \to \infty} \frac{K_n}{n} = 0 \quad P_0 a.s.$$

Proof. Denote S the support of the continuous part of P_0. Let $\epsilon > 0$. Let $\mu_1, \cdots, \mu_M$ de distinct support points such that $\mathbb{P}(\{\mu_1, \cdots, \mu_M\} \cup S) > 1 - \frac{\epsilon}{3}$. Then there exists an $m > 0$ such that $\forall n > N$ the number of observations of μ_i in the data, $N_{i,n}$ satisfies, for all $1 \leq i \leq M$,

$$|\frac{N_{i,n}}{n} - \mathbb{P}(\mu_i)| < \frac{\epsilon}{3M}$$

by the strong law of large numbers, almost surely. This means that there are at most $M + \frac{2\epsilon}{3}n$ distinct observations of the discrete part. Picking $K' = \max(K, \frac{3M}{\epsilon})$, we find that for all $n > K'$ we have

$$\frac{K_n}{n} \leq \epsilon$$

Hence $\frac{K_n}{n} \to 0$ P_0 almost surely.   □

Lemma 6.2.11. *Let* $\frac{K_n}{n} \to 0$. *Then* $(1 + O(\frac{1}{n}))^{K_n} \to 1$.

Proof.
$$(1 + O(\frac{1}{n}))^{K_n} = \exp(K_n \log(1 + O(\frac{1}{n}))).$$

Now because the exponential function is continuous it is enough to show that

$$K_n \log(1 + O(\frac{1}{n})) \to 0.$$

For this observe that we can Taylor expand $x \mapsto \log(1+x)$ by $O(x)$. Applying this to something of $O(\frac{1}{n})$ yields that $\log(1 + O(\frac{1}{n})) = O(\frac{1}{n})$ and hence $K_n O(\frac{1}{n})$ converges to zero. This is what we wanted.   □

Next comes a lemma which now seems arbitrary, but is relevant to compute limits of products appearing.

Lemma 6.2.12. *Let* K_n *be a sequence of positive natural numbers such that* $\frac{K_n}{n} \to 0$. *Let* $p_{1,n}$ *be a sequence converging to* $p_1 > 0$. *Let* $a, b > 0$. *Then, uniformly in*

- $p_{i,n}$ *positive real numbers such that* $\sum_{i=1}^{K_n} p_{i,n} = 1$,

- *all permutations* σ *of* $[K_n]$,

$$\frac{np_{1,n} + a}{n\sum_{i:\sigma(i)>\sigma(k)} p_i + b} \prod_{j=1}^{\sigma(1)} \frac{n\sum_{i:\sigma(i)>\sigma(j)} p_i + b}{\sum_{i:\sigma(i)\geq\sigma(j)} p_i + a + b} \to p_1.$$

Proof. We rewrite every term in

$$\frac{np_{1,n} + a}{n\sum_{i:\sigma(i)>\sigma(k)} p_i + b} \prod_{j=1}^{\sigma(1)} \frac{n\sum_{i:\sigma(i)>\sigma(j)} p_i + b}{\sum_{i:\sigma(i)\geq\sigma(j)} p_i + a + b}$$

as follows:

$$\frac{np_{1,n} + a}{n\sum_{i:\sigma(i)>\sigma(k)} p_i + a + b} = \frac{np_{1,n} + b}{n\sum_{i:\sigma(i)>\sigma(k)} p_i + a + b} \frac{np_{1,n} + a}{np_{1,n} + b}$$

and

$$\frac{n\sum_{i:\sigma(i)>\sigma(j)} p_i + b}{\sum_{i:\sigma(i)\geq\sigma(j)} p_i + a + b} = \frac{n\sum_{i:\sigma(i)>\sigma(j)} p_i + b}{n\sum_{i:\sigma(i)\geq\sigma(j)} p_i + b} \frac{n\sum_{i:\sigma(i)\geq\sigma(j)} p_i + b}{n\sum_{i:\sigma(i)\geq\sigma(j)} p_i + a + b}.$$

Now note that

$$\frac{np_{1,n} + b}{n\sum_{i:\sigma(i)>\sigma(k)} p_i + a + b} \prod_{j=1}^{\sigma(1)} \frac{n\sum_{i:\sigma(i)>\sigma(j)} p_i + b}{n\sum_{i:\sigma(i)\geq\sigma(j)} p_i + b}$$

telescopes to

$$\frac{np_{1,n} + b}{n\sum_{i=1}^{K_n} p_{i,n} + b}$$

and since $\sum_{i=1}^{K_n} p_{i,n} = 1$ per assumption, we can simplify this to

$$\frac{np_{1,n} + b}{n + b} \to p_1.$$

Now we only have to look at the error factors, we want to show that these converge to one. The first factor is easy, one shows directly that

$$\frac{np_{1,n} + a}{np_{1,n} + b} \to 1.$$

For the second term

$$\prod_{j=1}^{\sigma(1)} \frac{n\sum_{i:\sigma(i)\geq\sigma(j)} p_i + b}{n\sum_{i:\sigma(i)\geq\sigma(j)} p_i + a + b}$$

some more work is needed. Clearly this product is bounded above by 1, since
every term is less than 1. Now if we look at the smallest factor, this is

$$\frac{n \sum_{i:\sigma(i) \geq \sigma(1)} p_i + b}{n \sum_{i:\sigma(i) \geq \sigma(1)} p_i + a + b}.$$

We can study the worst case over all σ, namely $\sigma(1) = K_n$. Then this factor
becomes

$$\frac{np_{1,n} + b}{np_{1,n} + a + b}$$

This gives

$$\prod_{j=1}^{\sigma(1)} \frac{n \sum_{i:\sigma(i) \geq \sigma(j)} p_i + b}{n \sum_{i:\sigma(i) \geq \sigma(j)} p_i + a + b} \geq \left(\frac{np_{1,n} + b}{np_{1,n} + a + b} \right)^{K_n}.$$

If we can show this converges to 1 we are done. Note that we can make this
term even smaller by decreasing b to zero. So we want to study

$$\left(\frac{np_{1,n}}{np_{1,n} + a} \right)^{K_n}$$

Note that $\frac{np_{1,n}}{np_{1,n}+a} = 1 + O(\frac{1}{n})$ so we can apply Lemma 6.2.11, hence

$$\left(\frac{np_{1,n}}{np_{1,n} + a} \right)^{K_n} \to 0$$

Which means that

$$\frac{np_{1,n} + a}{n \sum_{i:\sigma(i) > \sigma(k)} p_i + b} \prod_{j=1}^{\sigma(1)} \frac{n \sum_{i:\sigma(i) > \sigma(j)} p_i + b}{\sum_{i:\sigma(i) \geq \sigma(j)} p_i + a + b} \to p_1.$$

$\square$

We also want a lemma to control how bad some other terms can become.
We shall show they have no influence on the limit. We are primarily interested
in the case where k is equal to 1 or 2. However, this lemma will probably be
relevant (in some way) for future research (read a Bernstein-von Mises like
theorem, see [1, Chapter 12] and [4, chapter 10]).

Lemma 6.2.13. *Let K_n be a sequence of positive natural numbers. Let $p_{1,n}$ be a sequence converging to $p_1 > 0$, such that $np_{1,n}$ are natural numbers. Let k be a natural number. Let V be a random variable with density f such that there exists $a, b > 0$ and a continuous function g bounded away from zero such that $f(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$. Then, uniformly in*

- *$p_{i,n}$ positive real numbers such that $\sum_{i=1}^{K_n} p_{i,n} = 1$ and $np_{i,n}$ is a natural number for all i.*

- *all permutations σ of $[K_n]$*

$$\prod_{j=1}^{\sigma(1)} \frac{1 - \mathbb{E}\left[(1-V)^{k+n\sum_{i:\sigma(i)\geq j} p_{i,n}}\right]}{1 - \mathbb{E}\left[(1-V)^{n\sum_{i:\sigma(i)\geq j} p_{i,n}}\right]} \to 1.$$

Proof. First observe that

$$\frac{1 - \mathbb{E}\left[(1-V)^{k+j}\right]}{1 - \mathbb{E}\left[(1-V)^{j}\right]} \geq 1$$

since $\mathbb{E}\left[(1-V)^j\right]$ is decreasing in j. This means that if we multiply by all the missing factors $\frac{1-\mathbb{E}\left[(1-V)^{k+j}\right]}{1-\mathbb{E}[(1-V)^j]}$ for j natural numbers larger than $np_{1,n}$, i.e. we multiply by

$$\frac{\prod_{j=np_{1,n}}^{n} \frac{1-\mathbb{E}\left[(1-V)^{k+j}\right]}{1-\mathbb{E}[(1-V)^j]}}{\prod_{j=1}^{\sigma(1)} \frac{1-\mathbb{E}\left[(1-V)^{k+n\sum_{i:\sigma(i)\geq j} p_{i,n}}\right]}{1-\mathbb{E}\left[(1-V)^{n\sum_{i:\sigma(i)\geq j} p_{i,n}}\right]}}$$

which is larger than one, we get

$$\prod_{j=np_{i,n}}^{n} \frac{1 - \mathbb{E}\left[(1-V)^{k+j}\right]}{1 - \mathbb{E}\left[(1-V)^{j}\right]}.$$

This is a telescoping product, telescoping to

$$\frac{\prod_{i=1}^{k}(1 - \mathbb{E}\left[(1-V)^{n+k}\right])}{\prod_{i=1}^{k}(1 - \mathbb{E}\left[(1-V)^{np_{1,n}+k-1}\right])}.$$

This converges to 1, so we have an upper bound converging to 1, and a lower bound of one, of

$$\prod_{j=1}^{\sigma(1)} \frac{1 - \mathbb{E}\left[(1-V)^{k+n\sum_{i:\sigma(i)\geq j} p_{i,n}}\right]}{1 - \mathbb{E}\left[(1-V)^{n\sum_{i:\sigma(i)\geq j} p_{i,n}}\right]}.$$

Hence this also converges to 1, uniformly for all σ and all assignments of $p_{i,n}$.   □

We need a bounded variation like property, but then for products instead of sums. The following lemma gives sufficient conditions when this happens.

Lemma 6.2.14 (Product of fractions of evaluations of continuous function remains bounded). *Let g be a continuously differentiable function bounded away from zero, then*

$$\sup_{0 \le s_1 < t_1 < \cdots < s_k < t_k} \prod_{i=1}^{k} \frac{g(s_i)}{g(t_i)}$$

and

$$\sup_{0 \le s_1 < t_1 < \cdots < s_k < t_k} \prod_{i=1}^{k} \frac{g(t_i)}{g(s_i)}$$

are finite.

Proof. Because g is continuously differentiable and bounded away from zero, we know that $\log g$ is continuously differentiable, and hence is bounded variation. Now observe that

$$\sup_{0 \le s_1 < t_1 < \cdots < s_k < t_k} \prod_{i=1}^{k} \frac{g(t_i)}{g(s_i)} = \exp \sup_{0 \le s_1 < t_1 < \cdots < s_k < t_k} \sum_{i=1}^{k} (\log(g(t_i)) - \log(g(s_i))).$$

And $\sup_{0 \le s_1 < t_1 < \cdots < s_k < t_k} \sum_{i=1}^{k} (\log(g(t_i)) - \log(g(s_i)))$ is smaller than the total variation norm, which is

$$\sup_{0 \le s_1 < t_1 < \cdots < s_k < t_k} \sum_{i=1}^{k} |g(t_i) - g(s_i)| < \infty$$

Because $\log g$ is continuously differentiable.   □

We will use this property of variances in the proof of consistency.

Lemma 6.2.15. *If $(X_n)_{n \in \mathbb{N}}$ and $(Y_n)_{n \in \mathbb{N}}$ are two sequences of random variables such that $Var(X_n) \to 0$ and $Var(Y_n) \to 0$. Then $Var(X_n + Y_n) \to 0$.*

Proof. $\mathrm{Var}(X_n + Y_n) = \mathrm{Var}(X_n) + \mathrm{Var}(Y_n) + 2\mathrm{Cov}(X_n, Y_n)$. Per assumption the first two terms converge to zero, so only need to check that $\mathrm{Cov}(X_n, Y_n) \to 0$. For this observe that $|\mathrm{Cov}(X_n, Y_n)| \leq \sqrt{\mathrm{Var}(X_n)\mathrm{Var}(Y_n)}$. Thus the last term converges to zero as well. □

The following lemma is a technique to find convergent sequences. This is a useful way to construct convergent sequences if you can create a double sequence such that, if we take limits in the first coordinate, and then limits in the second coordinate, we converge to the required answer. This is useful to control error terms.

Lemma 6.2.16. *Let $(x_{n,k})_{n,k \in \mathbb{N}}$ be a double sequence of elements in a metric space (X, d) such that for all k*

$$\lim_{n \to \infty} x_{n,k} = x_k,$$

and

$$\lim_{k \to \infty} x_k = x.$$

Then there exists a sequence $(k_n)_{n \in \mathbb{N}}$ such that

$$\lim_{n \to \infty} x_{n,k_n} = x.$$

Proof. Let $(\epsilon_n)_{n \in \mathbb{N}}$ be any decreasing sequence converging to zero. We want to define K_n iteratively. For iteration $i = 0$ we pick $N_0 = 1$. In iteration $i > 0$ we pick $M_i > 0$ such that for all $k \geq M_i$ we know $d(x_k, x) < \frac{\epsilon_i}{2}$. Now pick $N_i > 0$ such that $\forall n > N_i$ we have $d(x_{n,M_i}, x_{M_i}) < \frac{\epsilon_i}{2}$. Set $k_{N_i} = M_i$ and for $N_{i-1} < n < N_i$ define $k_n = K_{N_{i-1}}$.

Now we show that $\lim_{n \to \infty} x_{n,K_n} = x$. Let $\epsilon > 0$. Then there exists an i such that $\epsilon_i < \epsilon$ (by convergence to zero of ϵ_i). Let $n > N_i$. Suppose $N_j \leq n < N_{j+1}$. Then $d(x_{n,k_n}, x_{k_n}) < \frac{\epsilon_j}{2}$ and $d(x_{k_n}, x) < \frac{\epsilon_j}{2}$ by construction.

$$\begin{aligned}
d(x_{n,k_n}, x) &\leq d(x_{n,k_n}, x_{k_n}) + d(x_{k_n}, x) \\
&\leq \frac{\epsilon_j}{2} + \frac{\epsilon_j}{2} \\
&\leq \epsilon_j < \epsilon_i < \epsilon
\end{aligned}$$

at $N_{i-1} \leq n < N_i$. Then K_n has the required property. □

We recall a result on uniform convergent functions. We will apply this a couple of times.

Lemma 6.2.17 (Sum of uniformly convergent functions is uniformly convergent). *Suppose we have two sequences of functions $f_n, g_n : S_n \to \mathbb{R}$ such that f_n converges uniformly to λ and g_n converges uniformly to μ. Then $f_n + g_n$ converges uniformly to $\lambda + \mu$.*

Proof. Let $\epsilon > 0$. Since

$$\sup_{x \in S_n} |f_n(x) + g_n(x) - \lambda - \mu| \leq \sup_{x \in S_n} |f_n(x) - \lambda| + \sup_{x \in S_n} |g_n(x) - \mu|$$

Now pick $N > 0$ such that both

$$\sup_{x \in S_n} |f_n(x) - \lambda| < \frac{\epsilon}{2}$$

and

$$\sup_{x \in S_n} |g_n(x) - \mu| \leq \frac{\epsilon}{2}.$$

These exists by uniform convergence of f_n and g_n to λ and μ respectively.   □

6.2.5   Preparation

In order to show the consistency of the posterior, we first want to study the posterior. We will always work in the case that the base measure is atomless. Recall that the posterior is given by: Draw $I = (i_1, \cdots, i_m)$ proportional to

$$\mathbb{E}\left[\prod_{k=1}^{m} V_{i_k}^{n_k} \prod_{j=1}^{i_k-1} (1 - V_j)^{n_k} \mathbb{1}_{\text{ all values of } i_1, \cdots, i_m \text{ are distinct}}\right]$$

Conditional on I,

- Draw θ_i for $i \neq I_1, \cdots, I_m$ from α independently.

- Set $\theta_{I_i} = \hat{x}_i$.

- Define distribution $\tilde{D}_j$ to be the map $A \mapsto \frac{\int_A (1-V)^{\tilde{\#}f} V^{n_k \tilde{\#}'j} \, d\mathcal{D}(v)}{\int (1-V)^{\tilde{\#}f} V^{\tilde{\#}'j} \, d\mathcal{D}(v)}$.

- Draw $\tilde{V}_j$ from $\tilde{\mathcal{D}}_j$, independently.

In order to show the consistency of the posterior, we first want to study the posterior itself in more details.

Definition of the relevant random variables.

We will now study the posterior distribution in more detail. We are going to find some descriptions of relevant random variables. Let m denote the number of distinct observations, n the total number of observations, and p_i the sample frequency. We can compute properties of the posterior in this setting.

In order to reduce the amount of explicitly writing the conditioning on random variables, we introduce the following notation. $\tilde{\mathbb{E}}[\cdot] = \mathbb{E}[\cdot|X_1, \cdots, X_n]$, $\tilde{\mathrm{Var}}(\cdot) = \mathrm{Var}(\cdot|X_1, \cdots, X_n)$ and $\tilde{\mathbb{P}}(\cdot) = \mathbb{P}(\cdot|X_1, \cdots, X_n)$.

Computation of the normalization constants

For the probabilities $\tilde{\mathbb{P}}(I = (i_1, \cdots, i_m))$ we have an expression up to normalization constants. We compute these constants now.

Lemma 6.2.18 (Distribution of I). *Suppose we have K_n distinct observations in $X_1, \cdots, X_n$. Then*

$$\tilde{\mathbb{P}}(I = (i_1, \cdots, i_m)) = \frac{\tilde{\mathbb{E}}\left[\prod_{l=1}^{m} V_{i_l}^{np_l} \prod_{j=1}^{i_l-1} (1 - V_j)^{np_l} \mathbb{1}_{\text{ all values of } i_1, \cdots, i_m \text{ are distinct}}\right]}{\displaystyle\sum_{\sigma \in S^{K_n}} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}} (1-V)^{n \sum_{k=l+1}^{m} p_{\sigma^{-1}(k)}}\right]}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n \sum_{k=l}^{m} p_{\sigma^{-1}(k)}}\right]}}$$

Proof. What we observe is that summing over $i_1, \cdots, i_m$ all distinct is the same as summing over all permutations $\sigma : [m] \to [m]$, and then summing over all $i_{\sigma(1)} < i_{\sigma(2)} < \cdots < i_{\sigma(m)}$. We can then also permute the weights p and then we are done.

In details

$$\sum_{\substack{i_1, \cdots, i_m \\ \text{all distinct}}} \tilde{\mathbb{E}}\left[\prod_{k=1}^{m} V_{i_k}^{n_k} \prod_{j=1}^{i_k-1} (1 - V_j)^{n_k} \mathbb{1}_{\text{ all values of } i_1, \cdots, i_m \text{ are distinct}}\right]$$

Introducing the permutations to fix the order of $i_1, \cdots, i_m$ then gives

$$= \sum_{\sigma \in S^{K_n}} \sum_{i_{\sigma(1)} \in \mathbb{N}} \sum_{i_{\sigma(2)} > i_{\sigma(1)}} \cdots \sum_{i_{\sigma(m)} > i_{\sigma(m-1)}} \mathbb{P}(I = i_1, \cdots, i_m)$$

If we now also look at the swapped weights we get

$$
= \sum_{\sigma \in S^{K_n}} \sum_{i_1 < i_2 < \cdots < i_m} \tilde{\mathbb{E}} \left[\prod_{l=1}^{m} V_{i_l}^{np_{\sigma^{-1}(l)}} \prod_{j=1}^{i_l-1} (1 - V_j)^{np_{\sigma^{-1}(l)}} \right]
$$

Can now factorize this using independence and counting how many times terms appear, gives

$$
= \sum_{\sigma \in S^{K_n}} \sum_{i_1 < \cdots < i_m} \prod_{l=1}^{m} \tilde{\mathbb{E}} \left[V^{np_{\sigma^{-1}(l)}} (1 - V)^{n \sum\limits_{k=l+1}^{m} p_{\sigma^{-1}(k)}} \right] \tilde{\mathbb{E}} \left[(1 - V)^{n \sum\limits_{k=l}^{m} p_{\sigma^{-1}(k)}} \right]^{i_l - i_{l-1} - 1}
$$

Now if we take this sum over the integers, we get that this equals

$$
= \sum_{\sigma \in S^{K_n}} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}} \left[V^{np_{\sigma^{-1}(l)}} (1 - V)^{n \sum\limits_{k=l+1}^{m} p_{\sigma^{-1}(k)}} \right]}{1 - \tilde{\mathbb{E}} \left[(1 - V)^{n \sum\limits_{k=l}^{m} p_{\sigma^{-1}(k)}} \right]} .
$$

$\square$

From this lemma we can also deduce what the distribution of the permutations σ are, namely

Corollary 6.2.19. *Suppose we have K_n distinct observations in $X_1, \cdots, X_n$. Then*

$$
\tilde{\mathbb{P}}(\Sigma = \sigma) = \frac{\prod\limits_{l=1}^{m} \dfrac{\tilde{\mathbb{E}} \left[V^{np_{\sigma^{-1}(l)}} (1-V)^{n \sum\limits_{k=l+1}^{m} p_{\sigma^{-1}(k)}} \right]}{1 - \tilde{\mathbb{E}} \left[(1-V)^{n \sum\limits_{k=l}^{m} p_{\sigma^{-1}(k)}} \right]}}{\sum\limits_{\tau \in S^{K_n}} \prod\limits_{l=1}^{m} \dfrac{\tilde{\mathbb{E}} \left[V^{np_{\tau^{-1}(l)}} (1-V)^{n \sum\limits_{k=l+1}^{m} p_{\tau^{-1}(k)}} \right]}{1 - \tilde{\mathbb{E}} \left[(1-V)^{n \sum\limits_{k=l}^{m} p_{\tau^{-1}(k)}} \right]}} .
$$

The next lemma which is stated is a conjecture. We can show that there should not be to many observations from the continuous part picked before any specific observation X from the discrete part of the distribution. To be

more precise, if the set of the permutations for which there are at most K_n observations from the continuous part in front of the location of X. This lemma is needed to get the consistency in case there is a continuous part and the support of the discrete distribution is infinite.

The lemma we want is

Conjecture 6.2.20. *Let m_n denote the number of distinct observations. Let A_n denote the sample of observations from the continuous part of the distribution. Define S_n the space of permutations of $[m_n]$. Let $\hat{X}_k$ be an observation from the discrete part of the distribution. There almost surely exists a sequence K_n such that $\frac{K_n}{n} \to 0$ and such that the sequence of sets*

$$T_n = \{\sigma \in S_n | \#\{i \in A_n : \sigma(i) < \sigma(k)\} \le K_n\}.$$

is of asymptotic conditional probability 1.

One can observe from reversing the analysis for the Dirichlet process that lemma holds if the stick-breaking weights are beta$(1, M)$ distributed. We want to show this holds in general.

What one can observe is that we can prove a weaker version of this statement in case we have finitely many points with positive probability. This is done in the relevant section. However, if we can prove this general lemma we can extend the proofs of consistency to the last two cases as well.

We introduce some more notation, in order to keep the summations in powers in check. We denote $T_{\sigma,r} = \sum_{k=r}^{m} n p_{\sigma^{-1}(k)}$.

Explicit expression for moments of W_{I_k}

Note that we can compute $\tilde{\mathbb{E}}\left[W_{I_k}^q\right]$ by using the tower property of the expectation. This yields

$$\tilde{\mathbb{E}}\left[W_{I_k}^q\right] = \tilde{\mathbb{E}}\left[\tilde{\mathbb{E}}\left[W_{I_k}|I\right]\right] = \sum_{\substack{i_1, \cdots, i_m \\ \text{all distinct}}} \tilde{\mathbb{P}}(I = i)\tilde{\mathbb{E}}\left[W_{I_k}|I = i\right].$$

We know an explicit formula for the distribution of $\tilde{\mathbb{P}}(I = i)$, namely

$$\tilde{\mathbb{P}}(I = i) = \frac{\tilde{\mathbb{E}}\left[\prod_{l=1}^{m} V_{i_l}^{np_l} \prod_{j=1}^{i_l-1} (1 - V_j)^{np_l} \mathbb{1}_{\text{all values of } i_1, \cdots, i_m \text{ are distinct}}\right]}{\sum_{\sigma \in S^{K_n}} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}} (1-V)^{T_{\sigma,l+1}}\right]}{1 - \tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}.$$

Thus in order to compute the moments of W_{I_k} it is useful to compute the conditional moments first. The next lemma gives an expression for the conditional moments.

Lemma 6.2.21 (q-th moment of W_{I_k} given I)**.** *If $I = (i_1, \cdots, i_m)$, this induces a permutation σ so that $i_{\sigma(1)} < \cdots < i_{\sigma(m)}$. Then define $j_1 = i_{\sigma(1)}$ and $j_l = i_{\sigma(l)} - i_{\sigma(l-1)} - 1$.*

$$
\tilde{\mathbb{E}}\left[W_{I_k}^q | I = i\right] = \frac{\tilde{\mathbb{E}}\left[V^{np_k+q}(1-V)^{T_{\sigma,\sigma^{-1}(k)+1}}\right]}{\tilde{\mathbb{E}}\left[V^{np_k}(1-V)^{T_{\sigma,\sigma^{-1}(k)+1}}\right]}
$$

$$
\frac{\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma^{-1}(k)}}\right]^{j_{\sigma(k)}}}{\tilde{\mathbb{E}}\left[(1-V)^{nT_{\sigma,\sigma^{-1}(k)}}\right]^{j_{\sigma(k)}}}
$$

$$
\prod_{j<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}
$$

$$
\prod_{j<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]^{j_l}}{\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]^{j_l}}.
$$

Proof. For this, we do the same trick, look at permutations to fix the order of $I_1, \cdots, I_m$ and then look at $i_{\sigma(1)}, i_{\sigma(2)} - i_{\sigma(1)} - 1, \cdots, i_{\sigma(m)} - i_{\sigma(m-1)} - 1$, call these $j_1 \cdots, j_m$. Then we get the following 4 terms appearing in the conditional expectation of W_{I_k}

- $\dfrac{\tilde{\mathbb{E}}\left[V^{np_k+q}(1-V)^{T_{\sigma,\sigma^{-1}(k)+1}}\right]}{\tilde{\mathbb{E}}\left[V^{np_k}(1-V)^{T_{\sigma,\sigma^{-1}(k)+1}}\right]}.$

- $\dfrac{\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma^{-1}(k)}}\right]^{j_{\sigma(k)}}}{\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,\sigma^{-1}(k)}}\right]^{j_{\sigma(k)}}}$

- for $l < \sigma(k)$ we get terms $\dfrac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}.$

- for $l < \sigma(k)$ we get terms $\dfrac{\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]}{\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]^{j_l}}.$

$\square$

Now we have the conditional moment of W_{I_k} so we can compute the unconditional (with respect to I) moment of W_{I_k} given the data.

Lemma 6.2.22 (q-th moment of W_{I_k})**.** *Suppose we have K_n distinct observations in $X_1, \cdots, X_n$. Then*

$$\tilde{\mathbb{E}}\left[W_{I_k}^q\right] =$$

$$\frac{\displaystyle\sum_{\sigma \in S^{K_n}} \frac{\tilde{\mathbb{E}}\left[V^{np_k+q}(1-V)^{T_{\sigma,\sigma^{-1}(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma^{-1}(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\displaystyle\sum_{\sigma \in S^{K_n}} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}.$$

Proof. Recall that by using the tower formula

$$\tilde{\mathbb{E}}\left[W_{I_k}^q\right] = \tilde{\mathbb{E}}\left[\tilde{\mathbb{E}}\left[W_{I_k}^q|I\right]\right] = \sum_{\substack{i_1,\cdots,i_m \\ \text{all distinct}}} \tilde{\mathbb{P}}(I=i)\tilde{\mathbb{E}}\left[W_{I_k}^q|I=i\right].$$

Now we can combine everything into one expression for $\tilde{\mathbb{E}}\left[W_{I_k}^q\right]$ using the sum formula, note that for fixed $\sigma, j_1, \cdots, j_m$ we can compute the product, and then we can cancel terms, this gets rid of the normalization constants in the expression for $\tilde{\mathbb{E}}\left[W_{I_k}^q|I\right]$. We first ignore the normalization constant, because this is a constant factor which we can add in the end. Then, for fixed σ, j we get a product of the following 6 factors

- $\tilde{\mathbb{E}}\left[V^{np_k+q}(1-V)^{T_{\sigma,\sigma^{-1}(k)+1}}\right]$.

- $\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma^{-1}(k)}}\right]$.

- for $l < \sigma(k)$ we get terms $\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l}}\right]$.

- for $l < \sigma(k)$ we get terms $\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]^{j_l}$.

- for $l > \sigma(k)$ we get terms $\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]$.

- for $l > \sigma(k)$ we get terms $\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]^{j_l}$.

So after summing over all the assignments of j, and using the geometric series, we get , for fixed σ the product of the following 6 factors

- $\tilde{\mathbb{E}}\left[V^{np_k+q}(1-V)^{T_{\sigma,\sigma^{-1}(k)+1}}\right]$.

- $\dfrac{1}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma^{-1}(k)}}\right]}$.

- for $l < \sigma(k)$ we get a term $\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]$.

- for $l < \sigma(k)$ we get a term $\dfrac{1}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]}$.

- for $l > \sigma(k)$ we get a term $\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]$.

- for $l > \sigma(k)$ we get a term $\dfrac{1}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}$.

Now if we multiply these together, sum them over all values of σ, and multiply by the normalization constant. This gives the claimed value.

$\square$

We also want to know the mixed first moment of $W_{I_k}W_{I_j}$, so we first compute the conditional first moment.

Lemma 6.2.23 (first moment of $W_{I_k}W_{I_j}$ given I). *If $I = (i_1, \cdots, i_{K_n})$, this induces a permutation σ so that $i_{\sigma(1)} < \cdots < i_{\sigma(K)n)}$. Then define $j_1 = i_{\sigma(1)}$ and $j_l = i_{\sigma(l)} - i_{\sigma(l-1)} - 1$. If $I_k > I_m$ denote $s = k$ and $t = m$ and otherwise $s = m$ and $t = k$. Then*

$$
\tilde{\mathbb{E}}\left[W_{I_k}W_{I_m} \mid I = i\right] = \frac{\tilde{\mathbb{E}}\left[V^{np_s+q}(1-V)^{T_{\sigma,\sigma^{-1}(s)+1}}\right]}{\tilde{\mathbb{E}}\left[V^{np_s}(1-V)^{T_{\sigma,\sigma^{-1}(s)+1}}\right]} \frac{\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma^{-1}(s)}}\right]^{j_{\sigma(s)}}}{\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,\sigma^{-1}(s)}}\right]}
$$

$$
\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]} \frac{\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]}{\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}^{j_l}
$$

$$
\frac{\tilde{\mathbb{E}}\left[V^{np_t+q}(1-V)^{T_{\sigma,\sigma^{-1}(t)+1}}\right]}{\tilde{\mathbb{E}}\left[V^{np_t}(1-V)^{T_{\sigma,\sigma^{-1}(t)+1}}\right]} \frac{\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma^{-1}(t)}}\right]^{j_{\sigma(t)}}}{\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,\sigma^{-1}(t)}}\right]}
$$

$$
\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]} \frac{\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]}{\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}^{j_l}.
$$

Proof. For this, we do the same trick, look at permutations to fix the order of $I_1, \cdots, I_{K_n}$ and then look at $i_{\sigma(1)}, i_{\sigma(2)} - i_{\sigma(1)} - 1, \cdots, i_{\sigma(m)} - i_{\sigma(K_n-1)} - 1$, call these $j_1 \cdots, j_{K_m}$. We do the case where $I_k > I_m$, the other case follows by swapping roles. Then we get the following 8 terms appearing in the definition of W_{I_k}

- $\dfrac{\tilde{\mathbb{E}}\left[V^{np_k+q}(1-V)^{T_{\sigma,\sigma^{-1}(k)+1}}\right]}{\tilde{\mathbb{E}}\left[V^{np_k}(1-V)^{T_{\sigma,\sigma^{-1}(k)+1}}\right]}$.

- $\dfrac{\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma^{-1}(k)}}\right]^{j_{\sigma(k)}} \, j_{\sigma(k)}}{\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,\sigma^{-1}(k)}}\right]}$

- for $\sigma(m) < l < \sigma(k)$ we get terms $\dfrac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}$.

- for $l < \sigma(k)$ we get terms $\dfrac{\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]^{j_l}}{\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}$.

- $\dfrac{\tilde{\mathbb{E}}\left[V^{np_m+q}(1-V)^{T_{\sigma,\sigma^{-1}(m)+1}}\right]}{\tilde{\mathbb{E}}\left[V^{np_m}(1-V)^{T_{\sigma,\sigma^{-1}(m)+1}}\right]}$.

- $\dfrac{\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma^{-1}(m)}}\right]^{j_{\sigma(m)}} \, j_{\sigma(m)}}{\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,\sigma^{-1}(m)}}\right]}$

- for $l < \sigma(m)$ we get terms $\dfrac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}$.

- for $l < \sigma(m)$ we get terms $\dfrac{\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]^{j_l}}{\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}$.

$\square$

We now know the conditional first mixed moment, which allows us to compute the conditional first mixed moment.

Lemma 6.2.24 (first moment of $W_{I_k}W_{I_m}$)**.** *Let k and m be natural numbers. Suppose we have K_n distinct observations in $X_1, \cdots, X_n$. Then Note that the*

sums run into the second line of the expression.

$$\tilde{\mathbb{E}}\left[W_{I_k}W_{I_m}\right] =$$

$$\frac{\sum\limits_{\sigma\in S^{K_n}}\frac{\tilde{\mathbb{E}}\left[V^{np_s+1}(1-V)^{T_{\sigma,\sigma(s)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,\sigma(s)}}\right]}\frac{\tilde{\mathbb{E}}\left[V^{np_t+1}(1-V)^{T_{\sigma,\sigma(t)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,\sigma(t)}}\right]}\prod\limits_{l=1}^{\sigma(t)-1}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\sum\limits_{\sigma\in S^{K_n}}\prod\limits_{l=1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\frac{\prod\limits_{l=\sigma(t)+1}^{\sigma(s)-1}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]}\prod\limits_{l=\sigma(s)+1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}$$

Proof. Recall that by using the tower formula

$$\tilde{\mathbb{E}}\left[W_{I_k}W_{I_m}\right] = \tilde{\mathbb{E}}\left[\tilde{\mathbb{E}}\left[W_{I_k}W_{I_m}|I\right]\right] = \sum_{\substack{i_1,\cdots,i_{K_n}\\ \text{all distinct}}}\tilde{\mathbb{P}}(I=i)\tilde{\mathbb{E}}\left[W_{I_k}W_{I_m}|I=i\right].$$

Now we can combine everything into one expression for $\tilde{\mathbb{E}}\left[W_{I_k}W_{I_m}\right]$ using the sum formula, note that for fixed $\sigma, j_1, \cdots, j_{K_m}$ we can compute the product, and then we can cancel terms, this gets rid of the normalization constants in the expression for $\tilde{\mathbb{E}}\left[W_{I_k}^q|I\right]$. We first ignore the normalization constant, because this is a constant factor which we can add in the end. Then, for fixed σ, j we get a product of the following 10 factors

- $\tilde{\mathbb{E}}\left[V^{np_s+1}(1-V)^{T_{\sigma,\sigma(s)+1}}\right]$.

- $\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,\sigma(s)}}\right]^{j_s}$.

- $\tilde{\mathbb{E}}\left[V^{np_t+1}(1-V)^{T_{\sigma,\sigma(t)+1}}\right]$.

- $\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,\sigma(t)}}\right]^{j_l}$.

- for $l < \sigma(t)$ we get terms $\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l}}\right]$.

- for $l < \sigma(t)$ we get terms $\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]^{j_l}$.

- for $\sigma(t) < l < \sigma(s)$ we get terms $\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}}\right]$.

- for $\sigma(t) < l < \sigma(s)$ we get terms $\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]^{j_l}$.

- for $l > \sigma(s)$ we get terms $\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]$.

- for $l > \sigma(s)$ we get terms $\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]^{j_l}$.

So after summing over all the assignments of j, and using the geometric series, we get , for fixed σ the product of the following 10 factors

- $\tilde{\mathbb{E}}\left[V^{np_s+1}(1-V)^{T_{\sigma,\sigma(s)+1}}\right]$.

- $\dfrac{1}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,\sigma(s)}}\right]}$.

- $\tilde{\mathbb{E}}\left[V^{np_t+1}(1-V)^{T_{\sigma,\sigma(t)+1}}\right]$.

- $\dfrac{1}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,\sigma(t)}}\right]}$.

- for $l < \sigma(t)$ we get a term $\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]$.

- for $l < \sigma(t)$ we get a term $\dfrac{1}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}$.

- for $\sigma(t) < l < \sigma(k)$ we get a term $\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]$.

- for $\sigma(t)l < \sigma(k)$ we get a term $\dfrac{1}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]}$.

- for $l > \sigma(k)$ we get a term $\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]$.

- for $l > \sigma(k)$ we get a term $\dfrac{1}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}$.

Now if we multiply these together, sum them over all values of σ, and multiply by the normalization constant. This gives the claimed value. $\qquad\square$

We formulate a corollary from this discussion, which states that weights corresponding to observations which have the same sample frequency are identically distributed.

Corollary 6.2.25. *Suppose $N_i = N_j$. Then both I_i and I_j are have the same distribution and W_{I_i} and W_{I_j} are have the same distribution.*

Proof. Note that the distribution of I_i and I_j only cares about all the permutations and how many weight we assign to them, by Lemma 6.2.18 we see that the I_i and I_j are the same. Similarly, if we look to distribution of the weights, all that matters is how many times all the distinct observations have occurred. This shows that the distribution of $W_{I_i} = W_{I_j}$. $\qquad\square$

Note that the expressions for the first and second moment can alternatively be derived from the predictive probability functions and the expression for the EPPF as given in Lemma 4.3.1.

Lemma 6.2.26 (Bounded errors)**.** *The point-wise ratios in the fractions for the q-th moments remain bounded:*

$$
\frac{\frac{\tilde{\mathbb{E}}\left[V^{np_k+q}(1-V)^{T_{\sigma,\sigma(k)}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]}\prod_{l<\sigma(k)}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]}\prod_{l>\sigma(k)}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod_{l=1}^{m}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}\le 1.
$$

Proof. Observe that in numerator the terms in the expected value have nowhere lower exponent and in some places higher exponents. Because the expected value $\tilde{\mathbb{E}}\left[V^x(1-V)^y\right]$ is nonincreasing in x and y we get that the numerator is not larger than the denominator. This shows that the fraction is less than 1, which is what we wanted to show. $\square$

6.2.6   Finding well behaved sets of permutations

In order to apply many of our lemmas, we need to have at most K_n terms in the product, with $\frac{K_n}{n}$ converging to zero. For this we are going to prove a weaker version of the conjecture.

We first start by finding an explicit distribution of I.

Lemma 6.2.27 (Distribution of I)**.** *Suppose that there are m points which occur more than once, and $\lfloor \lambda n \rfloor$ points which occur exactly once. Then*

$$
\tilde{\mathbb{P}}(I=(i_1,\cdots,i_{m+\lfloor\lambda n\rfloor}))=
$$

$$
\frac{\tilde{\mathbb{E}}\left[\prod_{l=1}^{m+\lfloor\lambda n\rfloor}V_{i_l}^{np_l}\prod_{j=1}^{i_l}(1-V_j)^{np_l}\mathbb{1}_{\,all\ values\ of\ i_1\cdots,i_{m+\lfloor\lambda n\rfloor}\ are\ distinct}\right]}{\displaystyle\sum_{\sigma\in S^m}\ \sum_{\substack{j_0,\cdots,j_m\\ summing\ to\ \lfloor\lambda n\rfloor}}\prod_{k=1}^{m}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(k)}}(1-V)^{j_{\sigma(k)}+T_{\sigma,j,k+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,j,k}}\right]}\prod_{k=0}^{m}\prod_{r=0}^{j_{\sigma(k)}}\frac{\tilde{\mathbb{E}}\left[V(1-V)^{r+T_{\sigma,j,k+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{r+1+T_{\sigma,j,k+1}}\right]}}.
$$

Proof. The cleanest way to prove this, is to observe that for a fixed permutation, we get streaks of $j_0,\cdots,j_m$ weights corresponding to points we observed only once. These all have observed sample frequency $\frac{1}{n}$, so filling this in gives the claimed result. $\square$

This technical lemma is in essence the proof of a weakened version of the conjecture. We first do this for the case where the relative stick-breaking weights are beta distributed, and then extend the result to the general case.

Lemma 6.2.28 (No large values of $j_0, \cdots, j_{m-1}$). *Let k_n be a sequence such that $k_n \to \infty$. Suppose $V \sim Be(a, b)$. Then for all σ we will have*

$$\frac{\displaystyle\sum_{\substack{j_0,\cdots,j_{m-1}\leq k_n, j_m \geq 0 \\ \text{summing to } \lfloor \lambda n \rfloor}} \prod_{k=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(k)}}(1-V)^{j_{\sigma(k)}+T_{\sigma,j,k+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,j,k}}\right]} \prod_{k=0}^{m}\prod_{r=0}^{j_{\sigma(k)}} \frac{\tilde{\mathbb{E}}\left[V(1-V)^{r+T_{\sigma,j,k+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{r+1+T_{\sigma,j,k+1}}\right]}}{\displaystyle\sum_{\substack{j_0,\cdots,j_m \\ \text{summing to } \lfloor \lambda n \rfloor}} \prod_{k=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(k)}}(1-V)^{j_{\sigma(k)}+T_{\sigma,j,k+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,j,k}}\right]} \prod_{k=0}^{m}\prod_{r=0}^{j_{\sigma(k)}} \frac{\tilde{\mathbb{E}}\left[V(1-V)^{r+T_{\sigma,j,k+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{r+1+T_{\sigma,j,k+1}}\right]}}$$

converges to 1.

Proof. Fix $\sigma \in S^m$. We start with the most extreme case, $j_m = \lfloor \lambda n \rfloor$, and the other j_i are zero. Then swapping to the case where there are j_i' weights, we can compute this as follows: first we move all the weights from j_m to j_0', then to j_1', etc. This makes computation easier. Then observe, if we decrease j_k by one, and increase j_r by one, for $k > r$, the only thing that happens is that at the end of the j_k one weight disappears, all the weights between the k-th observation and r-th observation of the discrete part (after reordering via the permutation), have the exponent of the $(1 - V)$ component decrease by 1, and there appears a new term in the streak j_r, in the beginning. Nothing else changes, because before the leftmost observation of j_k we do not measure this value, as well as after the rightmost value of j_k. Also note that in the most extreme case the terms

$$\frac{1}{1 - \cdots}$$

are as large as they can get, so we can safely ignore these terms. Using this, we can compute what happens when we swap a continuous observation from j_m to j_k, when there are no weights from the continuous part between the k-th and the m-th discrete weight, we just get a factor

$$\frac{\tilde{\mathbb{E}}\left[V(1-V)^{T_{\sigma,j,k,n}}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{j_m-1}\right]} \prod_{r=k}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma(r)}}(1-V)^{j_k+T_{\sigma,j,k+1,n}}\right]}{\tilde{\mathbb{E}}\left[V^{np_{\sigma(r)}}(1-V)^{j_k-1+T_{\sigma,j,k+1,n}}\right]}.$$

The latter term we can compute as before. Because there are no continuous observations between the discrete weights, the factor we get from

$$\prod_{r=k}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma(r)}}(1-V)^{j_k+T_{\sigma,j,k+1,n}}\right]}{\tilde{\mathbb{E}}\left[V^{np_{\sigma(r)}}(1-V)^{j_k-1+T_{\sigma,j,k+1,n}}\right]}.$$

It is easy to compute, namely it is (asymptotic to)

$$\prod_{r=k}^{m} \frac{j_m + n\sum_{l=r+1}^{m} p_{\sigma(l),n}}{j_m + n\sum_{l=r}^{m} p_{\sigma(l),n}}.$$

This telescoping product, and is telescoping to

$$\frac{j_m}{j_m + n\sum_{l=k}^{m} p_{\sigma(l),n}}.$$

We can now pick the largest value that this can attain as an upper bound. This then becomes asymptotic to

$$\frac{\nu}{\nu + \sum_{l=k}^{m} q_{\sigma(l)}}.$$

If we want to bound the first term, again note that this term is largest when j_m is large. So then we apply Stirlings approximation to find what this is asymptotic to. This yields

$$\left(\frac{\nu + \sum_{l=k}^{m} q_{\sigma(l)}}{\nu}\right)^{1+\alpha}.$$

Thus in total we can (asymptotically) bound the mass you gain by doing one such swap from above by

$$\left(\frac{\nu}{\nu + \sum_{l=k}^{m} q_{\sigma(l)}}\right)^{\alpha}.$$

If we would now swap from the case where $j_0, \cdots, j_{m-1}$ are all zero and $j_m = \lfloor \lambda n \rfloor$ to all other possible assignments, it is just summing

$$\sum_{\substack{j_0,\cdots,j_m \\ \text{summing to } \lfloor \lambda n \rfloor}} \prod_{r=0}^{m-1} \left(\left(\frac{\nu}{\nu + \sum_{l=k}^{m} q_{\sigma(l)}}^{\alpha}\right)\right)^{j_r}.$$

If we now ignore the restriction on the sum and just sum over all the naturals, we can sum this to get another upper bound using geometric sums, which yields

$$\prod_{r=0}^{m-1} \frac{1}{1 - \left(\frac{\nu}{\nu+\sum_{l=k}^{m} q_{\sigma(l)}}\right)^{\alpha}}.$$

So we have bounded mass. Now we are looking at the first k_n observations, where $k_n \to \infty$. Since we actually have that every new term in the products is decreasing and smaller than the stated upper bound we are actually going to converge faster to the limit. So we are going to get asymptotic ratio 1 mass. $\qquad\square$

Lemma 6.2.29. *Let k_n be a sequence such that $k_n \to \infty$. Suppose V has a density f such that $f(v) = v^{a-1}(1-v)^{b-1}g(v)$ with g twice continuously differentiable and bounded away from zero, $a, b > 0$. Then for all σ we will have*

$$\frac{\displaystyle\sum_{\substack{j_0,\cdots,j_{m-1}\leq k_n, j_m\geq 0 \\ \text{summing to } \lfloor \lambda n \rfloor}} \prod_{k=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(k)}}(1-V)^{j_{\sigma(k)}+T_{\sigma,j,k+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,j,k}}\right]} \prod_{k=0}^{m}\prod_{r=0}^{j_{\sigma(k)}} \frac{\tilde{\mathbb{E}}\left[V(1-V)^{r+T_{\sigma,j,k+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{r+1+T_{\sigma,j,k+1}}\right]}}{\displaystyle\sum_{\substack{j_0,\cdots,j_m \\ \text{summing to } \lfloor \lambda n \rfloor}} \prod_{k=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(k)}}(1-V)^{j_{\sigma(k)}+T_{\sigma,j,k+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,j,k}}\right]} \prod_{k=0}^{m}\prod_{r=0}^{j_{\sigma(k)}} \frac{\tilde{\mathbb{E}}\left[V(1-V)^{r+T_{\sigma,j,k+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{r+1+T_{\sigma,j,k+1}}\right]}}$$

converges to 1.

Proof. The proof of Lemma 6.2.28 holds almost verbatim. The only difference appears when we are doing the computation for the moments, then we get extra error terms from the Stirling approximation. For any $\epsilon > 0$ we can pick the last $n\epsilon$ weights in the permutation and estimate them by one. Because we have exponential decay in n, they will contribute an error of at most ϵn times an exponential function to the total weight, so this converges to zero. Hence we can apply the Stirling approximation and estimate with $O(1/n)$ errors. This means that these errors contribute at most a constant factor to the total mass, so the total mass is still bounded and the rest of the argument holds. $\qquad\square$

Lemma 6.2.30 (Existence of good set)**.** *Suppose the true distirbution P_0 does not have both infinitely many points of positive probability and a continuous part. Suppose regularity conditions. Suppose $\hat{X}_i$ has positive probability*

under P_0, $P_0(\{\hat{X}_i\}) > 0$. *Then,* P_0 *almost surely, there exists a sequence* $(K_n)_{n \in \mathbb{N}}$ *such that* $\frac{K_n}{n} \to 0$ *and sets*

$$T_n = \{\sigma \in S_{M_n} | \sigma(i) \leq K_n\}$$

such that

$$\tilde{\mathbb{P}}(T_n) \to 1.$$

Proof. If the true distribution has infinitely many points of positive probability and no continuous part, by lemma Lemma 6.2.10 we know that the number of distinct observations K_n is such that $\frac{K_n}{n}$. Hence we can pick A_k to be the set of all permutations of the observations. Otherwise, suppose P_0 has d points of positive probability. If P_0 has no continuous part, we can again pick all the permutations of the distinct observations. If P_0 has a continuous part as well, we apply Lemma 6.2.29 so that we find such a K_n and sets of permutations. If P_0 has no points of positive probability this lemma is vacuously true.   □

6.2.7   Analysis of the asymptotic behavior of the moments of weights corresponding to observations from discrete part

Our proof of consistency uses that the mean and the variance of the weights corresponding to observations from the discrete part behave in the right way. Namely, what we want is that the weight of a observation $\hat{X}_i$ converges almost surely to the true probability of observing $\hat{X}_i$, and the variance of this weight converges almost surely to zero. This is the topic of this subsection.

We first show we converge uniformly to the right answer, which means that upper and lower bounds converge to the right answer.

Lemma 6.2.31. *Let* K_n *be a sequence such that* $\frac{K_n}{n} \to 0$. *Suppose that* $\lim_{n \to \infty} p_{l,n} = p_l$. *Let* $k > 0$. *Let* A_n *be a sequence of permutations such that* $\sigma(l) \leq K_n$ *for all* $\sigma \in A_n$. *Assuming regularity conditions then uniformly in all* $\sigma \in A_n$:

$$\frac{\frac{\tilde{\mathbb{E}}\left[V^{np_k+q}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}} \to p_l^k$$

Proof. We begin by canceling terms which appear both in the numerator and denominator.

$$\frac{\frac{\tilde{\mathbb{E}}\left[V^{np_k+q}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]}\prod_{l<\sigma(k)}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]}\prod_{l>\sigma(k)}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod_{l=1}^{m}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$=\frac{\frac{\tilde{\mathbb{E}}\left[V^{np_k+q}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]}\prod_{l<\sigma(k)}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]}}{\prod_{l=1}^{\sigma(k)}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

By Lemma 6.2.13 we can ignore the $\frac{1-\tilde{\mathbb{E}}\left[(1-V)^{x+q}\right]}{1-\tilde{\mathbb{E}}[(1-V)^x]}$ terms for computing the limit uniformly, i.e.

$$\frac{\frac{\tilde{\mathbb{E}}\left[V^{np_k+q}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]}\prod_{l<\sigma(k)}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]}}{\prod_{l=1}^{\sigma(k)}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$=\frac{\tilde{\mathbb{E}}\left[V^{np_k+q}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]\prod_{l<\sigma(k)}\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{\prod_{l=1}^{\sigma(k)}\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}(1+R_{\sigma,n})$$

where

$$\sup_{\sigma\in A_n}|R_{\sigma,n}|\to 0.$$

We now apply Lemma A.2 to every of these terms in the product. Thus for every term in the product we write

$$\frac{\tilde{\mathbb{E}}\left[V^x(1-V)^{y+k}\right]}{\tilde{\mathbb{E}}\left[V^x(1-V)^y\right]}=(1+r_{x,y,k})\prod_{i=0}^{k-1}\frac{y+b+i}{x+y+a+b-1+i}.$$

So we get

$$\frac{\tilde{\mathbb{E}}\left[V^{np_k+q}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]\prod_{l<\sigma(k)}\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{\prod_{l=1}^{\sigma(k)}\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}=$$

$$\prod_{i=0}^{q-1}\left(\frac{np_{k,n}+a+i}{\sum_{j:\sigma(j)\geq\sigma(k)}np_{j,n}+a+b-1+i}\;\prod_{l=1}^{\sigma(k)-1}\frac{\sum_{j:\sigma(j)\geq l+1}np_{j,n}+b+i}{\sum_{j:\sigma(j)\geq l}np_{j,n}+a+b-1+i}\right)$$

$$\prod_{l=1}^{\sigma(k)}(1+r_l)$$

Where we also know that $\prod_{l=1}^{\sigma(k)}(1+r_l)\to 1$ uniformly, thus, for $1+R'_{\sigma,n}=\prod_{l=1}^{\sigma(k)}(1+r_l)$ we know that $\sup_{\sigma\in A_n}|R'_{\sigma,n}|\to 0$.

We finally apply Lemma 6.2.12 to every of these factors conclude that

$$\prod_{i=0}^{q-1}\left(\frac{np_{k,n}+a+i}{\sum_{j:\sigma(j)\geq\sigma(k)}np_{j,n}+a+b-1+i}\;\prod_{l=1}^{\sigma(k)-1}\frac{\sum_{j:\sigma(j)\geq l+1}np_{j,n}+b+i}{\sum_{j:\sigma(j)\geq l}np_{j,n}+a+b-1+i}\right)$$

$$=p_l^k(1+R''_{\sigma,n})$$

where

$$\sup_{\sigma\in A_n}|R''_{\sigma,n}|\to 0.$$

So that

$$\frac{\frac{\tilde{\mathbb{E}}\left[V^{np_k+q}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]}\prod_{l<\sigma(k)}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]}\prod_{l>\sigma(k)}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod_{l=1}^{m}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}=p_l^k(1+R'''_{\sigma,n})$$

with

$$\sup_{\sigma\in A_n}|R'''_{\sigma,n}|\to 0.$$

Which shows that indeed uniformly in all σ we have that

$$\frac{\frac{\tilde{\mathbb{E}}\left[V^{np_k+q}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]}\prod_{l<\sigma(k)}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]}\prod_{l>\sigma(k)}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod_{l=1}^{m}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}\to p_l^k.$$

$\square$

This means that we have a good set, uniform convergence on the good set and bounded behavior, so we can compute the limit of the ratios which appear.

Lemma 6.2.32. *Suppose that the true distribution does not have both infinitely many points of positive probability and a continuous part. Suppose $P_0(\hat{X}_k) = p_k$. Then*

$$\mathbb{E}\left[W_{I_k}^q | X_1, \cdots, X_n\right] \to p_k^q \qquad P_0 \ a.s.$$

in the model where $X_1, \cdots, X_n | P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0, 1]$ which admits a Lebesgue density h and $a, b > 0$ and g a twice continuously differentiable function on $[0, 1]$ which is bounded away from zero such that $h(v) = v^{a-1}(1 - v)^{b-1}g(v)$ for all $v \in (0, 1)$.

Proof. By Lemma 6.2.22

$$\tilde{\mathbb{E}}\left[W_{I_k}^q\right] =$$

$$\frac{\displaystyle\sum_{\sigma \in S^m} \frac{\tilde{\mathbb{E}}\left[V^{np_k+q}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\displaystyle\sum_{\sigma \in S^m} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}.$$

By Lemma 6.2.26 the pointwise errors remain bounded.

By the strong law of large numbers the sample frequency $p_{k,n} = \frac{N_{k,n}}{n}$ converges P_0 almost surely to p_k. By Lemma 6.2.30 we can P_0 almost surely find a sequence K_n and sets T_n such that

$$\frac{\displaystyle\sum_{\sigma \in T_m} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\displaystyle\sum_{\sigma \in S^m} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}} \to 1.$$

So we restrict to the probability one set of $X_1, X_2, \cdots$ such that we indeed have that $p_{k_n} \to p_k$ and that there exists a sequence K_n and sets T_n as stated above. By Lemma 6.2.31 applied with such observations $X_1, X_2, \cdots$, a

sequence K_n and sets A_n we have uniform convergence for every permutation in A_n. This means we can apply Lemma 6.2.9 we conclude that

$$\tilde{\mathbb{E}}\left[W_{I_l}^k|X_1,\cdots,X_n\right] \to p_l^k \qquad P_0 \text{ a.s.}$$

$\square$

Now we know the expressions for the moments, we can just collect the results on the mean and the variance.

Lemma 6.2.33. *Suppose that the true distribution does not have both infinitely many points with positive probability and a continuous part. Suppose $p_k = P_0(\{\hat{X}_k\})$. Then*

$$\mathbb{E}\left[W_{I_k}|X_1,\cdots,X_n\right] \to p_k \qquad P_0 \text{ a.s.}$$

$$Var(W_{I_k}|X_1,\cdots,X_n) \to 0 \, P_0 \text{ a.s.}$$

in the model where $X_1,\cdots,X_n|P \sim P$ and $P \sim SBP(\alpha,\mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0,1]$ which admits a Lebesgue density h and $a,b > 0$ and g a twice continuously differentiable function on $[0,1]$ which is bounded away from zero such that $h(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$.

Proof. By Lemma 6.2.32 we know that

$$\mathbb{E}\left[W_{I_k}|X_1,\cdots,X_n\right] \to p_k$$

$$\mathbb{E}\left[W_{I_k}^2|X_1,\cdots,X_n\right] \to p_k^2$$

Hence

$$\begin{aligned}
Var(W_{I_k}|X_1,\cdots,X_n) &= \mathbb{E}\left[W_{I_k}^2|X_1,\cdots,X_n\right] - \mathbb{E}\left[W_{I_k}|X_1,\cdots,X_n\right]^2 \\
&\to p_k^2 - p_k^2 \\
&= 0.
\end{aligned}$$

$\square$

6.2.8 Analysis of the asymptotic behavior of first and second moment of weight of observations from continuous part

First we need some extra techniques to study the behavior. If you have these results, we can then use them to prove the results we want.

The expressions for the moments

The first step is getting expressions for the conditional moment of the sum of the weights corresponding to observations in the continuous part.

We start by computing the sum of the q-th moments corresponding to observations from the continuous part.

Lemma 6.2.34. *Suppose P_0 is the true distribution. Denote $C = \{i : \mathbb{P}(\{\hat{X}_i\}) = 0\}$.*

$$\tilde{\mathbb{E}}\left[\sum_{k \in C_n} W_{I_k}^q\right] =$$

$$\frac{\sum_{\sigma \in S^m} \sum_{k \in C_n} \frac{\tilde{\mathbb{E}}\left[V^{np_k, n+q}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\sum_{\sigma \in S^m} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}.$$

Proof. By linearity we can take the sum outside the expectation. Now we have an expression for the moments of W_{I_k} by Lemma 6.2.22. If we apply this we get

$$\tilde{\mathbb{E}}\left[\sum_{k \in C_n} W_{I_k}^q\right] =$$

$$\sum_{k \in C_n} \frac{\sum_{\sigma \in S^m} \frac{\tilde{\mathbb{E}}\left[V^{np_k, n+q}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\sum_{\sigma \in S^m} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}.$$

Note that every denominator in this sum is the same, so we can move the sum over all the variables inside. If we then move it inside the summation over all the permutations as well we get

$$\tilde{\mathbb{E}}\left[\sum_{k \in C_n} W_{I_k}^q\right] =$$

$$\frac{\sum_{\sigma \in S^m} \sum_{k \in C_n} \frac{\tilde{\mathbb{E}}\left[V^{np_k, n+q}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\sum_{\sigma \in S^m} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}.$$

$\square$

We can also find an expression for the first moment of products of weights corresponding to observations of the continuous part.

Lemma 6.2.35. *Let P_0 be a distribution. We define $C_n = \{i : P_0(\{\hat{X}_i\}) = 0, N_{i,n} = 1\}$. Suppose there are K_n distinct observations. Then*

$$\tilde{\mathbb{E}}\left[\sum_{i,j\in C, i\neq j} W_{I_k} W_{I_m}\right] =$$

$$\frac{\sum_{\sigma\in S^m} \sum_{s,t:\sigma^{-1}(s),\sigma^{-1}(t)\in C_n, s<t} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]} \prod_{l=1}^{t-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\sum_{\sigma\in S^{K_n}} \prod_{l=1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\frac{\prod_{l=t+1}^{s-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]} \prod_{l=s+1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}$$

in the model where $X_1, \cdots, X_n | P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0,1]$ which admits a Lebesgue density h and $a, b > 0$ and g a twice continuously differentiable function on $[0,1]$ which is bounded away from zero such that $h(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$.

Proof. From Lemma 6.2.24 We know that

$$\tilde{\mathbb{E}}\left[W_{I_k} W_{I_m}\right] =$$

$$2\frac{\sum_{\sigma\in S^m} \frac{\tilde{\mathbb{E}}\left[V^{np_s+1}(1-V)^{T_{\sigma,\sigma(s)}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,\sigma(s)}}\right]} \frac{\tilde{\mathbb{E}}\left[V^{np_t+1}(1-V)^{T_{\sigma,\sigma(t)}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,\sigma(t)}}\right]} \prod_{l=1}^{\sigma(t)-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\sum_{\sigma\in S^{K_n}} \prod_{l=1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\frac{\prod_{l=\sigma(t)+1}^{\sigma(s)-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]} \prod_{l=\sigma(s)+1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}$$

Now we use linearity twice to compute this for the sum. This gives

$$\tilde{\mathbb{E}}\left[\sum_{i,j\in C_n, i\neq j} W_{I_k} W_{I_m}\right] =$$

$$\frac{\sum_{\sigma\in S^m} \Sigma_{i,j\in C_n, i\neq j} \frac{\tilde{\mathbb{E}}\left[V^{np_s+1}(1-V)^{T_{\sigma,\sigma(s)}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,\sigma(s)}}\right]} \frac{\tilde{\mathbb{E}}\left[V^{np_t+1}(1-V)^{T_{\sigma,\sigma(t)}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,\sigma(t)}}\right]} \prod_{l=1}^{\sigma(t)-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\sum_{\sigma\in S^{K_n}} \prod_{l=1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\frac{\prod_{l=\sigma(t)+1}^{\sigma(s)-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]} \prod_{l=\sigma(s)+1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}$$

Because we are only summing observations from the continuous part which appeared once, we can simplify this a lot further using that the distribution of W_{I_i} is the same for all $i \in C_n$ Corollary 6.2.25. So we are only interested in the which term of the permutation we are summing, so we can also take a look at this sum, which means we can actually sum over the images of the permutation, i.e. those s, t such that $\sigma^{-1}(s)$ and $\sigma^{-1}(t)$ lie in C. Because we then know what the order is, we can restrict to the case where $t < s$ and double these values. This gives

$$\tilde{\mathbb{E}}\left[\sum_{i,j \in C_n, i \neq j} W_{I_k} W_{I_m}\right] =$$

$$2 \frac{\displaystyle\sum_{\sigma \in S^m} \sum_{s,t:\sigma^{-1}(s),\sigma^{-1}(t) \in C_n, s<t} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]} \prod_{l=1}^{t-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\displaystyle\sum_{\sigma \in S^{K_n}} \prod_{l=1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\frac{\displaystyle\prod_{l=t+1}^{s-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]} \prod_{l=s+1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}$$

$\square$

Another thing we will be interested in are expression for the sums of streaks of continuous observations. We will formulate this in the form of getting expressions of the moments of weights in case we have only observed observations from the continuous part. The next lemmas in this subsection are only used to compute those moments. The advantage of doing the computation like this is that we get far easier to manipulate expressions where we can compute the limiting behavior of.

We now study the distribution of the $\tilde{V}_k$, and the probabilities. For this we transform to differences in order statistics, and compute the distribution of that, and then account for the overcounting, but this drops as a normalization term. We drop this factor immediately. Denote J to be the differences in order statistics of I, so $J_1 = I_{(1)}$, $J_k = I_{(k)} - I_{(k-1)}$.

Lemma 6.2.36. *The probability that the differences in order statistics are $j_1, \cdots, j_n$ is given by:*

$$\mathbb{P}(J = (j_1, \cdots, j_n)) = \prod_{j=1}^{n} \tilde{\mathbb{E}}\left[(1-V)^{n+1-j}\right]^{j_j-1} \prod_{j=1}^{n} \left(1 - \tilde{\mathbb{E}}\left[(1-V)^{n-j}\right]\right)$$

Proof. First we look at the distribution of I, namely $\mathbb{P}(I = (i_1, \cdots, i_n))$.

This is proportional to

$$\tilde{\mathbb{E}}\left[\prod_{k=1}^{m}V_{i_k}^{n_k}\prod_{j=1}^{i_k-1}(1-V_j)^{n_k}\mathbb{1}_{\text{ all values of }i_1,\cdots,i_m\text{ are distinct}}\right].$$

We can count how many times each factor appears. The V_k with $k < j_1$ appear, n times as a factor $(1-V)$, V_{j_1} appears as a factor $V(1-V)^{n-1}$, the V_k with $J_1 < k < J_2$ appear with a factor $(1-V)^{n-1}$, V_{j_2} appears as a factor $V(1-V)^{n-2}$. If we generalize, if $J_l < k < J_{k+1}$, then V_k appears as a factor $(1-V)^{n-l}$, and V_{j_l} appear as a factor $V(1-V)^{n-l-1}$. So we can rewrite this expectation using the independence of the V_l. Namely we get

$$\mathbb{P}(J=(j_1,\cdots,j_n))\propto\prod_{l=1}^{n}\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]\prod_{l=1}^{n}\tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]^{j_l}$$

So we apply the geometric series n times to get the normalization. This gives

$$\mathbb{P}(J=(j_1,\cdots,j_n))=\prod_{j=1}^{n}\tilde{\mathbb{E}}\left[(1-V)^{n+1-j}\right]^{j_j-1}\prod_{j=1}^{n}(1-\tilde{\mathbb{E}}\left[(1-V)^{n-j}\right])$$

$$\square$$

In case we only have observations of the continuous part we want to have another expression.

Lemma 6.2.37. *Suppose every observation appears once. Then the weights of the l-th observation, ordering based on $i_1,\cdots,i_n$, are given by*

$$\left(\prod_{k=1}^{l}\frac{\tilde{\mathbb{E}}\left[(1-V)^{n+2-k}\right]}{\tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]}\right)^{j_k-1}\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n+1-l}\right]}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{n-l}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]}$$

Proof. The weights are given by

$$\tilde{\mathbb{E}}\left[\tilde{V}_{i_j}\prod_{k=1}^{i_j-1}(1-\tilde{V}_k)\right],$$

where the distribution of $\tilde{V}_k$ is given by

$$A\mapsto\frac{\int_A(1-V)^{n_k\#f}V^{n_k\#'j}\,\mathrm{d}\,\mathcal{D}(v)}{\int(1-V)^{\#f}V^{\#'j}\,\mathrm{d}\,\mathcal{D}(v)}$$

We get $I_k - 1$ times a factor of $\tilde{\mathbb{E}}\left[1 - \tilde{V}\right]$, which is

$$\frac{\tilde{\mathbb{E}}\left[(1 - V)^{n+2-j}\right]}{\tilde{\mathbb{E}}\left[(1 - V)^{n+1-j}\right]}.$$

We get for $k < j$ a factor

$$\frac{\tilde{\mathbb{E}}\left[V(1 - V)^{n-k}\right]}{\tilde{\mathbb{E}}\left[V(1 - V)^{n-k}\right]}.$$

And we get for $k = j$ a factor

$$\frac{\tilde{\mathbb{E}}\left[V^2(1 - V)^{n-j}\right]}{\tilde{\mathbb{E}}\left[V(1 - V)^{n-j}\right]}.$$

Combining all these factors gives the weight.

$$\left(\prod_{k=1}^{l} \frac{EE\left[(1 - V)^{n+2-k}\right]}{\tilde{\mathbb{E}}\left[(1 - V)^{n+1-k}\right]}\right)^{j_k - 1} \left(\prod_{k=1}^{l-1} \frac{\tilde{\mathbb{E}}\left[V(1 - V)^{n+1-k}\right]}{\tilde{\mathbb{E}}\left[V(1 - V)^{n-k}\right]}\right) \frac{\tilde{\mathbb{E}}\left[V^2(1 - V)^{n-l}\right]}{\tilde{\mathbb{E}}\left[V(1 - V)^{n-l}\right]}.$$

Now we can telescope the product

$$\prod_{k=1}^{l-1} \frac{\tilde{\mathbb{E}}\left[V(1 - V)^{n+1-k}\right]}{\tilde{\mathbb{E}}\left[V(1 - V)^{n-k}\right]}$$

into

$$\frac{\tilde{\mathbb{E}}\left[V(1 - V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1 - V)^{n+1-l}\right]}.$$

$\square$

Lemma 6.2.38. *Suppose every observations appears exactly once, then the sum of all the weights is given by*

$$\sum_{i_1, \cdots, i_n \text{ distinct}} \sum_{l=1}^{m} \tilde{\mathbb{E}}\left[\tilde{V}_{i_l} \prod_{k=1}^{i_l - 1}(1 - \tilde{V}_k)\right] \mathbb{P}(I = (i_1, \cdots, i_m)$$

$$= \sum_{j=0}^{n-1} \frac{\tilde{\mathbb{E}}\left[V(1 - V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1 - V)^{j+1}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1 - V)^j\right]}{\tilde{\mathbb{E}}\left[V(1 - V)^j\right]} \frac{1 - \tilde{\mathbb{E}}\left[(1 - V)^n\right]}{1 - \tilde{\mathbb{E}}\left[(1 - V)^{j+2}\right]}.$$

Proof. We multiply the formula for the weights and the probability together. From the previous two lemmas we know these, namely:

$$\left(\prod_{k=1}^{l} \frac{\tilde{\mathbb{E}}\left[(1-V)^{n+2-k}\right]}{\tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]}\right)^{j_k-1} \frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n+1-l}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{n-l}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]}$$

and

$$\mathbb{P}(J = (j_1, \cdots, j_n)) = \prod_{k=1}^{n} \tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]^{j_k-1} \prod_{k=1}^{n} (1 - \tilde{\mathbb{E}}\left[(1-V)^{n-k}\right]).$$

If we multiply these two we get

$$\left(\prod_{k=1}^{l} \tilde{\mathbb{E}}\left[(1-V)^{n+2-k}\right]\right)^{j_k-1} \frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n+1-l}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{n-l}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]}$$

$$\prod_{k=l+1}^{n} \tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]^{j_k} \prod_{k=1}^{n} (1 - \tilde{\mathbb{E}}\left[(1-V)^{n-k}\right])$$

We now look at the terms for which there are j_k terms, these are:

$$\left(\prod_{k=1}^{l} \tilde{\mathbb{E}}\left[(1-V)^{n+2-k}\right]\right)^{j_k-1} \prod_{k=l+1}^{n} \tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]^{j_k}.$$

Summing these over $j_1, \cdots, j_n \in \mathbb{N}$ yields

$$\left(\prod_{k=1}^{l} \frac{1}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+2-k}\right]}\right) \prod_{k=l+1}^{n} \frac{1}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]}.$$

The rest does not change, so after summing over $j_1, \cdots, j_n$ we are left with

$$\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n+1-l}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{n-l}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]} \prod_{k=1}^{n} (1 - \tilde{\mathbb{E}}\left[(1-V)^{n-k}\right])$$

$$\left(\prod_{k=1}^{l} \frac{1}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+2-k}\right]}\right) \prod_{k=1}^{n} \frac{1}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]}.$$

Now we want to simplify

$$\prod_{k=1}^{n}(1-\tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right])\left(\prod_{k=1}^{l}\frac{1}{1-\tilde{\mathbb{E}}\left[(1-V)^{n+2-k}\right]}\right)\prod_{k=1}^{n}\frac{1}{1-\tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]}.$$

We see that a part cancels, so we are left with

$$\prod_{k=1}^{l}\frac{1-\tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{n+2-k}\right]}.$$

This telescopes to

$$\frac{1-\tilde{\mathbb{E}}\left[(1-V)^{n}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{n+2-l}\right]}.$$

So in the end we are left with

$$\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n+1-l}\right]}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{n-l}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]}\frac{1-\tilde{\mathbb{E}}\left[(1-V)^{n}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{n+2-l}\right]}.$$

This is for a fixed l, and we sum over l, thus after summing we are left with

$$\sum_{l=1}^{n}\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n+1-l}\right]}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{n-l}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]}\frac{1-\tilde{\mathbb{E}}\left[(1-V)^{n}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{n+2-l}\right]}.$$

Now we can do a change of summation, by writing $j=n-l$, which yields

$$\sum_{j=0}^{n-1}\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{j+1}\right]}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{j}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{j}\right]}\frac{1-\tilde{\mathbb{E}}\left[(1-V)^{n}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{j+2}\right]}.$$

$\square$

Next we can compute the expectation of the second moment. First observe that we can split the second moment into

$$\tilde{\mathbb{E}}\left[\sum_{i=1}^{n}W_{I_i}^2+\sum_{1\leq i<j\leq n}W_{I_i}W_{I_j}\right].$$

Now we can compute each of these, note that swapping to the differences in order statistics has no influence on this sum, because every term appears. So with this observation we can repeat the previous proofs.

Lemma 6.2.39 (Second moment of W_{J_l} given J). *Suppose every observation appears exactly once. Then the second moment of W_{J_l} conditional on J is given by*

$$\tilde{\mathbb{E}}\left[W_{J_l}^2 | J\right] =$$
$$\prod_{k=1}^{l-1} \left(\frac{\tilde{\mathbb{E}}\left[(1-V)^{n+3-k}\right]}{\tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]}\right)^{j_k-1} \frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+2}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n+1-l}\right]} \cdot \frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]} \frac{\tilde{\mathbb{E}}\left[V^3(1-V)^{n-l}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]}.$$

The proof follows practically the same style as before, just note we take second moment instead of first moment.

Proof.   • We get, for $k < l$, a factor

$$\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n-k+2}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-k}\right]}.$$

• We get, for $k = l$, a factor

$$\frac{\tilde{\mathbb{E}}\left[V^3(1-V)^{n-l}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]}.$$

• And we get, for $k \le j$, $I_k - 1$ times a factor

$$\frac{\tilde{\mathbb{E}}\left[(1-V)^{n+3-k}\right]}{\tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]}.$$

So we can telescope the products of the factors $k < l$. Namely

$$\prod_{k=1}^{l-1} \frac{\tilde{\mathbb{E}}\left[V(1-V)^{n-k+2}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-k}\right]}.$$

telescopes into
$$\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+2}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n+1-l}\right]} \cdot \frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]}.$$

So in total we get

$$\prod_{k=1}^{l-1} \left(\frac{\tilde{\mathbb{E}}\left[(1-V)^{n+3-k}\right]}{\tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]}\right)^{j_k-1} \frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+2}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n+1-l}\right]} \cdot \frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]} \frac{\tilde{\mathbb{E}}\left[V^3(1-V)^{n-l}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]}.$$

$\square$

Lemma 6.2.40 (First moment of $W_{J_i} W_{J_j}$ for $i < j$ conditional on J). *In case every observation appears exactly once, the product of $W_{I_i} W_{I_j}$ for $i < j$ has, conditional on J, expectation*

$$\tilde{\mathbb{E}}\left[W_{J_i} W_{J_j} | J\right] =$$

$$= \prod_{k=1}^{i} \left(\frac{\tilde{\mathbb{E}}\left[(1-V)^{n+3-k}\right]}{\tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]} \right)^{J_k - 1} \prod_{k=i+1}^{j} \left(\frac{\tilde{\mathbb{E}}\left[(1-V)^{n+2-k}\right]}{\tilde{\mathbb{E}}\left[(1-V)^{n-k}\right]} \right)^{J_k - 1}$$

$$\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+2}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-i+1}\right]} \frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-i}\right]} \frac{\tilde{\mathbb{E}}\left[V(1-V)^{n-i+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-j}\right]}$$

$$\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{n-i+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-i+1}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{n-j+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-j+1}\right]}.$$

Proof. Here one has to distinguish a bit more cases, but the proof is essentially the same as the previous two computations of the moments combined.

- We get, for $k < i$, a factor

$$\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n-k+2}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-k}\right]}.$$

- We get, for $k = i$, a factor

$$\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{n-i}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-i}\right]}.$$

- We get, for $i < k < j$, a factor

$$\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n-k+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-k}\right]}.$$

- We get, for $k = j$, a factor

$$\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{n-j}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-j}\right]}$$

- We get, for $k \le i$, $J_k - 1$ times a factor

$$\frac{\tilde{\mathbb{E}}\left[(1-V)^{n+3-j}\right]}{\tilde{\mathbb{E}}\left[(1-V)^{n+1-j}\right]}.$$

- And finally we get, for $i < k \le j$, $J_k - 1$ times a factor

$$\frac{\tilde{\mathbb{E}}\left[(1-V)^{n+2-j}\right]}{\tilde{\mathbb{E}}\left[(1-V)^{n+1-j}\right]}.$$

We can again telescope the factors smaller than i, and the factors between i and j. This leads to

$$\tilde{\mathbb{E}}\left[W_{J_i}W_{J_j}|J\right] =$$

$$= \prod_{k=1}^{i}\left(\frac{\tilde{\mathbb{E}}\left[(1-V)^{n+3-j}\right]}{\tilde{\mathbb{E}}\left[(1-V)^{n+1-j}\right]}\right)^{J_k-1} \prod_{k=i+1}^{j}\left(\frac{\tilde{\mathbb{E}}\left[(1-V)^{n+2-j}\right]}{\tilde{\mathbb{E}}\left[(1-V)^{n-j}\right]}\right)^{J_k-1}$$

$$\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+2}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-i+1}\right]}\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-i}\right]}\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n-i+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-j}\right]}$$

$$\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{n-i}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-i}\right]}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{n-j}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-j}\right]}.$$

$\square$

Lemma 6.2.41 (Second moment of W_{J_k})**.** *Suppose every observation occurs exactly once, then*

$$\tilde{\mathbb{E}}\left[\sum_{j=1}^{n}W_{J_j}^2\right] = \sum_{j=1}^{n}\frac{1-\tilde{\mathbb{E}}\left[(1-V)^n\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{n+3-l}\right]}\frac{1-\tilde{\mathbb{E}}\left[(1-V)^{n-1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{n+2-l}\right]}$$

$$\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+2}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n+1-l}\right]}\cdot\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]}\frac{\tilde{\mathbb{E}}\left[V^3(1-V)^{n-l}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]}.$$

Proof. We know an expression for the second moment of W_{J_k} given J. We now multiply this by the probability of J, and then sum over all the possible assignments of J. Recall that the distribution of J is given by

$$\mathbb{P}(J = (j_1, \cdots, j_n)) = \prod_{k=1}^{n}\tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]^{j_k-1}\prod_{k=1}^{n}(1-\tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]).$$

This combined with the lemma on the conditional second moment of W_{J_k}, which states

$$\tilde{\mathbb{E}}\left[W_{J_l}^2|J\right] =$$

$$\prod_{k=1}^{l-1}\left(\frac{\tilde{\mathbb{E}}\left[(1-V)^{n+3-k}\right]}{\tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]}\right)^{j_k-1}\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+2}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n+1-l}\right]}\cdot\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]}\frac{\tilde{\mathbb{E}}\left[V^3(1-V)^{n-l}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]}.$$

allows us to compute the expected value. We first only restrict ourselves to the terms with an j_k in it. This is the same as before for the first moment, take the product, sum everything, get the $\frac{1}{1-..}$ factors. Now observe that a part cancels and we are left with

$$\prod_{k=1}^{l} \frac{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+3-k}\right]}.$$

Now this telescopes to

$$\frac{1 - \tilde{\mathbb{E}}\left[(1-V)^{n}\right]}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+3-l}\right]} \frac{1 - \tilde{\mathbb{E}}\left[(1-V)^{n-1}\right]}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+2-l}\right]}.$$

The rest does not change with j_k, so we can sum over this. This yields

$$\frac{1 - \tilde{\mathbb{E}}\left[(1-V)^{n}\right]}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+3-l}\right]} \frac{1 - \tilde{\mathbb{E}}\left[(1-V)^{n-1}\right]}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+2-l}\right]} \frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+2}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n+1-l}\right]}$$
$$\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]} \frac{\tilde{\mathbb{E}}\left[V^{3}(1-V)^{n-l}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]}.$$

Summing over all the j leads to

$$\tilde{\mathbb{E}}\left[\sum_{j=1}^{n} W_{J_j}^{2}\right] = \sum_{j=1}^{n} \frac{1 - \tilde{\mathbb{E}}\left[(1-V)^{n}\right]}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+3-l}\right]} \frac{1 - \tilde{\mathbb{E}}\left[(1-V)^{n-1}\right]}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+2-l}\right]}$$
$$\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+2}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n+1-l}\right]} \frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]} \frac{\tilde{\mathbb{E}}\left[V^{3}(1-V)^{n-l}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]}.$$

$\square$

Lemma 6.2.42 (First moment of $W_{I_i} W_{I_j}$)**.** *Suppose every observation ap-*

pears exactly once, then

$$\tilde{\mathbb{E}}\left[\sum_{1\le i<j\le n} W_{J_i}W_{J_j}\right]$$

$$= \sum_{1\le i<j\le n} \frac{1 - \tilde{\mathbb{E}}\left[(1-V)^n\right]}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+3-l}\right]} \frac{1 - \tilde{\mathbb{E}}\left[(1-V)^{n-1}\right]}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+2-l}\right]} \frac{1 - \tilde{\mathbb{E}}\left[(1-V)^{n-i}\right]}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+2-j}\right]}$$

$$\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+2}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-i+1}\right]} \frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-i}\right]} \frac{\tilde{\mathbb{E}}\left[V(1-V)^{n-i+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-j}\right]}$$

$$\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{n-i}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-i}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{n-j}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-j}\right]}.$$

Proof. We know an expression for the first moment of $W_{J_i}W_{J_j}$ given J. We now multiply this by the probability of J, and then sum over all the possible assignments of J. Recall that the distribution of J is given by

$$\mathbb{P}(J = (j_1, \cdots, j_n)) = \prod_{k=1}^{n} \tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]^{j_k-1} \prod_{k=1}^{n} (1 - \tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]).$$

This combined with the lemma on the conditional first moment of $W_{J_i}W_{J_j}$, which states

$$\tilde{\mathbb{E}}\left[W_{J_i}W_{J_j}|J\right] =$$

$$= \prod_{k=1}^{i}\left(\frac{\tilde{\mathbb{E}}\left[(1-V)^{n+3-k}\right]}{\tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]}\right)^{J_k-1} \prod_{k=i+1}^{j}\left(\frac{\tilde{\mathbb{E}}\left[(1-V)^{n+2-k}\right]}{\tilde{\mathbb{E}}\left[(1-V)^{n-k}\right]}\right)^{J_k-1}$$

$$\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+2}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-i+1}\right]} \frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-i}\right]} \frac{\tilde{\mathbb{E}}\left[V(1-V)^{n-i+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-j}\right]}$$

$$\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{n-i+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-1}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{n-i}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-i}\right]}.$$

So we again take the product, sum everything, get the $\frac{1}{1-..}$ factors. We again only look at the terms where there was a J_k term, because the rest does not change. We can compute these, and then look what resulting factor we get. First observe that a part cancels and we are left with

$$\prod_{k=1}^{i}\frac{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+3-k}\right]} \cdot \prod_{k=i+1}^{j}\frac{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+1-k}\right]}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+2-k}\right]}.$$

Now we do the telescoping per product, the first yields a telescoping product of two terms, the second of one term. This yields

$$\frac{1 - \tilde{\mathbb{E}}\left[(1-V)^n\right]}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+3-l}\right]} \frac{1 - \tilde{\mathbb{E}}\left[(1-V)^{n-1}\right]}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+2-l}\right]} \frac{1 - \tilde{\mathbb{E}}\left[(1-V)^{n-i}\right]}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+2-j}\right]}.$$

Combining this with the previous expression and summing over all the i, j yields

$$\tilde{\mathbb{E}}\left[\sum_{1 \leq i < j \leq n} W_{J_i} W_{J_j}\right]$$

$$= \sum_{1 \leq i < j \leq n} \frac{1 - \tilde{\mathbb{E}}\left[(1-V)^n\right]}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+3-i}\right]} \frac{1 - \tilde{\mathbb{E}}\left[(1-V)^{n-1}\right]}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+2-i}\right]} \frac{1 - \tilde{\mathbb{E}}\left[(1-V)^{n-i}\right]}{1 - \tilde{\mathbb{E}}\left[(1-V)^{n+2-j}\right]}$$

$$\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+2}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-i+1}\right]} \frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-i}\right]} \frac{\tilde{\mathbb{E}}\left[V(1-V)^{n-i+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-j}\right]}$$

$$\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{n-i+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-1}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{n-i}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-i}\right]}.$$

$\square$

Development of tools

We want to better understand the situation where all the observations have occurred exactly once. It will turn out that later we can reduce to this case. The following lemmas are the main tools we need to compute the limiting behavior, in case every observation is unique. For the general case we will need more techniques, which are then used to return to this computation.

This lemma is used to control the terms appearing when we analyze the first moment.

Lemma 6.2.43 (First moment converges to 1.)**.** *Suppose $V \sim \mathcal{D}$ where $\mathcal{D}$ admits a density f such that there exists constants $a, b > 0$ and a twice continuously differentiable function g bounded away from zero and $f(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$. Then*

$$\sum_{j=0}^{n-1} \frac{\mathbb{E}\left[V(1-V)^{n+1}\right]}{\mathbb{E}\left[V(1-V)^{j+1}\right]} \frac{\mathbb{E}\left[V^2(1-V)^j\right]}{\mathbb{E}\left[V(1-V)^j\right]} \frac{1 - \mathbb{E}\left[(1-V)^n\right]}{1 - \mathbb{E}\left[(1-V)^{j+2}\right]} \to 1.$$

Proof. In view of Lemma A.4 we can estimate

$$K_j = \frac{\mathbb{E}\left[V(1-V)^{n+1}\right]}{\mathbb{E}\left[V(1-V)^{j+1}\right]} \frac{\mathbb{E}\left[V^2(1-V)^j\right]}{\mathbb{E}\left[V(1-V)^j\right]} \frac{1-\mathbb{E}\left[(1-V)^n\right]}{1-\mathbb{E}\left[(1-V)^{j+2}\right]}.$$

By

$$K_j \leq \frac{C}{c}(\frac{j+1}{n+1})^{1+\alpha}\frac{1}{1-\mathbb{E}\left[(1-V)^3\right]}.$$

So then

$$\sum_{j=1}^{J} K_j \leq \mathrm{Const}(\frac{j+1}{n+1})^{1+\alpha}$$

If $\frac{J}{n^{\frac{1+\alpha}{2+\alpha}}} \to 0$ this sum converges to zero. This means we can ignore the first J terms and apply Lemma A.5. Thus for all $\epsilon > 0$, there exists an $N > 0$ such that foor all $n > N$, the error of estimating the $\mathbb{E}\left[V^s(1-V)^j\right]$ for $j > J_n$ is

$$(1-\epsilon)g(0)\frac{\Gamma(s+\alpha)}{j^{s+\alpha}} \leq \mathbb{E}\left[V^s(1-V)^j\right] \leq (1+\epsilon)g(0)\frac{\Gamma(s+\alpha)}{j^{s+\alpha}}$$

And further

$$1-\epsilon \leq \frac{1}{1-\mathbb{E}\left[(1-V)^{j+2}\right]} \leq 1+\epsilon$$

If we apply this we get

$$(\frac{1+\epsilon}{1-\epsilon})^3(\frac{j+1}{n+1})^{1+\alpha}\frac{1+\alpha}{j} \leq K_j \leq (\frac{1-\epsilon}{1+\epsilon})^3(\frac{j+1}{n+1})^{1+\alpha}\frac{1+\alpha}{j}.$$

If we now compute the sum $\sum_{j=J}^{n} K_j$, we can estimate this by

$$\sum_{j=J}^{n}(\frac{j+1}{n+1})^{1+\alpha}\frac{1+\alpha}{j}$$

where the errors are given above. If we compute this sum, we get something of the order

$$\frac{(n+1)^{1+\alpha}-(J+1)^{1+\alpha}}{(n+1)^{1+\alpha}} \sim 1.$$

So combining all this yields

$$\sum_{i=1}^{n} W_{I_i} \to 1.$$

$\square$

Next, we start working on the convergence of the second moment. For this will use two lemmas controlling the convergence in two different cases, and then approximate.

This lemma is used to control the terms appearing when we analyze the sum of second moments.

Lemma 6.2.44 (Sum of second moments has zero contribution). *Suppose $V \sim \mathcal{D}$ where $\mathcal{D}$ admits a density f such that there exists constants $a, b > 0$ and a twice continuously differentiable function g bounded away from zero and $f(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$. Then*

$$\sum_{j=1}^{n} \frac{1 - \mathbb{E}\left[(1-V)^n\right]}{1 - \mathbb{E}\left[(1-V)^{n+3-l}\right]} \frac{1 - \mathbb{E}\left[(1-V)^{n-1}\right]}{1 - \mathbb{E}\left[(1-V)^{n+2-l}\right]}$$

$$\frac{\mathbb{E}\left[V(1-V)^{n+2}\right]}{\mathbb{E}\left[V(1-V)^{n+1-l}\right]} \cdot \frac{\mathbb{E}\left[V(1-V)^{n+1}\right]}{\mathbb{E}\left[V(1-V)^{n-l}\right]} \frac{\mathbb{E}\left[V^3(1-V)^{n-l}\right]}{\mathbb{E}\left[V(1-V)^{n-l}\right]}$$

$$= \sum_{k=1}^{n} K_{n+1-j} \to 0.$$

Proof. As said before, we can again approximate using Lemma A.4 to find an upper bound. This yields

$$K_j \leq \frac{C}{c} \frac{1}{1 - \mathbb{E}\left[(1-V)^3\right]} \frac{1}{1 - \mathbb{E}\left[(1-V)^2\right]} \frac{(k+1)^{1+\alpha}}{(n+2)^{1+\alpha}} \frac{k^{1+\alpha}}{(n-1)^{1+\alpha}} \frac{1}{k^2}$$

So if we look at

$$\sum_{j=1}^{J} K_j$$

we see that this converges to zero as $n \to \infty$ and $\frac{J}{n^{\frac{1+2\alpha}{2+2\alpha}}} \to 0$. This means that we can bound the estimates as in Lemma A.5 to get upper and lower bounds. Let $\epsilon > 0$. Then there exists an $N > 0$ such that for all $n > N$ and $j > J_n$ we have

$$(1-\epsilon)^5 \frac{(k+1)^{1+\alpha}}{(n+2)^{1+\alpha}} \frac{k^{1+\alpha}}{(n-1)^{1+\alpha}} \frac{1}{k^2} \leq K_j \leq (1+\epsilon)^5 \frac{(k+1)^{1+\alpha}}{(n+2)^{1+\alpha}} \frac{k^{1+\alpha}}{(n-1)^{1+\alpha}} \frac{1}{k^2}$$

So if we sum the K_j it is enough to compute what happens when we sum

$$\sum_{j=J}^{n} \frac{(k+1)^{1+\alpha}}{(n+2)^{1+\alpha}} \frac{k^{1+\alpha}}{(n-1)^{1+\alpha}} \frac{1}{k^2}$$

Observe that this behaves as

$$\sum_{j=1}^{n} \frac{j^{2+2\alpha}}{n^{2+2\alpha}} \frac{1}{j^2} \sim \frac{n^{1+2\alpha} - J^{1+2\alpha}}{n^{2+2\alpha}} \sim \frac{1}{n}.$$

So indeed the sum of second moments, $\mathbb{E}\left[\sum_{i=1}^{n} W_{I_i}^2\right]$, converges to zero as $n \to \infty$. □

This lemma is used to control the terms appearing when we analyze the sum of mixed moments.

Lemma 6.2.45 (Sum of first moments of $W_{I_i} W_{I_j}$ converges to 1)**.** *Suppose $V \sim \mathcal{D}$ where $\mathcal{D}$ admits a density f such that there exists constants $a, b > 0$ and a twice continuously differentiable function g bounded away from zero and $f(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$. Then*

$$\sum_{1 \le i < j \le n} \frac{1 - \mathbb{E}\left[(1-V)^n\right]}{1 - \mathbb{E}\left[(1-V)^{n+3-i}\right]} \frac{1 - \mathbb{E}\left[(1-V)^{n-1}\right]}{1 - \mathbb{E}\left[(1-V)^{n+2-i}\right]} \frac{1 - \mathbb{E}\left[(1-V)^{n-i}\right]}{1 - \mathbb{E}\left[(1-V)^{n+2-j}\right]}$$

$$\frac{\mathbb{E}\left[V(1-V)^{n+2}\right]}{\mathbb{E}\left[V(1-V)^{n-i+1}\right]} \frac{\mathbb{E}\left[V(1-V)^{n+1}\right]}{\mathbb{E}\left[V(1-V)^{n-i}\right]} \frac{\mathbb{E}\left[V(1-V)^{n-i+1}\right]}{\mathbb{E}\left[V(1-V)^{n-j}\right]}$$

$$\frac{\mathbb{E}\left[V^2(1-V)^{n-i+1}\right]}{\mathbb{E}\left[V(1-V)^{n-1}\right]} \frac{\mathbb{E}\left[V^2(1-V)^{n-i}\right]}{\mathbb{E}\left[V(1-V)^{n-i}\right]} \to 1.$$

Proof. We can first bound this by observing that the factors

$$\frac{1 - \mathbb{E}\left[(1-V)^n\right]}{1 - \mathbb{E}\left[(1-V)^{n+3-i}\right]} \frac{1 - \mathbb{E}\left[(1-V)^{n-1}\right]}{1 - \mathbb{E}\left[(1-V)^{n+2-i}\right]} \frac{1 - \mathbb{E}\left[(1-V)^{n-i}\right]}{1 - \mathbb{E}\left[(1-V)^{n+2-j}\right]}$$

are always between

$$\frac{1 - \mathbb{E}\left[(1-V)^n\right]}{1 - \mathbb{E}\left[(1-V)^{n+2}\right]} \frac{1 - \mathbb{E}\left[(1-V)^{n-1}\right]}{1 - \mathbb{E}\left[(1-V)^{n+1}\right]} \frac{1 - \mathbb{E}\left[(1-V)^1\right]}{1 - \mathbb{E}\left[(1-V)^n\right]}$$

and

$$\frac{1}{1 - \mathbb{E}\left[(1-V)^4\right]} \frac{1}{1 - \mathbb{E}\left[(1-V)^3\right]} \frac{1}{1 - \mathbb{E}\left[(1-V)^2\right]}.$$

Because the lower bound is convergent to 1 and always positive, we can find a constant const > 0 such this constant is a lower bound for all n. This

allows us to ignore these factors for now. Now look at

$$\sum_{1 \leq i < j \leq n} \frac{\mathbb{E}\left[V(1-V)^{n+2}\right]}{\mathbb{E}\left[V(1-V)^{n-i+1}\right]} \frac{\mathbb{E}\left[V(1-V)^{n+1}\right]}{\mathbb{E}\left[V(1-V)^{n-i}\right]} \frac{\mathbb{E}\left[V(1-V)^{n-i+1}\right]}{\mathbb{E}\left[V(1-V)^{n-j}\right]}$$
$$\frac{\mathbb{E}\left[V^2(1-V)^{n-i+1}\right]}{\mathbb{E}\left[V(1-V)^{n-1}\right]} \frac{\mathbb{E}\left[V^2(1-V)^{n-i}\right]}{\mathbb{E}\left[V(1-V)^{n-i}\right]}.$$

We first get an upper bound by using Lemma A.4. This yields

$$\frac{C}{c} \sum_{1 \leq i < j \leq n} \frac{(n+1-i)^{1+\alpha}}{(n+2)^{1+\alpha}} \frac{(n-i)^{1+\alpha}}{(n+1)^{1+\alpha}} \frac{(n-j)^{1+\alpha}}{(n-i+1)^{1+\alpha}} \frac{(n-i+1)^{1+\alpha}}{(n-i+1)^{2+\alpha}} \frac{(n-j)^{1+\alpha}}{(n-j)^{2+\alpha}}.$$

If we simplify this we get

$$2\frac{C}{c} \sum_{1 \leq i < j \leq n} \frac{(n-j)^{1+\alpha}}{(n+2)^{1+\alpha}} \frac{(n-i)^{1+\alpha}}{(n+1)^{1+\alpha}} \frac{1}{n-i+1} \frac{1}{n-j}.$$

Note that

$$2 \sum_{1 \leq i < j \leq n} \frac{(n-j)^{1+\alpha}}{(n+2)^{1+\alpha}} \frac{(n-i)^{1+\alpha}}{(n+1)^{1+\alpha}} \frac{1}{n-i+1} \frac{1}{n-j}$$
$$\sim (\sum_{k=1}^{n} \frac{k^\alpha}{n^{1+\alpha}})^2 - \sum_{k=1}^{n} \frac{k^{2\alpha}}{n^{2+2\alpha}}$$

where we see that $\sum_{k=1}^{n} \frac{k^{2\alpha}}{n^{2+2\alpha}} \to 0$ as $n \to \infty$. The rest of this sum converges. So if we now study

$$\sum_{k=1}^{J} \frac{k^\alpha}{n^{1+\alpha}} \sim \frac{1}{1+\alpha} \frac{J^{1+\alpha}}{n^{1+\alpha}}$$

we see that picking J such that $J \to \infty$ and $\frac{J}{n} \to 0$ this part of the sum converges to zero as $n \to \infty$. Looking back at what i and j correspond to $k \leq J$ we see that this means $n - i$ and $n - j$ should be less than J. By Lemma A.5 we can now approximate all the expectations uniformly. This

yields upper and lower bounds of the form

$$
2(\frac{1-\epsilon}{1+\epsilon})^8(\frac{1}{1+\alpha})^2
$$

$$
\sum_{1\leq i<j\leq n} \frac{(n+1-i)^{1+\alpha}}{(n+2)^{1+\alpha}} \frac{(n-i)^{1+\alpha}}{(n+1)^{1+\alpha}} \frac{(n-j)^{1+\alpha}}{(n-i+1)^{1+\alpha}} \frac{(n-i+1)^{1+\alpha}}{(n-i+1)^{2+\alpha}} \frac{(n-j)^{1+\alpha}}{(n-j)^{2+\alpha}}
$$

$$
\leq 2\mathbb{E}\left[\sum_{1\leq i<j\leq n} W_{J_i}W_{J_j}\right] \leq
$$

$$
2(\frac{1+\epsilon}{1-\epsilon})^8(\frac{1}{1+\alpha})^2
$$

$$
\sum_{1\leq i<j\leq n} \frac{(n+1-i)^{1+\alpha}}{(n+2)^{1+\alpha}} \frac{(n-i)^{1+\alpha}}{(n+1)^{1+\alpha}} \frac{(n-j)^{1+\alpha}}{(n-i+1)^{1+\alpha}} \frac{(n-i+1)^{1+\alpha}}{(n-i+1)^{2+\alpha}} \frac{(n-j)^{1+\alpha}}{(n-j)^{2+\alpha}}.
$$

So we study those two bounds. For this it is enough to study

$$
2\frac{1}{1+\alpha}^2 \frac{(n+1-i)^{1+\alpha}}{(n+2)^{1+\alpha}} \frac{(n-i)^{1+\alpha}}{(n+1)^{1+\alpha}} \frac{(n-j)^{1+\alpha}}{(n-i+1)^{1+\alpha}} \frac{(n-i+1)^{1+\alpha}}{(n-i+1)^{2+\alpha}} \frac{(n-j)^{1+\alpha}}{(n-j)^{2+\alpha}}.
$$

If we use the same analysis as before to simplify etc, we find that this is again asymptotic to

$$
(\frac{1}{1+\alpha}\sum_{k=1}^{n}\frac{k^\alpha}{n^{1+\alpha}})^2.
$$

And we know the term inside the square converges to one. This shows

$$
2\mathbb{E}\left[\sum_{1\leq i<j\leq n} W_{J_i}W_{J_j}\right] \to 1.
$$

$\square$

These three lemmas form the core of the argument.

Analysis of the moments

The analysis of the weights of the continuous part is only done in case the true distribution only has finitely many points with positive probability.

For the main tool to compute the limiting behavior of the fractions we need to identify good sets, show we converge uniformly on those sets and we stay everywhere bounded. This will now be the topic of the lemmas. We already have the good sets, so we start the boundedness results.

Lemma 6.2.46 (Errors in mean of weights of continuous part stay bounded)**.**
*The pointwise ratios in the derived expression for the sum of the mean of
the weights corresponding to observations from the continuous part remain
bounded, i.e. there exists an $L > 0$ such that for all $\sigma \in S^{K_n}$*

$$
\frac{\sum_{k \in C_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+1}(1-V)^{T_{\sigma,\sigma(k)}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}} \leq L.
$$

*P_0 almost surely in the model where $X_1, \cdots, X_n | P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$
where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0,1]$ which
admits a Lebesgue density h and $a, b > 0$ and g a twice continuously dif-
ferentiable function on $[0,1]$ which is bounded away from zero such that
$h(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$.*

Proof. We want to prove that in every fraction as derived in Lemma 6.2.34
the errors stay bounded. So by Lemma 6.2.34 we know

$$
\tilde{\mathbb{E}}\left[\sum_{k \in C_n} W_{I_k}\right] =
$$

$$
\frac{\sum_{\sigma \in S^m} \sum_{k \in C_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+1}(1-V)^{T_{\sigma,\sigma(k)}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\sum_{\sigma \in S^m} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}.
$$

So we want to show that uniformly in all $\sigma \in S^{K_n}$

$$
\frac{\sum_{k \in C_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+1}(1-V)^{T_{\sigma,\sigma(k)}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}
$$

remains bounded. Observe that it does not matter for the distribution of
the weights of the continuous part, if we replace the weights of the i-th
distinct discrete observation by N_i observations which occur only once, since
we can bound uniformly by a constant factor, by Lemma 6.2.11. If we then
increase the sum by adding all these non-negative weights. Then P_0 almost

surely, we have only distinct observations, i.e. every observations occurs
only once. This means that we are in the situation where we only have
continuous observations. We then ask for the first moment of the weights of
all the continuous observations in the case where we only have continuous
observations. This is, as follows from Lemma 6.2.43 and Lemma 6.2.38,
convergent to 1. Hence indeed we stay bounded.   □

Lemma 6.2.47 (Errors of second moment of weights of continuous part stay
bounded)**.** *The errors in the ratios in the derived form for the sum of the
second moment of weights corresponding to observations from the continuous
part remain bounded. I.e. there exists an $L > 0$ such that for all $\sigma \in S^{K_n}$*

$$\frac{\sum_{k \in C_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+2}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}} \leq L$$

*P_0 almost surely in the model where $X_1, \cdots, X_n | P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$
where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0,1]$ which
admits a Lebesgue density h and $a, b > 0$ and g a twice continuously dif-
ferentiable function on $[0,1]$ which is bounded away from zero such that
$h(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$.*

Proof. We want to prove that in every fraction as derived in Lemma 6.2.34
the errors stay bounded. So by Lemma 6.2.34 we know

$$\tilde{\mathbb{E}}\left[\sum_{k \in C_n} W_{I_k}^2\right] =$$

$$\frac{\sum_{\sigma \in S^m} \sum_{k \in C_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+2}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\sum_{\sigma \in S^m} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}} .$$

This means we want to show that uniformly in all $\sigma \in S^{K_n}$ we have that

$$\frac{\sum_{k \in C_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+2}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

remains bounded. Observe that it does not matter for the distribution of the weights of the continuous part, if we replace the weights of the i-th distinct discrete observation by N_i observations which occur only once, since we can bound uniformly by a constant factor, by Lemma 6.2.11. If we then increase the sum by adding all these non-negative weights. P_0 almost surely we have every observation exactly once. This means we are in the situation where we only have continuous observations. We then ask for the sum of second moments of the weights of all the continuous observations in the case where we only have continuous observations. This is, as computed in Lemma 6.2.44 and Lemma 6.2.41, convergent to 0. Hence indeed we stay bounded P_0 almost surely. □

Lemma 6.2.48 (Errors of mixed moments of weights of continuous part stay bounded)**.** *Suppose that the true distribution $P_0 = P_d + \lambda P_c$ where P_d is a finite discrete distribution. Suppose we have K_n distinct observations. For all permutations $\sigma \in S^{K_n}$*

$$
\tilde{\mathbb{E}}\left[\sum_{i,j\in C, i\neq j} W_{I_k} W_{I_m}\right] =
$$

$$
2\frac{\displaystyle\sum_{s,t:\sigma^{-1}(s),\sigma^{-1}(t)\in C, s<t} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]}\prod_{l=1}^{t-1}\frac{\tilde{\mathbb{E}}\left[V^{n p_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\displaystyle\prod_{l=1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{n p_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}
$$

$$
\frac{\displaystyle\prod_{l=t+1}^{s-1}\frac{\tilde{\mathbb{E}}\left[V^{n p_l}(1-V)^{1+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]}\prod_{l=s+1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{n p_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1} \leq L
$$

P_0 almost surely in the model where $X_1, \cdots, X_n|P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0,1]$ which admits a Lebesgue density h and $a, b > 0$ and g a twice continuously differentiable function on $[0,1]$ which is bounded away from zero such that $h(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$.

Proof. Observe that it does not matter for the distribution of the weights of the continuous part, if we replace the weights of the i-th distinct discrete observation by N_i observations which occur only once, since we can bound uniformly by a constant factor, by Lemma 6.2.11. If we then increase the sum by adding all these non-negative weights, we are in the situation where we only have continuous observations. We then ask for the sum of all mixed first moment of the weights of all the continuous observations in the case where we only have continuous observations. This is, as we have seen in

Lemma 6.2.45 and Lemma 6.2.42, convergent to 1. Hence indeed we stay
bounded.   □

In order to collect the results on uniform convergence, we split the result
into two parts, the "head" of the permutation and the "tail" of the permu-
tation. For those we need different techniques to control them. Namely,
everything in the "head" of the permutation should converge to zero, and
everything in the "tail" has to converge to the right value. This is what we
will do now.

Lemma 6.2.49 (Weights in front of discrete observations have zero contri-
bution). *Suppose $P_0 = P_d + \lambda P_c$ with $\lambda > 0$. Define $C_n = \{i : P_0(\{\hat{X}_i\}) = 0\}$.
Let T_n be a sequence of permutations such that for all $\sigma \in T_n$ $D_n = \{i \in
C_n : \exists j : P_0(\{\hat{X}_j\}) > 0 \text{ and } \sigma(i) \le \sigma(j)\}$ has cardinality at most K_n with
$\frac{K_n}{n} \to 0$. Then*

$$\sum_{k \in D_n} \frac{\tilde{\mathbb{E}}\left[V^{np_k,n+q}(1-V)^{T_{\sigma,\sigma(k)}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}$$

$$\overline{\sum_{\sigma \in S^m} \prod_{l=1}^m \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\sup_{\sigma \in T_n} \sum_{i \in D_n} \tilde{\mathbb{E}}\left[W_{I_i}^k\right] \to 0 \qquad P_0 \ a.s.$$

*in the model where $X_1, \cdots, X_n | P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is
an atomless base measure and $\mathcal{D}$ is a distribution on $[0,1]$ which admits a
Lebesgue density h and $a, b > 0$ and g a twice continuously differentiable
function on $[0,1]$ which is bounded away from zero such that $h(v) = v^{a-1}(1-
v)^{b-1}g(v)$ for all $v \in (0,1)$.*

Proof. We denote the weights in front of discrete observations by $D_n = \{i \in
C_n : \exists j : P_0(\{\hat{X}_j\}) > 0 \text{ and } \sigma(i) \le \sigma(j)\}$. Then $|D_n| \le K_n$. We first observe
that by Lemma A.3

$$\frac{\tilde{\mathbb{E}}\left[V^q(1-V)^n\right]}{\tilde{\mathbb{E}}\left[V(1-V)^n\right]} = (1+r_{q,n}) \prod_{i=0}^{q-1} \frac{1+a+i}{1+a+b+cn+i}.$$

With

$$\lim_{n \to \infty} (1+r_{q,n})^{K_n} \to 1.$$

By the strong law of large numbers, the number of the observations from the continuous part of the true distribution divided by the total number of observations approaches λ. Because there are at most K_n weights corresponding to observations of the continuous part of the true distribution in front of weights corresponding to observations from the discrete part, and $\frac{K_n}{n} \to 0$, eventually there are at least $\frac{\lambda n}{2}$ observations placed behind the weight we are inspecting, hence the y in

$$\frac{\tilde{\mathbb{E}}\left[V^q(1-V)^y\right]}{\tilde{\mathbb{E}}\left[V(1-V)^y\right]}$$

is at least $\frac{\lambda n}{2}$. If $\lambda > 0$ then $\frac{\lambda n}{2} \to \infty$. Note that every factor other factor appearing is less than 1, and hence

$$\sup_{\sigma \in T_n} \sum_{i \in D_n} \tilde{\mathbb{E}}\left[W_{I_i}\right] \le K_n \frac{\tilde{\mathbb{E}}\left[V^q(1-V)^{\frac{\lambda}{2}n}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{\frac{\lambda}{2}n}\right]} \qquad P_0 \text{ a.s.}.$$

Now since $\frac{K_n}{n} \to 0$ we know that $K_n \frac{1+a}{1+a+b+cn} \to 0$. Therefore $K_n \prod_{i=0}^{k-1} \frac{1+a+i}{1+a+b+cn+i} \to 0$. We also know and $(1 + r_{q, \frac{\lambda}{2}n}) \to 1$. Hence

$$\sup_{\sigma \in T_n} \sum_{i \in D_n} \tilde{\mathbb{E}}\left[W_{I_i}\right] \to 0 \qquad P_0 \text{ a.s.}$$

$\square$

Lemma 6.2.50 (Weights in the tail have asymptotic mass λ). *Suppose $P_0 = P_d + \lambda P_c$ with $\lambda > 0$. Define $C_n = \{i : P_0(\{\hat{X}_i\}) = 0\}$. Suppose T_n is a sequence of sets of permutations such that $D_n = \{i \in C_n : \exists j : P_0(\{\hat{X}_j\}) > 0 \text{ and } \sigma(i) \le \sigma(j)\}$ has cardinality at most K_n with $\frac{K_n}{n} \to 0$. Let $E_n = C_n \setminus D_n$. Then*

$$\frac{\sum_{k \in E_n} \frac{\tilde{\mathbb{E}}\left[V^{np_k,n+1}(1-V)^{T_{\sigma,\sigma(k)}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

Converges uniformly to λ on T_n P_0 almost surely in the model where $X_1, \cdots, X_n | P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0,1]$ which admits a Lebesgue density h and $a, b > 0$ and g

a twice continuously differentiable function on $[0,1]$ *which is bounded away from zero such that* $h(v) = v^{a-1}(1-v)^{b-1}g(v)$ *for all* $v \in (0,1)$.

Proof. Observe that the image of E_n under σ is the tail of the permutation. This means that there exists a natural number k such that $i \in E_n$ if and only if $\sigma(i) > s$. This means that, for all $k \in E_n$, the first s terms in the product

$$\frac{\frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+1}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

are the same. Since they are all in front of k, they all have the $+1$ term in the power. Note that $s \le K_n + d$, and $\frac{K_n+d}{n} \to 0$. We want to study the limiting behavior of this product. First observe that this product is

$$\frac{\prod_{l<s} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]}}{\prod_{l=1}^{s} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}.$$

Now by Lemma 6.2.13 we can ignore the $\frac{1}{1-\tilde{\mathbb{E}}[\cdots]}$ terms as they will converge uniformly to 1. Hence we want to study

$$\frac{\prod_{l<s} \tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{\prod_{l=1}^{s} \tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}.$$

Now we can apply Lemma A.2. This gives that

$$\frac{\prod_{l<s} \tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{\prod_{l=1}^{s} \tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]} = \prod_{l=1}^{s} \frac{T_{\sigma,l+1}+b}{T_{\sigma,l}+a+b}(1+R_\sigma)$$

with $\sup_\sigma |R_\sigma| \to 0$. Since $\frac{T_{\sigma,l+1}}{n} \to \lambda\, P_0$ almost surely, i.e. the number of observations from the continuous part minus K_n divided by the total number of observations converges to the probability of the continuous part by the strong law of large numbers, we can apply Lemma 6.2.12 to conclude that

$$\prod_{l=1}^{s} \frac{T_{\sigma,l+1}+b}{T_{\sigma,l}+a+b}$$

converges uniformly to λ. Therefore

$$\frac{\prod\limits_{l<s}\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{\prod\limits_{l=1}^{s}\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}\to\lambda.$$

uniformly in all $\sigma\in T_n$ P_0 almost surely.

Now we study the other part of the product, namely

$$\frac{\frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+1}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,\sigma(k)}}\right]}\prod\limits_{s<l<\sigma(k)}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]}\prod\limits_{l>\sigma(k)}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod\limits_{l=s+1}^{m}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

Now we study the other part of the product, namely

$$\frac{\frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+2}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,\sigma(k)}}\right]}\prod\limits_{s<l<\sigma(k)}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}\prod\limits_{l>\sigma(k)}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod\limits_{l=s+1}^{m}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

Note that P_0 almost surely, we only look at observations which have occurred exactly one time. Therefore this is just the same as studying the weight corresponding to the case where we only have a continuous part and have $|E_n|$ many observations. Since $E_n\to\infty$ uniformly on T_n ($|C_n|\geq|E_n|\geq|C_n|-K_n$). So we can apply Lemma 6.2.38 to find that this product is equal to

$$\sum_{j=0}^{n-1}\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{j+1}\right]}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^j\right]}{\tilde{\mathbb{E}}\left[V(1-V)^j\right]}\frac{1-\tilde{\mathbb{E}}\left[(1-V)^n\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{j+2}\right]}.$$

And then by Lemma 6.2.43 this converges uniformly on T_n to 1. Hence

$$\frac{\frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+1}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,\sigma(k)}}\right]}\prod\limits_{l<\sigma(k)}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]}\prod\limits_{l>\sigma(k)}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod\limits_{l=1}^{m}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

converges uniformly on T_n to λ P_0 almost surely $\qquad\square$

Lemma 6.2.51 (Second moment of weights in the tail converge to 0). *Let $P_0 = P_d + \lambda P_c$ where P_d is a discrete subprobabiity distribution and P_c is a continuous probability distribution.*

$$\frac{\sum_{k \in E_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+2}(1-V)^{T_{\sigma,\sigma(k)}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\sum_{\sigma \in S^m} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

Converges uniformly to zero in the model where $X_1, \cdots, X_n | P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0,1]$ which admits a Lebesgue density h and $a, b > 0$ and g a twice continuously differentiable function on $[0,1]$ which is bounded away from zero such that $h(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$.

The proof of this lemma resembles the proof of Lemma 6.2.50 a lot. The only thing that changes is that we get higher moments, and thus the expressions differ a tiny bit, and we need to refer to the lemma regarding the sum of second moments instead of sum of first moments.

Proof. Observe that the image of E_n under σ is the tail of the permutation. This means that there exists a natural number k such that $i \in E_n$ if and only if $\sigma(i) > s$. This means that, for all $k \in E_n$, the first s terms in the product

$$\frac{\frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+2}(1-V)^{T_{\sigma,\sigma(k)}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\sum_{\sigma \in S^m} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

are the same. Since they are all in front of k, they all have the $+2$ term in the power. Note that $s \leq K_n + d$, and $\frac{K_n+d}{n} \to 0$. We want to study the limiting behavior of this product. First observe that this product is

$$\frac{\prod_{l<s} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\prod_{l=1}^{s} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}.$$

Now by Lemma 6.2.13 we can ignore the $\frac{1}{1-\tilde{\mathbb{E}}[\cdots]}$ terms as they will converge uniformly to 1. Hence we want to study

$$\frac{\prod\limits_{l<s} \tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{\prod\limits_{l=1}^{s} \tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}.$$

Now we can apply Lemma A.2. This gives that

$$\frac{\prod\limits_{l<s} \tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{\prod\limits_{l=1}^{s} \tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]} = \prod_{l=1}^{s} \frac{T_{\sigma,l+1}+b+1}{T_{\sigma,l}+a+b+1}\frac{T_{\sigma,l+1}+b}{T_{\sigma,l}+a+b}(1+R_\sigma)$$

with $\sup_\sigma |R_\sigma| \to 0$. Since $\frac{T_{\sigma,l+1}}{n} \to \lambda\, P_0$ almost surely, i.e. the number of observations from the continuous part minus K_n divided by the total number of observations converges to the probability of the continuous part by the strong law of large numbers, we can apply Lemma 6.2.12 to conclude that

$$\prod_{l=1}^{s} \frac{T_{\sigma,l+1}+b+1}{T_{\sigma,l}+a+b+1}\frac{T_{\sigma,l+1}+b}{T_{\sigma,l}+a+b}$$

converges uniformly to λ^2. Therefore

$$\frac{\prod\limits_{l<s} \tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{\prod\limits_{l=1}^{s} \tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]} \to \lambda^2.$$

uniformly in all $\sigma \in T_n\, P_0$ almost surely. Note that P_0 almost surely, we only look at observations which have occurred exactly one time. Therefore this is just the same as studying the weight corresponding to the case where we only have a continuous part and have $|E_n|$ many observations. Since $E_n \to \infty$ uniformly on T_n ($|C_n| \geq |E_n| \geq |C_n| - K_n$). So we can apply Lemma 6.2.41 to find that this product is equal to.

$$\sum_{j=1}^{n} \frac{1-\tilde{\mathbb{E}}\left[(1-V)^n\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{n+3-l}\right]}\frac{1-\tilde{\mathbb{E}}\left[(1-V)^{n-1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{n+2-l}\right]}$$

$$\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+2}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n+1-l}\right]}\cdot\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]}\frac{\tilde{\mathbb{E}}\left[V^3(1-V)^{n-l}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-l}\right]}.$$

And then by Lemma 6.2.44 this converges uniformly on T_n to 0. Hence

$$\frac{\dfrac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+1}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,\sigma(k)}}\right]}\displaystyle\prod_{l<\sigma(k)}\dfrac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]}\prod_{l>\sigma(k)}\dfrac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\displaystyle\prod_{l=1}^{m}\dfrac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

Converges uniformly on T_n to 0 P_0 almost surely   □

Lemma 6.2.52 (Mixed terms in front of discrete weights do not contribute anything on good sets). *Suppose $P_0 = P_d + \lambda P_c$ with $\lambda > 0$. Define $C_n = \{i : P_0(\{\hat{X}_i\}) = 0\}$. Suppose T_n is a sequence of sets of permutations such that $D_n = \{i \in C_n : \exists j : P_0(\{\hat{X}_j\}) > 0$ and $\sigma(i) \leq \sigma(j)\}$ has cardinality at most K_n with $\frac{K_n}{n} \to 0$. Then*

$$2\frac{\displaystyle\sum_{i,j\in E_n}\dfrac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]}\dfrac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]}\displaystyle\prod_{l=1}^{t-1}\dfrac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\displaystyle\prod_{l=1}^{K_n}\dfrac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\frac{\displaystyle\prod_{l=t+1}^{s-1}\dfrac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]}\displaystyle\prod_{l=s+1}^{K_n}\dfrac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1} \to 0 \qquad P_0 \ a.s.$$

in the model where $X_1, \cdots, X_n | P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0,1]$ which admits a Lebesgue density h and $a, b > 0$ and g a twice continuously differentiable function on $[0,1]$ which is bounded away from zero such that $h(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$.

The main argument is the same as in Lemma 6.2.49. However, now we have terms which are not yet in the "convergent" part so we need to act a bit more clever. The trick is to sum all the weights which are behind all the discrete weights.

Proof. We split the problem into two parts, the sum such that both $i, j \in D_n$ and the sum such that $i \in D_n, j \notin D_n$. These require a bit different analysis. The first is a straightforward modification of the proof of Lemma 6.2.49, but now with two terms. This gives

$$\sup_{\sigma \in T_n} 2 \frac{\sum_{i,j \in D_n, j:\sigma(j)>\sigma(i)} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]} \prod_{l=1}^{t-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\prod_{l=1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\frac{\prod_{l=t+1}^{s-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]} \prod_{l=s+1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}$$

$$\leq K_n^2 \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{\frac{\lambda}{2}n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{\frac{\lambda}{2}n+1}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{\frac{\lambda}{2}n}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{\frac{\lambda}{2}n}\right]} \qquad P_0 \text{ a.s.}$$

The same arguments as before apply so that this term is $o(n^2)$ and we have K_n^2 terms, hence

$$K_n^2 \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{\frac{\lambda}{2}n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{\frac{\lambda}{2}n+1}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{\frac{\lambda}{2}n}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{\frac{\lambda}{2}n}\right]} \to 0$$

And thus

$$\sup_{\sigma \in T_n} 2 \frac{\sum_{i,j \in D_n, j:\sigma(j)>\sigma(i)} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]} \prod_{l=1}^{t-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\prod_{l=1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\frac{\prod_{l=t+1}^{s-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]} \prod_{l=s+1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}$$

$$\to 0 \qquad P_0 \text{ a.s.}$$

For the other part of the claim, fix $i \in D_n$ and then we sum all $j \notin D_n$.

Using that all other factors which appear are less than one, we get

$$
\sup_{\sigma \in T_n} 2 \frac{\sum_{i \in D_n, j \notin D_n} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]} \prod_{l=1}^{t-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\prod_{l=1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}
$$

$$
\frac{\prod_{l=t+1}^{s-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]} \prod_{l=s+1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}
$$

$$
\leq \sup_{\sigma \in T_n} \sum_{i \in D_n} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,\sigma(i)+1}}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{T_{\sigma,\sigma(i)+1}}\right]}
$$

$$
\frac{\sum_{k \in E_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n+q}}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}} .
$$

From Lemma 6.2.50 we know that uniformly on all $\sigma \in T_n$ we have

$$
\frac{\sum_{k \in E_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n+q}}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}} \to \lambda .
$$

We know that uniformly over all $\sigma \in T_n$ we have $\sum_{i \in D_n} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,\sigma(i)+1}}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{T_{\sigma,\sigma(i)+1}}\right]} \to 0$ by Lemma 6.2.49 hence

$$
\sup_{\sigma \in T_n} 2 \frac{\sum_{i \in D_n, j \notin D_n} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]} \prod_{l=1}^{t-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\prod_{l=1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}
$$

$$
\frac{\prod_{l=t+1}^{s-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]} \prod_{l=s+1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}
$$

$$
\to 0 \qquad P_0 \text{ a.s.}
$$

This shows

$$
2 \frac{\sum_{i,j \in E_n} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]} \prod_{l=1}^{t-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\prod_{l=1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}
$$

$$
\frac{\prod_{l=t+1}^{s-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]} \prod_{l=s+1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1} \to 0 \qquad P_0 \text{ a.s.}
$$

$\square$

Lemma 6.2.53 (Mixed terms in the tail converge to λ^2). *Suppose the true distribution $P_0 = P_d + \lambda P_c$ where P_c is a continuous probability distribution and P_d is a finite discrete subprobability measure. Then*

$$
2 \frac{\sum\limits_{s,t:\sigma^{-1}(s),\sigma^{-1}(t)\in E_n,s<t} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]} \prod\limits_{l=1}^{t-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\prod\limits_{l=1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}
$$

$$
\frac{\prod\limits_{l=t+1}^{s-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]} \prod\limits_{l=s+1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}
$$

converges uniformly to λP_0 almost surely in the model where $X_1, \cdots, X_n|P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0,1]$ which admits a Lebesgue density h and $a, b > 0$ and g a twice continuously differentiable function on $[0,1]$ which is bounded away from zero such that $h(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$.

The proof of this lemma resembles the proof of Lemma 6.2.50 a lot. The only thing that changes is that we get higher moments, and thus the expressions differ a tiny bit, and we need to refer to the lemma regarding the sum of second moments instead of sum of first moments.

Proof. Observe that the image of E_n under σ is the tail of the permutation. This means that there exists a natural number k such that $i \in E_n$ if and only if $\sigma(i) > s$. This means that, for all s, t such that $s < t$ and $\sigma^{-1}(s), \sigma^{-1}(t) \in E_n$, the first u terms in the product

$$
\frac{\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]} \prod\limits_{l=1}^{t-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\prod\limits_{l=1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}
$$

$$
\frac{\prod\limits_{l=t+1}^{s-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]} \prod\limits_{l=s+1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}
$$

are the same. Since they are all in front of i, j, they all have the $+2$ term in the power. Note that $s \leq K_n + d$, and $\frac{K_n+d}{n} \to 0$. We want to study the

limiting behavior of this product. First observe that this product is

$$\frac{\displaystyle\prod_{l<u}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\displaystyle\prod_{l=1}^{u}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}.$$

Now by Lemma 6.2.13 we can ignore the $\frac{1}{1-\tilde{\mathbb{E}}[\cdots]}$ terms as they will converge uniformly to 1. Hence we want to study

$$\frac{\displaystyle\prod_{l<u}\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{\displaystyle\prod_{l=1}^{u}\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}.$$

Now we can apply Lemma A.2. This gives that

$$\frac{\displaystyle\prod_{l<u}\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{\displaystyle\prod_{l=1}^{u}\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}=\prod_{l=1}^{s}\frac{T_{\sigma,l+1}+b+1}{T_{\sigma,l}+a+b+1}\frac{T_{\sigma,l+1}+b}{T_{\sigma,l}+a+b}(1+R_\sigma)$$

with $\sup_\sigma|R_\sigma|\to 0$. Since $\frac{T_{\sigma,l+1}}{n}\to\lambda$ P_0 almost surely, i.e. the number of observations from the continuous part minus K_n divided by the total number of observations converges to the probability of the continuous part by the strong law of large numbers, we can apply Lemma 6.2.12 to conclude that

$$\prod_{l=1}^{u}\frac{T_{\sigma,l+1}+b+1}{T_{\sigma,l}+a+b+1}\frac{T_{\sigma,l+1}+b}{T_{\sigma,l}+a+b}$$

converges uniformly to λ^2. Therefore

$$\frac{\displaystyle\prod_{l<u}\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{\displaystyle\prod_{l=1}^{u}\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}\to\lambda^2.$$

uniformly in all $\sigma\in T_n$ P_0 almost surely.

So now we study the other part of this expression

$$\frac{\displaystyle\sum_{s,t:\sigma^{-1}(s),\sigma^{-1}(t)\in E_n,s<t}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]}\prod_{l=u+1}^{t-1}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\displaystyle\prod_{l=u+1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\frac{\displaystyle\prod_{l=t+1}^{s-1}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]}\prod_{l=s+1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}$$

Note that P_0 almost surely, we only look at observations which have occurred exactly one time. Therefore this is just the same as studying the weight corresponding to the case where we only have a continuous part and have $|E_n|$ many observations. Since $E_n \to \infty$ uniformly on T_n ($|C_n| \geq |E_n| \geq |C_n| - K_n$). So we can apply Lemma 6.2.42 to find that this product is equal to.

$$\sum_{1\leq i<j\leq n}\frac{1-\tilde{\mathbb{E}}\left[(1-V)^n\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{n+3-l}\right]}\frac{1-\tilde{\mathbb{E}}\left[(1-V)^{n-1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{n+2-l}\right]}\frac{1-\tilde{\mathbb{E}}\left[(1-V)^{n-i}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{n+2-j}\right]}$$

$$\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+2}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-i+1}\right]}\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-i}\right]}\frac{\tilde{\mathbb{E}}\left[V(1-V)^{n-i+1}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-j}\right]}$$

$$\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{n-i}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-i}\right]}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{n-j}\right]}{\tilde{\mathbb{E}}\left[V(1-V)^{n-j}\right]}.$$

And then by Lemma 6.2.45 this converges uniformly on T_n to 1. Hence

$$2\frac{\displaystyle\sum_{s,t:\sigma^{-1}(s),\sigma^{-1}(t)\in E_n,s<t}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]}\prod_{l=1}^{t-1}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\displaystyle\prod_{l=1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\frac{\displaystyle\prod_{l=t+1}^{s-1}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]}\prod_{l=s+1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}$$

Converges uniformly on T_n to 0 P_0 almost surely $\qquad\square$

Now we have shown that the terms in the "head" of the permutations converge uniformly to zero and the terms in the "tail" of the permutations converge uniformly to the right answer. This allows us to compute the total influence, which is basically collecting the results from the previous lemmas.

Lemma 6.2.54 (First moments of weights have asymptotic mass λ). *Suppose $P_0 = P_d + \lambda P_c$ with $\lambda > 0$. Define $C_n = \{i : P_0(\{\hat{X}_i\}) = 0\}$. Suppose T_n is a sequence of permutations such that $D_n = \{i \in C_n : \exists j : P_0(\{\hat{X}_j\}) > 0 \text{ and } \sigma(i) \le \sigma(j)\}$ has cardinality at most K_n with $\frac{K_n}{n} \to 0$. Then*

$$\frac{\sum_{k \in C_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+q}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

converges uniformly on T_n to λP_0 almost surely in the model where $X_1, \cdots, X_n | P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0,1]$ which admits a Lebesgue density h and $a, b > 0$ and g a twice continuously differentiable function on $[0,1]$ which is bounded away from zero such that $h(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$.

Proof. We use linearity to split this sum. Define $E_n = C_n \setminus D_n$. This gives

$$\frac{\sum_{k \in C_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+q}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$= \frac{\sum_{k \in D_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+q}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$+ \frac{\sum_{k \in E_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+q}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\tilde{\mathbb{E}}\left[\sum_{i \in C_n} W_{I_i}\right] = \tilde{\mathbb{E}}\left[\sum_{i \in D_n} W_{I_i}\right] + \tilde{\mathbb{E}}\left[\sum_{i \in E_n} W_{I_i}\right]$$

Because

$$\frac{\sum_{k \in D_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+q}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

converges uniformly on T_n to zero by Lemma 6.2.49 and

$$\sum_{k\in E_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+q}(1-V)^{T_{\sigma,\sigma(k)}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}$$
$$\Big/ \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}$$

converges uniformly on T_n to λ by Lemma 6.2.50 we can apply Lemma 6.2.17 which implies that

$$\tilde{\mathbb{E}}\left[\sum_{i\in C_n} W_{I_i}\right]$$

converges uniformly on T_n to λP_0 almost surely. $\square$

Lemma 6.2.55 (Second moments of weights do not contribute anything on good sets)**.** *Suppose* $P_0 = P_d + \lambda P_c$ *with* $\lambda > 0$*. Define* $C_n = \{i : P_0(\{\hat{X}_i\}) = 0\}$*. Suppose* T_n *is a sequence of permutations such that* $D_n = \{i \in C_n : \exists j : P_0(\{\hat{X}_j\}) > 0 \text{ and } \sigma(i) \le \sigma(j)\}$ *has cardinality at most* K_n *with* $\frac{K_n}{n} \to 0$*. Then*

$$\sum_{k\in C_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+2}(1-V)^{T_{\sigma,\sigma(k)}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}$$
$$\Big/ \sum_{\sigma\in S^m} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}$$

converges uniformly on T_n *to* 0 P_0 *almost surely in the model where* $X_1, \cdots, X_n | P \sim P$ *and* $P \sim SBP(\alpha, \mathcal{D})$ *where* α *is an atomless base measure and* $\mathcal{D}$ *is a distribution on* $[0,1]$ *which admits a Lebesgue density* h *and* $a, b > 0$ *and* g *a twice continuously differentiable function on* $[0,1]$ *which is bounded away from zero such that* $h(v) = v^{a-1}(1-v)^{b-1}g(v)$ *for all* $v \in (0,1)$*.*

Proof. We use linearity to split this sum. Define $E_n = C_n \setminus D_n$. This gives

$$\frac{\sum_{k \in C_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+2}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\sum_{\sigma \in S^m} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$= \frac{\sum_{k \in D_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+2}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\sum_{\sigma \in S^m} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$+ \frac{\sum_{k \in E_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+2}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\sum_{\sigma \in S^m} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

Because

$$\frac{\sum_{k \in D_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+2}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\sum_{\sigma \in S^m} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

converges uniformly on T_n to zero by Lemma 6.2.49 and

$$\frac{\sum_{k \in E_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+2}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\sum_{\sigma \in S^m} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

converges uniformly on T_n to zero by Lemma 6.2.51 we can apply Lemma 6.2.17 which implies

$$\frac{\sum_{k \in C_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+2}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\sum_{\sigma \in S^m} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

converges uniformly on T_n to 0 P_0 almost surely.   □

Lemma 6.2.56 (Mixed terms have asymptotic mass λ^2 on good sets)**.** *Suppose $P_0 = P_d + \lambda P_c$ with $\lambda > 0$. Define $C_n = \{i : P_0(\{\hat{X}_i\}) = 0\}$. Suppose T_n is a sequence of permutations such that $D_n = \{i \in C_n : \exists j : P_0(\{\hat{X}_j\}) > 0 \text{ and } \sigma(i) \le \sigma(j)\}$ has cardinality at most K_n with $\frac{K_n}{n} \to 0$. Then*

$$2\frac{\sum\limits_{s,t:\sigma^{-1}(s),\sigma^{-1}(t)\in C_n,s<t}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]}\prod\limits_{l=1}^{t-1}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\prod\limits_{l=1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\frac{\prod\limits_{l=t+1}^{s-1}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]}\prod\limits_{l=s+1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}$$

converges uniformly on T_n to $\lambda^2 P_0$ almost surely in the model where $X_1, \cdots, X_n | P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0,1]$ which admits a Lebesgue density h and $a, b > 0$ and g a twice continuously differentiable function on $[0,1]$ which is bounded away from zero such that $h(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$.

Proof. We use linearity to split this sum. Define $E_n = C_n \setminus D_n$. This gives

$$2\frac{\sum\limits_{s,t:\sigma^{-1}(s),\sigma^{-1}(t)\in C_n,s<t}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]}\prod\limits_{l=1}^{t-1}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\prod\limits_{l=1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\frac{\prod\limits_{l=t+1}^{s-1}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]}\prod\limits_{l=s+1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}$$

$$= 2\frac{\sum\limits_{i,j\in E_n}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]}\prod\limits_{l=1}^{t-1}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\prod\limits_{l=1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\frac{\prod\limits_{l=t+1}^{s-1}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]}\prod\limits_{l=s+1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}$$

$$+ 2\frac{\sum\limits_{i\in D_n,j:\sigma(j)>\sigma(i)}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]}\prod\limits_{l=1}^{t-1}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\prod\limits_{l=1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\frac{\prod\limits_{l=t+1}^{s-1}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]}\prod\limits_{l=s+1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}$$

Because

$$2\frac{\sum_{i,j\in E_n}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]}\prod_{l=1}^{t-1}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\prod_{l=1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\frac{\prod_{l=t+1}^{s-1}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]}\prod_{l=s+1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}$$

converges uniformly on T_n to zero P_0 almost surely by Lemma 6.2.52 and

$$2\frac{\sum_{i\in D_n,j:\sigma(j)>\sigma(i)}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]}\prod_{l=1}^{t-1}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\prod_{l=1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\frac{\prod_{l=t+1}^{s-1}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]}\prod_{l=s+1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}$$

converges uniformly on T_n to λ^2 P_0 almost surely by Lemma 6.2.53 we can apply Lemma 6.2.17
    Hence

$$2\frac{\sum_{s,t:\sigma^{-1}(s),\sigma^{-1}(t)\in C_n,s<t}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]}\frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]}\prod_{l=1}^{t-1}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\prod_{l=1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\frac{\prod_{l=t+1}^{s-1}\frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]}\prod_{l=s+1}^{K_n}\frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}$$

converges uniformly on T_n to λ^2 P_0 almost surely $\square$

    Now we have shown that everything behaves well on the good sets and all the terms appearing remain bounded, we can conclude that the ratios in the expressions for the moments converge to the right answer. This will be the subject of the next three lemmas.

Lemma 6.2.57. *Suppose that the true distribution $P_0 = P_d + \lambda P_c$ where P_d is a finite discrete distribution. Then*

$$\tilde{\mathbb{E}}\left[\sum_{k\in C}W_{I_k}\right]\to\lambda\qquad P_0\ a.s..$$

in the model where $X_1, \cdots, X_n | P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0,1]$ which admits a Lebesgue density h and $a, b > 0$ and g a twice continuously differentiable function on $[0,1]$ which is bounded away from zero such that $h(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$.

Proof. We use the expression of

$$\tilde{\mathbb{E}}\left[\sum_{k \in C} W_{I_k}\right]$$

derived in Lemma 6.2.34

$$\tilde{\mathbb{E}}\left[\sum_{k \in C_n} W_{I_k}^q\right] =$$

$$\frac{\sum_{\sigma \in S^m} \sum_{k \in C_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+q}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\sum_{\sigma \in S^m} \prod_{l=1}^m \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}.$$

Now we apply Lemma 6.2.30 to find good sets. This gives us a sequence $(K_n)_{n\in\mathbb{N}}$ and sets T_n of asymptotic probability one. This gives exactly what we need to apply Lemma 6.2.54, which shows that on T_n

$$\frac{\sum_{k \in C_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+q}(1-V)^{T_{\sigma,\sigma(k)+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{q+T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{q+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\prod_{l=1}^m \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l+1}}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

converges uniformly to λP_0 almost surely. By Lemma 6.2.46 the errors remain bounded and hence we can apply Lemma 6.2.9 to conclude that almost surely

$$\tilde{\mathbb{E}}\left[\sum_{k \in C} W_{I_k}\right] \to \lambda \qquad P_0 \text{ a.s..}$$

$\square$

Lemma 6.2.58. *Suppose that the true distribution $P_0 = P_d + \lambda P_c$ where P_d is a finite discrete distribution. Then*

$$\tilde{\mathbb{E}}\left[\sum_{k \in C} W_{I_k}\right] \to 0 \qquad P_0 \ a.s..$$

in the model where $X_1, \cdots, X_n | P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0,1]$ which admits a Lebesgue density h and $a, b > 0$ and g a twice continuously differentiable function on $[0,1]$ which is bounded away from zero such that $h(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$.

Proof. By Lemma 6.2.34 we can find an expression for $\tilde{\mathbb{E}}\left[\sum_{k \in C_n} W_{I_k}^2\right]$.

$$\tilde{\mathbb{E}}\left[\sum_{k \in C_n} W_{I_k}^2\right] =$$

$$\frac{\displaystyle\sum_{\sigma \in S^m} \sum_{k \in C_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+2}(1-V)^{T_{\sigma,\sigma(k)}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\displaystyle\sum_{\sigma \in S^m} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}.$$

Now we apply Lemma 6.2.30 to find good sets. This gives us a sequence $(K_n)_{n \in \mathbb{N}}$ and sets T_n of asymptotic probability one. This gives exactly what we need to apply Lemma 6.2.55, which shows that on T_n

$$\frac{\displaystyle\sum_{k \in C_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{k,n}+2}(1-V)^{T_{\sigma,\sigma(k)}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,\sigma(k)}}\right]} \prod_{l<\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]} \prod_{l>\sigma(k)} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{\displaystyle\sum_{\sigma \in S^m} \prod_{l=1}^{m} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

converges uniformly to 0 P_0 almost surely. By Lemma 6.2.47 the errors remain bounded and hence we can apply Lemma 6.2.9 to conclude that almost surely

$$\tilde{\mathbb{E}}\left[\sum_{k \in C} W_{I_k}\right] \to 0 \qquad P_0 \ \text{a.s..}$$

$\square$

Lemma 6.2.59. *Suppose that the true distribution $P_0 = P_d + \lambda P_c$ where P_d is a finite discrete distribution. Then*

$$\tilde{\mathbb{E}}\left[\sum_{k \in C} W_{I_k}\right] \to \lambda^2 \qquad P_0 \ a.s..$$

in the model where $X_1, \cdots, X_n | P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0,1]$ which admits a Lebesgue density h and $a, b > 0$ and g a twice continuously differentiable function on $[0,1]$ which is bounded away from zero such that $h(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$.

Proof. By Lemma 6.2.35 we know an expression for the moment of the mixed weights. This is

$$\tilde{\mathbb{E}}\left[\sum_{i,j \in C, i \neq j} W_{I_k} W_{I_m}\right] =$$

$$2 \frac{\sum_{\sigma \in \tilde{S}^m} \sum_{s,t:\sigma^{-1}(s),\sigma^{-1}(t) \in C, s<t} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]} \prod_{l=1}^{t-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\sum_{\sigma \in S^{K_n}} \prod_{l=1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\frac{\prod_{l=t+1}^{s-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]} \prod_{l=s+1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}$$

Now we apply Lemma 6.2.30 to find good sets. This gives us a sequence $(K_n)_{n \in \mathbb{N}}$ and sets T_n of asymptotic probability one. This gives exactly what we need to apply Lemma 6.2.56, which shows that on good sets

$$2 \frac{\sum_{s,t:\sigma^{-1}(s),\sigma^{-1}(t) \in C_n, s<t} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,s}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,s}}\right]} \frac{\tilde{\mathbb{E}}\left[V^2(1-V)^{T_{\sigma,t}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,t}}\right]} \prod_{l=1}^{t-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{2+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{2+T_{\sigma,l}}\right]}}{\prod_{l=1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}$$

$$\frac{\prod_{l=t+1}^{s-1} \frac{\tilde{\mathbb{E}}\left[V^{np_l}(1-V)^{1+T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{1+T_{\sigma,l}}\right]} \prod_{l=s+1}^{K_n} \frac{\tilde{\mathbb{E}}\left[V^{np_{\sigma^{-1}(l)}}(1-V)^{T_{\sigma,l}+1}\right]}{1-\tilde{\mathbb{E}}\left[(1-V)^{T_{\sigma,l}}\right]}}{1}$$

converge uniformly to λ^2 P_0 almost surely. By Lemma 6.2.48 the errors remain bounded and hence we can apply Lemma 6.2.9 to conclude that almost surely

$$\tilde{\mathbb{E}}\left[\sum_{k \in C} W_{I_k}\right] \to \lambda^2 \qquad P_0 \ a.s..$$

$\square$

Lemma 6.2.60. *Suppose that the true distribution $P_0 = P_d + \lambda P_c$ where P_d is a finite discrete distribution. Then*

$$\tilde{\mathbb{E}}\left[\left(\sum_{k\in C} W_{I_k}\right)^2\right] \to \lambda^2 \qquad P_0 \ a.s..$$

in the model where $X_1, \cdots, X_n | P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0,1]$ which admits a Lebesgue density h and $a, b > 0$ and g a twice continuously differentiable function on $[0,1]$ which is bounded away from zero such that $h(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$.

Proof. We first rewrite the square, this yields

$$\tilde{\mathbb{E}}\left[\left(\sum_{k\in C} W_{I_k}\right)^2\right] = \tilde{\mathbb{E}}\left[\sum_{k\in C} W_{I_k}^2 + \sum_{i,j\in C, i\neq j} W_{I_i}W_{I_j}\right].$$

Then by Lemma 6.2.58 we know that

$$\tilde{\mathbb{E}}\left[\sum_{k\in C} W_{I_k}^2\right] \to 0 \qquad P_0 \ a.s..$$

and by Lemma 6.2.59 we we also know that

$$\tilde{\mathbb{E}}\left[\sum_{i,j\in C, i\neq j} W_{I_i}W_{I_j}\right] \to \lambda^2 \qquad P_0 \ a.s..$$

Hence

$$\tilde{\mathbb{E}}\left[\left(\sum_{k\in C} W_{I_k}\right)^2\right] \to \lambda^2 \qquad P_0 \ a.s..$$

$\square$

Finally we want to show that the variance of the weights in the continuous part converge to zero.

Lemma 6.2.61. *Suppose that the true distribution $P_0 = P_d + \lambda P_c$ where P_d is a finite discrete distribution. Then*

$$\tilde{Var}\left(\left(\sum_{k \in C} W_{I_k}\right)^2\right) \to 0 \qquad P_0 \ a.s..$$

in the model where $X_1, \cdots, X_n | P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0, 1]$ which admits a Lebesgue density h and $a, b > 0$ and g a twice continuously differentiable function on $[0, 1]$ which is bounded away from zero such that $h(v) = v^{a-1}(1 - v)^{b-1}g(v)$ for all $v \in (0, 1)$.

Proof. By Lemma 6.2.57 and Lemma 6.2.60 we know that

$$\tilde{\mathbb{E}}\left[\sum_{k \in C} W_{I_k}\right] \to \lambda \qquad P_0 \ a.s.$$

and

$$\tilde{\mathbb{E}}\left[\left(\sum_{k \in C} W_{I_k}\right)^2\right] \to \lambda^2 \qquad P_0 \ a.s..$$

Hence

$$\tilde{Var}\left(\left(\sum_{k \in C} W_{I_k}\right)^2\right) = \tilde{\mathbb{E}}\left[\left(\sum_{k \in C} W_{I_k}\right)^2\right] - \tilde{\mathbb{E}}\left[\sum_{k \in C} W_{I_k}\right] \to \lambda^2 - \lambda^2 = 0.$$

$\square$

6.2.9 The convergence of the posterior mean

So we have shown the weights in the posterior act nicely. We now use this to show that indeed the posterior mean converges to the right answer.

The upcoming lemma is used to extend the convergence on arbitrary large sets to convergence on the total space.

Lemma 6.2.62. *Let P_n be a sequence of random probability measures and P a probability measure. Let f be a bounded measurable function. Suppose there exists a sequence of measurable sets A_k such that*

$$\mathbb{E}[P_n(\mathbb{1}_{A_k}f)] \to P(\mathbb{1}_{A_k}f) \qquad \forall k,$$

and
$$P(A_k) \to 1$$

then
$$\mathbb{E}\left[P_n(f)\right] \to P(f).$$

Proof. We begin by splitting f into $\mathbb{1}_{A_k} f + \mathbb{1}_{A_k^c} f$ and then using linearity of integration and expected values.

$$\begin{aligned}
\mathbb{E}\left[P_n(f)\right] &= \mathbb{E}\left[P_n(\mathbb{1}_{A_k} f + \mathbb{1}_{A_k^c} f)\right] \\
&= \mathbb{E}\left[P_n(\mathbb{1}_{A_k} f)\right] + \mathbb{E}\left[P_n(\mathbb{1}_{A_k^c} f)\right].
\end{aligned}$$

If we denote an upper bound of f by F we can bound the second term:

$$\mathbb{E}\left[P_n(\mathbb{1}_{A_k^c} f)\right] \le F\mathbb{E}\left[P_n(\mathbb{1}_{A_k^c})\right].$$

Now we look at the double sequence $x_{n,k} = \left(\mathbb{E}\left[P_n(\mathbb{1}_{A_k} f)\right], \mathbb{E}\left[P_n(\mathbb{1}_{A_k^c})\right]\right)$ with the product metric. Note that $\lim_{n\to\infty} x_{n,k} = x_k = \left(\mathbb{E}\left[P(\mathbb{1}_{A_k} f)\right], \mathbb{E}\left[P(\mathbb{1}_{A_k^c})\right]\right)$, and $\lim_{k\to\infty} x_k = x = (P(f), 0)$ by dominated convergence theorem. By applying Lemma 6.2.16 we can extract a sequence k_n such that $\lim_{n\to\infty} x_{n,k_n} = (P(f), 0)$. This implies that

$$\mathbb{E}\left[P_n(\mathbb{1}_{A_{k_n}} f)\right] \to P(f)$$

and
$$\mathbb{E}\left[P_n(\mathbb{1}_{A_{k_n}^c} f)\right] \to 0.$$

Thus
$$\mathbb{E}\left[P_n(f)\right] \to p(f).$$

$\square$

The next lemma allows us to use the results on the total weight coming from the first moment to find the convergence of mean over the continuous part of the true distribution.

Lemma 6.2.63. *Let* $C_n = \{i : P_0(\{\hat{X}_i\}) = 0\}$ *be the set of observations coming from the continuous part of the true distribution. Then*

$$\mathbb{E}\left[\sum_{i \in C_n} f(x_i) W_{I_i}\right] = \frac{\sum_{i=1}^m f(x_i)}{|C_n|} \sum_{i=1}^m \mathbb{E}\left[W_{I_i}\right].$$

Proof. First remark that the distribution of W_{I_i} is the same for every $i \in C_n$. Let $k \in C_n$ be arbitrary, then we can rewrite as follows

$$
\begin{aligned}
\mathbb{E}\left[\sum_{i \in C_n} f(x_i) W_{I_i}\right] &= \sum_{i \in C_n} \mathbb{E}\left[f(x_i) W_{I_i}\right] \\
&= \sum_{i \in C_n} f(x_i) \mathbb{E}\left[W_{I_i}\right] \\
&= \sum_{i \in C_n} f(x_i) \mathbb{E}\left[W_{I_1}\right] \\
&= \mathbb{E}\left[W_{I_k}\right] \sum_{i \in C_n} f(x_i) \\
&= \left(\frac{\sum_{i \in C_n} \mathbb{E}\left[W_{I_i}\right]}{|C_n|}\right) \sum_{i \in C_n} f(x_i) \\
&= \left(\sum_{i \in C_n} \mathbb{E}\left[W_{I_i}\right]\right) \frac{\sum_{i \in C_n} f(x_i)}{|C_n|}.
\end{aligned}
$$

$\square$

This lemma states that the influence of the prior disappears.

Lemma 6.2.64. *Let P_0 be a distribution on $\mathcal{X}$ which does not have both infinitely many points with positive probability and a continuous part.*

$$
\mathbb{E}\left[\sum_{i : i \neq I_1, \cdots, I_{K_n}} W_i | X_1, \cdots, X_n\right] \to 0.
$$

in the model where $X_1, \cdots, X_n | P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0, 1]$ which admits a Lebesgue density h and $a, b > 0$ and g a twice continuously differentiable function on $[0, 1]$ which is bounded away from zero such that $h(v) = v^{a-1}(1 - v)^{b-1}g(v)$ for all $v \in (0, 1)$.

Proof. Define $C_n = \{i : P_0(\{\hat{X}_i\}) = 0\}$. Because we know $\sum_{i=1}^{\infty} W_i = 1$ almost surely, and for every $\epsilon > 0$ we can pick discrete support points of the true distribution $\mu_1, \cdots, \mu_k$ such that $P_0\left(\{\mu_1, \cdots, \mu_k\} \cup C_n\right) \geq 1 - \frac{\epsilon}{2}$. Now because we know the weights of these atoms and the weights corresponding

to the continuous part both converge to the true probability, we can find an $N > 0$ so that the total distance between the total weight given to the continuous plus the weights corresponding to the selected discrete points always stays within $\epsilon/2$ of $1 - \frac{\epsilon}{2}$. Therefore, the total weight assigned to the observations is at least $1 - \epsilon$. Thus we know that the total mass the other weights can have is bounded above by ϵ. $\square$

The next lemma states that the posterior mean of converges on a collection of atoms.

Lemma 6.2.65. *Let P_0 be a distribution on $\mathcal{X}$ which does not have both infinitely many points with positive probability and a continuous part. Let $A = \{\mu_1, \cdots, \mu_m\}$ be a subset of $\mathcal{X}$. Let f be a nonnegative bounded measurable function on $\mathcal{X}$. Then*

$$\mathbb{E}\left[P(\mathbb{1}_A f)|X_1, \cdots, X_n\right] \to P_0(\mathbb{1}_A f) \qquad a.s.$$

in the model where $X_1, \cdots, X_n|P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0,1]$ which admits a Lebesgue density h and $a, b > 0$ and g a twice continuously differentiable function on $[0,1]$ which is bounded away from zero such that $h(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$.

Proof. The predictive posterior distribution gives that

$$\mathbb{E}\left[P\mathbb{1}_A f|X_1, \cdots, X_n\right] = \sum_{i:\hat{X}_i \in A} \mathbb{E}\left[W_{I_i}|X_1, \cdots, X_n\right] f(\hat{X}_i)$$
$$+ \alpha(\mathbb{1}_A f) \sum_{i:i \neq I_1, \cdots, I_{K_n}} \mathbb{E}\left[W_{I_i}|X_1, \cdots, X_n\right].$$

By Lemma 6.2.64 we know that the latter term converges to zero. So it is enough to show

$$\sum_{i:\hat{X}_i \in A} \mathbb{E}\left[W_{I_i}|X_1, \cdots, X_n\right] f(\hat{X}_i \to P_0(\mathbb{1}_A f) \qquad P_0\text{a.s.}$$

Now by Lemma 6.2.33 the weights $\mathbb{E}\left[W_{I_i}|X_1, \cdots, X_n\right]$ converge almost surely to the true probability $P_0(\hat{X}_k)$. Since we want to take the limit of the (finite) sum we are done. $\square$

The next lemma states that the posterior mean converges on the continuous part.

Lemma 6.2.66. *Let P_0 be a distribution on $\mathcal{X}$ which does not have both infinitely many points with positive probability and a continuous part. Let $A = \{\mu \in \mathcal{X} : P_0(\{\mu\}) > 0\}$ be the collection of points of positive probability. Let f be a nonnegative bounded measurable function on $\mathcal{X}$. Then*

$$\mathbb{E}\left[P(\mathbb{1}_{A^c} f)|X_1, \cdots, X_n\right] \to P_0(\mathbb{1}_{A^c} f) \qquad P_0 \text{ a.s.}$$

in the model where $X_1, \cdots, X_n|P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0,1]$ which admits a Lebesgue density h and $a, b > 0$ and g a twice continuously differentiable function on $[0,1]$ which is bounded away from zero such that $h(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$.

Proof. Suppose that the true distribution can be written as $P_0 = P_d + \lambda P_c$, where P_d is a discrete (sub)probability distribution and P_c is an atomless probability distribution. The predictive posterior distribution gives that

$$\mathbb{E}\left[P\mathbb{1}_A f|X_1, \cdots, X_n\right] = \mathbb{E}\left[\sum_{i:\hat{X}_i \notin A} f(\hat{X}_i)W_{I_i}|X_1, \cdots, X_n\right]$$
$$+ \alpha(\mathbb{1}_A f) \sum_{i:i \neq I_1, \cdots, I_{K_n}} \mathbb{E}\left[W_{I_i}|X_1, \cdots, X_n\right].$$

Again by Lemma 6.2.64 we know that the second term converges to zero. So it is enough to show that the first term converges to $P_0(\mathbb{1}_{A^c} f)$. By Lemma 6.2.63 we know that for $S = \{i : \hat{X}_i \in A^c\}$ and m being the cardinality of S we can rewrite the in the following way:

$$\mathbb{E}\left[\sum_{i:\hat{X}_i \notin A} f(\hat{X}_i)W_{I_i}|X_1, \cdots, X_n\right] = \frac{\sum_{i=1}^m f(\hat{X}_i)}{|S|} \sum_{i=1}^m \mathbb{E}\left[W_{I_i}\right].$$

Now we have to compute two terms. First we show that

$$\frac{\sum_{i=1}^m f(\hat{X}_i)}{|S|} \to P_C(f).$$

Note that the $\hat{X}_i$ which lie in S are an iid sample from P_c, so by the strong law of large numbers we know that indeed

$$\frac{\sum_{i=1}^m f(\hat{X}_i)}{|S|} \to P_C(f) \qquad P_0 \text{ a.s.}$$

Note that by Lemma 6.2.57

$$\mathbb{E}\left[\sum_{i=1}^m \mathbb{E}\left[W_{I_i}\right] | X_1, \cdots, X_n\right] \to \lambda \qquad P_0 \text{ a.s.}$$

Hence in total we get

$$\mathbb{E}\left[P(\mathbb{1}_{A^c} f) | X_1, \cdots, X_n\right] \to P_0(\mathbb{1}_{A^c} f) \qquad P_0 \text{ a.s.}$$

$\square$

The next lemma states that the posterior mean converges on the total space.

Lemma 6.2.67. *Suppose P_0 is a distribution with does not have both infinitely many points of positive probability and a continuous part. Suppose that the base measure α is atomless and the relative stick-breaking weight distribution $\mathcal{D}$ admits a density h and constants $a, b > 0$ and a twice continuously differentiable function $g : [0, 1] \to \mathbb{R}$ which is bounded away from zero such that $h(v) = v^{a-1}(1-v)^{b-1}g(v)$ for $v \in (0, 1)$. Let f be a non-negative bounded measurable function. If $P \sim SBP(\alpha, \mathcal{D})$ then*

$$\mathbb{E}\left[P(f) | X_1, \cdots, X_n\right] \to P_0(f) \qquad P_0 \text{ a.s.}$$

Proof. Let P_0 be a distribution which does not have both infinitely many points of positive probability and a continuous part. Define $A = \{\mu \in \mathcal{X} : P_0(\{\mu\}) > 0\}$ the set of points of positive probability. For every $n \in \mathbb{N}$ we pick a finite set $A_k = \{\mu_1, \cdots, \mu_m\}$ such that $P_0(A_k \cup A^c) > 1 - \frac{1}{k}$. Then $P_0(A_k \cup A^c) \to 1$.

Let f be a nonnegative bounded measurable function. Then by Lemma 6.2.65 we know
$$\mathbb{E}\left[P(\mathbb{1}_{A_k} f) | X_1, \cdots, X_n\right] \to P_0(\mathbb{1}_{A_k} f) \quad P_0 \text{ a.s.}$$

Similarly from Lemma 6.2.66 it follows that

$$\mathbb{E}\left[P(\mathbb{1}_{A^c} f) | X_1, \cdots, X_n\right] \to P_0(\mathbb{1}_{A^c} f) \quad P_0 \text{ a.s.}$$

Because A_k and A^c are disjoint, we can use linearity to conclude that

$$\mathbb{E}\left[P(\mathbb{1}_{A_k \cup A^c} f) | X_1, \cdots, X_n\right] \to P_0(\mathbb{1}_{A_k \cup A^c} f) \qquad P_0 \text{ a.s.}$$

This works for every k, so we can apply Lemma 6.2.62 to conclude

$$\mathbb{E}\left[P(f) | X_1, \cdots, X_n\right] \to P_0(f) \qquad P_0 \text{ a.s.}$$

$\square$

6.2.10 The convergence of the posterior variance

We have shown that the variance of the weights behave nicely. We now want to extend this to control the variance integrals with respect to the posterior.

We want to show that we can bound the posterior variance on the full set if we can control it on a sufficiently large set.

Lemma 6.2.68. *Let P_n be a sequence of random probability measures and P be a distribution such that $P_n(\mathbb{1}_A) \to P(\mathbb{1}_A)$ for every measurable set A. Suppose that there exists a sequence of measurable sets A_k such that $\mathbb{E}\left[P_n(\mathbb{1}_{A_k} f)\right] \to P(\mathbb{1}_{A_k} f)$, $P(A_k) \to 1$ and $Var(P_n(A_k f)) \to 0$ for all k. Then $Var(P_0(f)) \to 0$.*

Proof.

$$\begin{aligned}
\mathrm{Var}(P_n(f)) &= \mathrm{Var}(P_n(\mathbb{1}_{A_k} f + \mathbb{1}_{A_k^c} f) \\
&= \mathrm{Var}(P_n(\mathbb{1}_{A_k} f)) + \mathbb{P}_n(\mathbb{1}_{A_k^c} f)))
\end{aligned}$$

So in view of Lemma 6.2.15 we only have to show that we can find a sequence k_n such that $\mathrm{Var}(P_n(\mathbb{1}_{A_{k_n}} f)) \to 0$ and $\mathrm{Var}(P_n(\mathbb{1}_{A_{k_n}^c} f))$ converge to zero. We denote $X_n = P_n(\mathbb{1}_{A_{k_n}} f)$ and $Y_n = P_n(\mathbb{1}_{A_{k_n}^c} f)$, not denoting the dependence on k_n.

We start with the second requirement: $\mathrm{Var}(Y_n) \to 0$. For this, we compute the first and second moment. We denote the bound of f by F.

$$\begin{aligned}
\mathbb{E}\left[Y_n\right] &= \mathbb{E}\left[P_n(\mathbb{1}_{A_{k_n}^c} f)\right] \\
&\leq \mathbb{E}\left[P_n(\mathbb{1}_{A_{k_n}^c} F)\right] \\
&= F\mathbb{E}\left[P_n(\mathbb{1}_{A_{k_n}^c})\right]
\end{aligned}$$

Now compute the second moment:

$$\mathbb{E}\left[Y_n^2\right] = \mathbb{E}\left[(P_n(\mathbb{1}_{A_{k_n}^c} f))^2\right]$$

$$\leq \mathbb{E}\left[(FP_n(\mathbb{1}_{A_{k_n}^c}))^2\right]$$

$$= F^2\mathbb{E}\left[(P_n(\mathbb{1}_{A_{k_n}^c}))^2\right]$$

$$\leq F^2\mathbb{E}\left[p_n(\mathbb{1}_{A_{k_n}^c}))\right].$$

So to show that $\mathrm{Var}(Y_n^2) \to 0$ it is enough to show that $\mathbb{E}\left[p_n(\mathbb{1}_{A_{k_n}^c}))\right] \to 0$. We are going to construct a sequence k_n which does this and such that $\mathrm{Var}(X_n) \to 0$ as well. Define $z_{n,k} = (\mathbb{E}\left[P_n(\mathbb{1}_{A_{k_n}} f)\right], \mathbb{E}\left[p_n(\mathbb{1}_{A_{k_n}^c}))\right])$ a sequence in $\mathbb{R}^2$. Note that $\lim_{n\to\infty} z_{n,k} = z_k = (\mathbb{E}\left[p_n(f\mathbb{1}_{A_{k_n}}))\right], \mathbb{E}\left[p_n(\mathbb{1}_{A_{k_n}^c}))\right])$, and $\lim_{k\to\infty} z_k = (0,0)$. Now if we apply Lemma 6.2.16 we get a sequence k_n such that $\mathbb{E}\left[p_n(\mathbb{1}_{A_{k_n}^c}))\right] \to 0$ and $\mathrm{Var}(P_n(\mathbb{1}_{A_{k_n}} f)) \to 0$. Thus $\mathrm{Var}(Y_n) \to 0$ and $\mathrm{Var}(X_n) \to 0$. This is exactly what is needed to conclude that $\mathrm{Var}(P_n(f)) \to 0$. $\qquad\square$

The next lemma allows us to bound the covariances appearing from the observations of the continuous part.

Lemma 6.2.69. *Let P_0 be a true distribution with $P_0 = P_d + \lambda P_c$ where P_d is a finite discrete subprobability distribution and P_c is a continuous probability distribution. Denote $C_n = \{i : P_0(\{\hat{X}_i\}) = 0\}$. Then for all $i, j \in C_n$*

$$\sum_{i,j\in C_n, i\neq j} |Cov(W_i, W_j | X_1, \cdots, X_n)| \to 0 \qquad P_0 \ a.s.$$

in the model where $X_1, \cdots, X_n | P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0,1]$ which admits a Lebesgue density f and $a, b > 0$ and g a twice continuously differentiable function on $[0,1]$ which is bounded away from zero such that $f(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$.

Proof. Because $i, j \in C_n$ we know that almost surely, $N_i = N_j = 1$. By exchangebility, Corollary 6.2.25, we know that

$$\sum_{i,j\in C_n, i\neq j} |\mathrm{Cov}(W_i, W_j | X_1, \cdots, X_n)|$$

is just
$$n(n-1)|\text{Cov}(W_i, W_j | X_1, \cdots, X_n)|$$

for $i \neq j, i, j \in C_n$. From Lemma 6.2.59 we know that $\tilde{\mathbb{E}}\left[W_{I_j} W_{I_i}\right] = \frac{\lambda^2 + o(1)}{n(n-1)}$ and from Lemma 6.2.57 we know that $\tilde{\mathbb{E}}\left[W_{I_i}\right] = \frac{\lambda + o(1)}{n}$. Now observe that $\text{Cov}(W_i, W_j | X_1, \cdots, X_n) = \frac{\lambda^2 + o(1)}{2n(n-1)} - \frac{1}{n}^2 (\lambda^2 + o(1))$. Hence

$$2n(n-1)|\text{Cov}(W_i, W_j | X_1, \cdots, X_n)| = |\lambda^2 + o(1) - \frac{n(n-1)}{n^2}(\lambda^2 + o(1))| \to 0.$$

Thus
$$\sum_{i,j \in C_n, i \neq j} |\text{Cov}(W_i, W_j | X_1, \cdots, X_n)| \to 0.$$

$\square$

We first recall a little result on the variance of bounded random variables.

Lemma 6.2.70. *If X_n is a sequence of random variables on $[0, F]$ such that $\mathbb{E}[X_n] \to 0$, then $Var(X_n) \to 0$.*

Proof. We know that $\mathbb{E}[X_n] 0$. Now we want to compute $\mathbb{E}[X_n^2]$. This is $\mathbb{E}[\mathbb{1}_{X_n < 1} X_n^2 + \mathbb{1}_{X_n \geq 1} X_n^2]$. We can split this using linearity. On $X_n < 1$ we know that $X_n^2 < X_n$ so $\mathbb{E}[\mathbb{1}_{X_n < 1} X_n^2] < \mathbb{E}[\mathbb{1}_{X_n < 1} X_n]$. Because X_n is non-negative, we know that $\mathbb{E}[\mathbb{1}_{X_n < 1} X_n] < \mathbb{E}[X_n] \to 0$. Hence the first term converges to zero. For the second term, observe that on $\mathbb{1}_{X_n \geq 1}$ we know that $X_n \geq 1$. Suppose there exists $c > 0$ such that $P(X_n \geq 1) > c$ for arbitrarily large n. Then $\mathbb{E}[\mathbb{1}X_n \geq 1 X_n] > c$ for those n. However, this means $\mathbb{E}[X_n] > c$ for those n, which is in contradiction with our assumptions. Hence $\mathbb{P}(X_n \geq 1) \to 0$. Because $X_n \leq F$ we can conclude $\mathbb{E}[\mathbb{1}_{X_n \geq 1} X_n^2] \leq F^2 \mathbb{E}[\mathbb{1}_{X_n \geq 1}] \to 0$. Hence the first and second moment of X_n converge to zero, and therefore the variance of X_n converges to zero. $\square$

The variance coming from the prior contributes zero to the variance.

Lemma 6.2.71. *Let P_0 be a distribution on $\mathcal{X}$ which does not have both infinitely many points whit positive probability and a continuous part. Then*

$$Var(\sum_{i:i \neq I_1, \cdots, I_{K_n}} f(\theta_i) W_i) \to 0.$$

in the model where $X_1, \cdots, X_n | P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0, 1]$ which admits a Lebesgue density f and $a, b > 0$ and g a twice continuously differentiable function on $[0, 1]$ which is bounded away from zero such that $f(v) = v^{a-1}(1 - v)^{b-1}g(v)$ for all $v \in (0, 1)$.

Proof. From Lemma 6.2.64 we know that the first moment of $\sum_{i:i \neq I_1, \cdots, I_{K_n}} f(\theta_i)W_i$ conditional on the observations is asymptotically zero. Because f is bounded, we know that

$$\sum_{i:i \neq I_1, \cdots, I_{K_n}} f(\theta_i)W_i$$

is a bounded random variable. Hence we can apply Lemma 6.2.70 to conclude that the variance converges to zero as well.   □

This lemma allows us to show that the variance of the integral on the discrete part converges to zero.

Lemma 6.2.72. *Let P_0 be a distribution on $\mathcal{X}$ which does not have both infinitely many points whit positive probability and a continuous part. Let $A = \{\mu_1, \cdots, \mu_m\}$ be a subset of $\mathcal{X}$. Let f be a nonnegative bounded measurable function on $\mathcal{X}$. Then*

$$Var(P(\mathbb{1}_A f)|X_1, \cdots, X_n) \to 0 \qquad P_0 \ a.s.$$

in the model where $X_1, \cdots, X_n | P \sim P$ and $P \sim SBP(\alpha, \mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0, 1]$ which admits a Lebesgue density f and $a, b > 0$ and g a twice continuously differentiable function on $[0, 1]$ which is bounded away from zero such that $f(v) = v^{a-1}(1 - v)^{b-1}g(v)$ for all $v \in (0, 1)$.

Proof. From the posterior predictive model we find

$$\text{Var}(P(\mathbb{1}_A f)|X_1, \cdots, X_n) = \text{Var}(\sum_{i:\hat{X}_i \in A} f(\hat{X}_i)W_{I_i}$$

$$+ \sum_{i:i \neq I_1, \cdots, I_{K_n}} f(\theta_i)W_i | X_1, \cdots, X_n)$$

This is the conditional variance of a sum of two random variables:

$$\sum_{i:\hat{X}_i \in A} f(\hat{X}_i)W_{I_i}$$

and

$$\sum_{i:i\neq I_1,\cdots,I_{K_n}} f(\theta_i)W_i.$$

From Lemma 6.2.71 we know that the second term vanishes asymptotically. In view of Lemma 6.2.70 it is enough to show that the variance of the first term vanishes asymptotically as well. Note that this term is again a finite sum. So we compute the conditional variance of every term. This is

$$\mathrm{Var}(f(\hat{X}_i)W_{I_i}|X_1,\cdots,X_n) = f(\hat{X}_i)^2\mathrm{Var}(W_{I_i}|X_1,\cdots,X_n).$$

The variance of this vanishes by Lemma 6.2.33, hence again by Lemma 6.2.70 we see that the variance of

$$\sum_{i:\hat{X}_i\in A} f(\hat{X}_i)W_{I_i}$$

converges to zero as well. This shows that

$$\mathrm{Var}(P(\mathbb{1}_A f)|X_1,\cdots,X_n) \to 0 \qquad P_0 \text{ a.s.}$$

$\square$

The variance coming from the observations from the continuous part of the true distributions have zero variance:

Lemma 6.2.73. *Let P_0 be a distribution on $\mathcal{X}$ which does not have both infinitely many points with positive probability and a continuous part. Let $A = \{\mu \in \mathcal{X} : P_0(\{\mu\}) > 0\}$ be the collection of points of positive probability. Let f be a nonnegative bounded measurable function on $\mathcal{X}$. Then*

$$Var(P(\mathbb{1}_{A^c}f)|X_1,\cdots,X_n) \to 0 \qquad P_0 \text{ a.s.}$$

in the model where $X_1,\cdots,X_n|P \sim P$ and $P \sim SBP(\alpha,\mathcal{D})$ where α is an atomless base measure and $\mathcal{D}$ is a distribution on $[0,1]$ which admits a Lebesgue density f and $a,b > 0$ and g a twice continuously differentiable function on $[0,1]$ which is bounded away from zero such that $f(v) = v^{a-1}(1-v)^{b-1}g(v)$ for all $v \in (0,1)$.

Proof. The predictive posterior distribution gives that

$$\text{Var}(P(\mathbb{1}_{A^c}f|X_1,\cdots,X_n) = \text{Var}(\sum_{i:\hat{X}_i\notin A}f(\hat{X}_i)W_{I_i}$$

$$+\sum_{i:i\neq I_1,\cdots,I_{K_n}}f(\theta_i)W_i|X_1,\cdots,X_n)$$

Again this is conditional variance of a sum of two random variables

$$\sum_{i:\hat{X}_i\notin A}f(\hat{X}_i)W_{I_i}$$

and

$$\sum_{i:i\neq I_1,\cdots,I_{K_n}}f(\theta_i)W_i.$$

The conditional variance of the second term vanishes asymptotically, so we only have to study the first.

$$\text{Var}(\sum_{i:\hat{X}_i\notin A}f(\hat{X}_i)W_{I_i}|X_1,\cdots,X_n) = \sum_{i=1}^{\infty}f(\hat{X}_i)^2\text{Var}(W_{I_i}|X_1,\cdots,X_n)$$

$$+\sum_{i\neq j}f(\hat{X}_i)f(\hat{X}_j)\text{Cov}(W_{I_i},W_{I_j})$$

Now we want to find an upper bound, so we estimate this from above by taking the absolute values of the covariances and the maximal value f can attain, say F. This gives

$$\text{Var}(\sum_{i:\hat{X}_i\notin A}f(\hat{X}_i)W_{I_i}|X_1,\cdots,X_n) \le \sum_{i=1}^{\infty}F^2\text{Var}(W_{I_i}|X_1,\cdots,X_n)$$

$$+F^2\sum_{i\neq j}|\text{Cov}(W_{I_i},W_{I_j})|$$

Now because the sum of the absolute values of these covariances converges to zero, we can study the limiting behavior of this bound by studying

$$\text{Var}(\sum_{i:\hat{X}_i\notin A}W_{I_i}|X_1,\cdots,X_n) = \sum_{i=1}^{\infty}\text{Var}(W_{I_i}|X_1,\cdots,X_n)$$

$$+\sum_{i\neq j}\text{Cov}(W_{I_i},W_{I_j})$$

And we have shown in Lemma 6.2.61 that this converges to zero. Hence

$$\text{Var}(\sum_{i:\hat{X}_i \notin A} f(\hat{X}_i)W_{I_i}|X_1, \cdots, X_n)$$

converges P_0 almost surely to zero, which is what was needed to show that

$$\text{Var}(P(\mathbb{1}_{A^c}f)|X_1, \cdots, X_n) \to 0 \qquad P_0 \text{ a.s.}$$

$\square$

The variance converges to zero.

Lemma 6.2.74. *Suppose P_0 is a distribution with does not have both infinitely many points of positive probability and a continuous part. Suppose that the base measure α is atomless and the relative stick-breaking weight distribution $\mathcal{D}$ admits a density f and constants $a, b > 0$ and a twice continuously differentiable function $g : [0, 1] \to \mathbb{R}$ which is bounded away from zero such that $f(v) = v^{a-1}(1-v)^{b-1}g(v)$ for $v \in (0, 1)$. Let h be a non-negative bounded measurable function. If $P \sim SBP(\alpha, \mathcal{D})$ then*

$$Var(P(f)|X_1, \cdots, X_n) \to 0 \qquad P_0 a.s.$$

Proof. Let P_0 be a distribution which does not have both infinitely many points of positive probability and a continuous part. Denote $A = \{\mu \in \mathcal{X} : P_0(\{\mu\} > 0\}$ the set of points with positive probability. For every $N \in \mathbb{N}$ we pick a finite set $A_k = \{\mu_1, \cdots, \mu_m\}$ such that $P_0(A_k \cup A^c) > 1 - \frac{1}{k}$. Then $P_0(A_k \cup A^c) \to 1$. Let f be a nonnegative bounded measurable function. Then by Lemma 6.2.72 we know

$$\text{Var}\left(P(\mathbb{1}_{A_k}f)|X_1, \cdots, X_n\right) \to 0 \qquad P_0 \text{ a.s.}$$

Similarly from Lemma 6.2.73

$$\text{Var}\left(P(\mathbb{1}_{A^c}f)|X_1, \cdots, X_n\right) \to 0 \qquad P_0 \text{ a.s.}$$

Thus by Lemma 6.2.15 , and using that A_k and A^C are disjoint, it follows that

$$\text{Var}\left(P(\mathbb{1}_{A_k \cup A^c}f)|X_1, \cdots, X_n\right) \to 0 \qquad P_0 \text{ a.s.}$$

This holds for every k thus we can apply Lemma 6.2.68 to conclude that

$$\text{Var}\left(P(f)|X_1, \cdots, X_n\right) \to 0 \qquad P_0 \text{ a.s.}$$

$\square$

6.2.11   General statement

Here we prove the general statement. It is just a collection of the results so far, with a bit of work to combine the last two sections into a general statement.

Theorem 6.2.75 (The stick-breaking process is consistent under regularity conditions). *Let α be an atomless measure. Suppose that the stick-breaking distribution $\mathcal{D}$ admits a density f such that there exists constants $a, b > 0$ and a twice continuously differentiable function g on $[0, 1]$ such that $f(v) = v^{a-1}(1-v)^{b-1}g(v)$ for $v \in (0, 1)$. Let P_0 be any distribution such that P_0 does not both have infinitely many points of positive probability and a continuous part. Then the posterior in the model $X_1, \cdots, X_n | P \sim P$, where $P \sim SBP(\alpha, \mathcal{D})$, as given in Theorem 5.4.4 is consistent with respect to the topology of pointwise convergence on bounded measurable functions at P_0.*

Proof. The proof consists of case checking and some computations. The main idea is to apply Lemma 6.2.2. So fix f a bounded measurable function. Now we want to show that $\mathbb{E}\left[P(f)|X_1, \cdots, X_n\right] \to P_0(f)$ P_0 almost surely and $\mathrm{Var}\left(P(f)|X_1, \cdots, X_n\right) \to 0$ P_0 almost surely.

By Lemma 6.2.67 we know that $\mathbb{E}\left[P(f)|X_1, \cdots, X_n\right] \to P_0(f)$ P_0 almost surely and by Lemma 6.2.74 we know that $\mathrm{Var}\left(P(f)|X_1, \cdots, X_n\right) \to 0$ P_0 almost surely. This shows consistency. $\square$

If we also assume the conjecture, then we can extend the results further to the following theorem:

Conjecture 6.2.76 (The stick-breaking process is consistent under regularity conditions). *Let α be an atomless measure. Suppose that the stick-breaking distribution $\mathcal{D}$ admits a density f such that there exists constants $a, b > 0$ and a twice continuously differentiable function g on $[0, 1]$ such that $f(v) = v^{a-1}(1-v)^{b-1}g(v)$ for $v \in (0, 1)$. Let P_0 be any distribution. Then the posterior in the model $X_1, \cdots, X_n | P \sim P$, where $P \sim SBP(\alpha, \mathcal{D})$, as given in Theorem 5.4.4 is consistent with respect to the topology of pointwise convergence on bounded measurable functions at P_0.*

The proof is mostly the same, except we also need to refer to the convergence of the first and second moment if there are infinitely many points of positive probability and a continuous part. The same arguments would apply except we need a new lemma to show that the bad event only occurs

with asymptotic probability zero. Currently we have shown a proof that the weights corresponding to observations from the continuous part in the tail converge almost surely to the right value if there are finitely many points of positive probability in the true distribution. However, these proofs can be adapted by showing that if you fix d points $\mu_1, \cdots, \mu_d$ of positive probability under the true distribution, and one looks at the weights in the tail, i.e. behind those, then one can bound the influence of the rest by considering upper and lower bounds. The rest of the arguments do not need to be changed except that we now need to refer to the more general lemma.

6.3 Consistency of Mixtures

We will first quote a theorem which we will state without proof, then we quote a theorem on the consistency of Dirichlet Processes, and then we show how to modify the proof to work in a more general situation.

We first need some introductory notation. d_H is the *Hellinger metric*, $N(\epsilon, \mathcal{P}_n, d)$ is the *covering number* of $\mathcal{P}_n$ by ϵ sized d balls. For more information on covering numbers, we refer to $[1, Appendix C]$. This allows us to formulate the consistency by *entropy* Theorem, as stated in $[1, theorem 6.23]$. Finally we denote $KL(\Pi)$ to be the set of all densities which admit the *Kullback Leibler* property. The Kullback-Leibler property states that every ϵ ball in the space of densities around a density p_0 with respect to the *Kullback Leibler divergence* $d_{\mathrm{KL}}(p, p_0)$ gets positive prior mass. The Kullback Leibler divergence $d_{\mathrm{KL}}(p, p_0) = \int \log \frac{p(x)}{p_0(x)} p(x) \, \mathrm{d}\, x$, which is the *relative entropy* of p with respect to p_0. We will later study how big the set of densities with the Kullback Leibler property is.

Theorem 6.3.1. *Given a distance d that generates convex balls and satisfies* $d(p_0, p) \leq d_H(p_0, p)$ *for every p, suppose that for every* $\epsilon > 0$ *there exists partitions* $\mathbb{P} = \mathcal{P}_{n,1} \cup \mathcal{P}_{n,2}$ *of the parameter space and a constant* $C > 0$, *such that, for sufficiently large n,*

- $\log N(\epsilon, \mathcal{P}_{n,1}, d) \leq n\epsilon^2$.

- $\Pi(\mathcal{P}_{n,2}) \leq e^{-Cn}$.

Then the posterior distribution in the model $X_1, \cdots X_n | p \overset{iid}{\sim} p$ *and* $p \sim \Pi$ *is strongly consistent relative to d at every* $p_0 \in KL(\Pi)$.

Using this theorem, we can prove the main theorem on the consistency of Dirichlet process mixtures. Again we first introduce some notation and then cite the theorem. Let $p_{F,\phi}(x) = \int \psi(x; \theta, \phi) \, d\, F(\theta)$ for a given family of probability densities $x \mapsto \phi(x; \theta, \phi)$, indexed by the two parameters $\theta \in \Theta \subset \mathbb{R}^k$ and $\phi \in \Phi \subset \mathbb{R}^l$. Equip F with the Dirichlet process prior and ϕ by some other prior π. This yields [1, theorem 7.15].

Theorem 6.3.2. *If for any given $\epsilon > 0$ and n, there exists subsets $\Theta_n \subset \mathbb{R}^k$ and $\Phi_n \subset \mathbb{R}^l$ and constants $a_n, A_n, b_n, B_n > 0$ such that*

- $||\psi(\cdot; \theta, \phi) - \psi(\cdot; \theta', \phi')||_1 \leq a_n ||\theta - \theta'|| + b_n ||\phi - \phi'||$, *for all $\theta, \theta' \in \Theta_n$ and $\phi, \phi' \in \Phi_n$.*

- $Diam(\Theta_n) \leq A_n$ *and* $Diam(\Phi_n) \leq B_n$.

- $\log(a_n A_n) \leq c \log n$ *for some $C > 0$ and $\log(b_n B_n) \leq \frac{n\epsilon^2}{8l}$.*

- $\max(\alpha(\Theta_n^c), \pi(\Phi_n^c)) \leq e^{-Cn}$ *for some $C > 0$.*

Then the posterior distribution $\Pi_n(\cdot | X_1, \cdots, X_n)$ for $p_{F,\phi}$ in the model $X_1, \cdots, X_n | (F, \phi) \overset{iid}{\sim} p_{F,\phi}$, for $(F, \phi) \sim DP(M\alpha) \times \pi$, is strongly consistent relative to the total variation norm at every p_0 in the Kullback-Leibler support of the prior of $p_{F,\phi}$.

Proof. We want to apply the consistency by entropy theorem Theorem 6.3.1, with d equal to the $\mathbb{L}_1$-distance divided by 2. This satisfies $d(p_0, p) \leq d_H(p_0, p)$. For given $\epsilon > 0$, we set $N_n \sim \frac{n\delta}{\log n}$ and δ small enough, to be determined later. Then we define

$$\mathcal{P}_{n,1} = \{\sum_{j=1}^{\infty} w_j \phi(\cdot; \theta_j, \phi) : (w_j) \in \mathbb{S}_\infty, \sum_{j \geq N} w_j \leq \frac{\epsilon}{8}, \theta_1, \cdots, \theta_N \in \theta_n, \phi \in \Phi_n\}.$$

Then the prior density $p_{F,\phi}$ is contained in $\mathcal{P}_{n,1}$, unless $\sum_{j \geq N_n} w_j \geq \frac{\epsilon}{8}$, or at least one of $\theta_1, \cdots, \theta_{N_n} \overset{iid}{\sim} \alpha$ fall outside Θ_n, or $\phi \notin \Phi_n$. It follows that

$$\Pi(\mathcal{P}_{n,1}^c) \leq P(\sum_{j > N_n} W_j \geq \frac{\epsilon}{8}) + N_n \alpha(\Theta_n^c) + \pi(\Phi_n^c).$$

The last two terms are exponentially small by assumption and the choice of N. We delegate the proof that $P(\sum_{j > N_n} W_j \geq \frac{\epsilon}{8})$ is exponentially small to the end.

Now we give a bound for the ϵ covering number of $\mathcal{P}_{n,1}$. The functions of the form $\sum_{j=1}^{N} w_j \psi(\cdot; \theta_j, \phi)$ with $(w_1, \cdots, w_{N_n}) \in \mathbb{S}_N$ form an $\frac{\epsilon}{4}$-net over $\mathcal{P}_{n,1}$ for the $\mathbb{L}_1$ norm. To construct an $\frac{3\epsilon}{4}$-net over these finite sums we restrict $(w_1, \cdots, w_{N_n})$ to an $\frac{\epsilon}{4}$-net over $\mathbb{S}_{N_n}$, restrict $(\theta_1, \cdots, \theta_{N_n})$ to an $\frac{\epsilon}{4a_n}$-net over Θ_n and ϕ to an $\frac{\epsilon}{4b_n}$-net over Φ_n. The cardinality of such a net is bounded above by

$$(\frac{20}{\epsilon})^{N_n}(\frac{12A_n a_n}{\epsilon})^{kN_n}(\frac{12B_n b_n}{\epsilon})^l.$$

Taking logs and rewriting yields $\log(N(\epsilon, \mathcal{P}_{n,1}, || \cdot ||_1) \leq n\epsilon^2$ as required, for δ small enough.

Now we tackle the problem of showing that $P(\sum_{j>N_n} W_j \geq \frac{\epsilon}{8})$ is exponentially small. The stick-breaking weights satisfy $W_j = V_j \prod_{s=1}^{j-1}(1-V_s)$, for $V_s \overset{iid}{\sim} \text{Be}(1, \alpha)$, and $sum_{j>N_n} w_j = \prod_{j=1}^{N_n}(1-V_j)$. Since $-\log(1-V_l)$ possesses an exponential distribution, $R_n := -\log \sum_{j>N_n} W_j$ is gamma distributed with parameters N_n and M, and hence $P(R_n \leq r) \leq \frac{(Mr)^{N_n}}{N_n!}$. Therefore, the first term is bounded above by $(\frac{eM\log(\frac{8}{\epsilon})}{N_n})^{N_n}$, which is exponentially small, by the choice of N_n, for any $\delta > 0$. $\qquad\square$

We will generalize this theorem to the general case of stick-breaking weights, not just the $\beta(1, M)$ case.

Theorem 6.3.3. *Let Π be a stick-breaking process with base measure α and stick-breaking weights distributed according to $\mathcal{D}$. Suppose that either D admits a bounded Lebesgue density or $\mathcal{D}$ is a discrete measure which does not put mass arbitrarily close to 0. If for any given $\epsilon > 0$ and n, there exists subsets $\Theta_n \subset \mathbb{R}^k$ and $\Phi_n \subset \mathbb{R}^l$ and constants $a_n, A_n, b_n, B_n > 0$ such that*

- *$||\psi(\cdot; \theta, \phi) - \psi(\cdot; \theta', \phi')||_1 \leq a_n ||\theta - \theta'|| + b_n ||\phi - \phi'||$, for all $\theta, \theta' \in \Theta_n$ and $\phi, \phi' \in \Phi_n$.*

- *$Diam(\Theta_n) \leq A_n$ and $Diam(\Phi_n) \leq B_n$.*

- *$\log(a_n A_n) \leq c \log n$ for some $C > 0$ and $\log(b_n B_n) \leq \frac{n\epsilon^2}{8l}$.*

- *$\max(\alpha(\Theta_n^c), \pi(\Phi_n^c)) \leq e^{-Cn}$ for some $C > 0$.*

Then the posterior distribution $\Pi_n(\cdot | X_1, \cdots, X_n)$ for $p_{F,\phi}$ in the model where $X_1, \cdots, X_n$ conditional on (F, ϕ) is iid from $p_{F,\phi}$, for $(F, \phi) \sim \Pi \times \pi$, is strongly consistent relative to the total variation norm at every p_0 in the Kullback-Leibler support of the prior of $p_{F,\phi}$.

In order to generalize the theorem, we observe that we only need to change the proof to get exponentially small decay, and we can pick $N_n = \lceil \frac{n\delta}{\log n} \rceil$. The next lemma gives us a way to find exponential decay.

We use the notation from the proof of theorem 7.15, except we work with the general process, so $V_l \sim \mathcal{D}$ for some distribution D.

Lemma 6.3.4. *Suppose there exists a $C > 0$ and a sequence $k(n)$ which solves $\mathbb{E}(1 - V_1)^{k(n)} \leq e^{\log(n)\left(\frac{\log \lambda k(n) - cn}{n\delta} \right)}$ for all natural numbers $n > m$. Then $P(\prod_{i=1}^{N_n} V_i \geq \lambda) \leq e^{-Cn}$ for all $n \geq m$.*

Proof. We apply the Markov inequality. We often work with $N_n = \frac{n\delta}{\log n}$, which is smaller, but at most 1 smaller.

$$
P(\prod_{i=1}^{N_n}(1 - V_i) \leq \lambda) \leq \frac{\mathbb{E}(\prod_{i=1}^{N_n}(1 - V_i))^{k(n)}}{\lambda^{k(n)}}
$$

$$
\overset{iid}{=} \frac{\prod_{i=1}^{N_n} \mathbb{E}(1 - V_i)^{k(n)}}{\lambda^{k(n)}}
$$

$$
= \frac{\left(\mathbb{E}\left[(1 - V_i)^{k(n)}\right] \right)^{N_n}}{\lambda^{k(n)}}
$$

$$
\overset{assumption}{\leq} \frac{\left(e^{\log(n)\left(\frac{\log \lambda k(n) - cn}{n\delta} \right)} \right)^{N_n}}{\lambda^{k(n)}}
$$

$$
\leq \frac{e^{\log(n)\left(\frac{\log \lambda k(n) - cn}{n\delta} \right) \frac{n\delta}{\log n}}}{\lambda^{k(n)}}
$$

$$
= \frac{e^{\left(\frac{\log \lambda k(n) - cn}{n\delta} \right) n\delta}}{\lambda^{k(n)}}
$$

$$
= \frac{e^{\log \lambda k(n) - cn}}{\lambda^{k(n)}}
$$

$$
= \frac{e^{\log \lambda k(n)} e^{-cn}}{\lambda^{k(n)}}
$$

$$
= \frac{\lambda^{k(n)} e^{-cn}}{\lambda^{k(n)}}
$$

$$
= e^{-cn}.
$$

So indeed, we have exponential decreasing probability.   □

We observe that we can rewrite the bound in the following ways:

$$e^{\log(n)\left(\frac{\log \lambda k(n) - cn}{n\delta}\right)} = n^{-\frac{c}{\delta}} \lambda^{\frac{k(n)\log n}{\delta n}}$$

$$= n^{\frac{\log \lambda k(n) - cn}{n\delta}}$$

$$= n^{\frac{\log \lambda}{\delta} \frac{k(n)}{n} - \frac{c}{\delta}}$$

We will use the first equality now:

Next question is to actually show that there are distributions which solve this decay. Read X as $1 - V$.

Lemma 6.3.5. *Suppose that* $EX^k \leq \frac{M}{k}$. *Then we have exponential decay*

$$P(\prod_{i=1}^{N_n} X_i \geq \lambda) \leq e^{-cn}$$

Proof. We want to show that for a fixed we can find a $C > 0$ and a $k : \mathbb{N} \to \mathbb{N}$, where k can depend on λ and δ and C can depend on δ.

$$\frac{M}{k} \leq n^{-\frac{c}{\delta}} \lambda^{\frac{k \log n}{\delta n}}$$

This is equivalent with (by multiplying left and right by $kn^{\frac{C}{\delta}}$)

$$Mn^{\frac{c}{\delta}} \leq k\lambda^{\frac{k \log n}{\delta n}}$$

We can take logarithms on both sides to get

$$\frac{c}{\delta} \log n + M \leq \log k + \frac{k \log n}{\delta n} \log \lambda$$

If we pick $k = \frac{-\delta n}{\log n \log \lambda}$. Filling this out into the equation gives

$$\frac{c}{\delta} \log n + M \leq \log(\frac{-\delta n}{\log n \log \lambda}) + \frac{\left(\frac{-\delta n}{\log n \log \lambda}\right) \log n}{\delta n} \log \lambda$$

We can simplify this into

$$\frac{c}{\delta} \log n + M \leq \log n + \log \delta - \log(-\log n \log \lambda) - 1$$

which holds when $C < \delta$ for large enough n. $\square$

Lemma 6.3.6. *Let V have an almost surely bounded density f with respect to the Lebesgue measure, then V has exponential decay:*

Proof. we estimate the $k - th$ moment of $1 - V$ by

$$
\begin{aligned}
\mathbb{E} X^k &= \int_0^1 x^k f(1-x) \, \mathrm{d} x \\
&\leq \int_0^1 \sup_{y \in [0,1]} f(y) \, \mathrm{d} x \\
&= \sup_{y \in [0,1]} f(y) \int_0^1 x^k \, \mathrm{d} x \\
&= \sup_{y \in [0,1]} f(y) \frac{1}{k+1} \\
&= O(\frac{1}{k})
\end{aligned}
$$

Now we can apply the second lemma to gain what we want. $\qquad \square$

Corollary 6.3.7. *Let V have a continuous density f. Then we have exponential decay.*

Proof. Because $[0,1]$ is compact, f attains a maximum, and it is positive so it bounded. Now apply previous lemma. $\qquad \square$

Note that these bounds are not the sharpest they can get, in the sense that there is most likely a bigger collection of probability measures which would lead to a consistent posterior. For example, densities which blow up at 1 or 0 might still be fine, and densities which have decay like $\frac{1}{\log n}$ are probably the tightest we can get. Furthermore, we would like to point out that most discrete distributes are fine. As long as they do not put mass in every neighborhood around zero, since, for a given λ, the product of $\prod_{i=1}^n (1 - V_i)$ is smaller than λ. let m be smallest value V_i can attain, and set $N = \frac{\log \lambda}{\log(1-m)}$. Then almost surely, for enough observations, i.e. $N_n \geq N$ we have that the probability that the product exceeds λ is zero, so we indeed have exponential decay. This works for all λ so we also get consistency there. We like to point out that discrete distributions V which do not put mass near 0 have k-th moments of $1 - V$ which decreases like $(1 - m)^k$.

Further, note that attaining zero with positive probability just leads to throwing away the δ_{X_i} terms where the V_i was zero, so we can also just sum

over all the weights where we did not attain the zero. Those weights are distributed according to the reweighed distribution where we just removed zero, which leads to consistency. The only problem is if V_i can attain values arbitrarily close to 0. In this case, the simple argument fails and you need to find another method of checking consistency.

This shows that if we use numerical simulations, we get the same consistency results as for the density case, because if we approximate densities on floats, we cannot get arbitrarily close to zero, so we actually are consistent.

In the end we have proven theorem Theorem 6.3.3.

Now the question is, how large is the space of densities in the Kullback Leibler support. For this there exists a general theoreom which gives what we want, namely [1, Theorem 7.3]. This states

Theorem 6.3.8. *Assume that*

- *χ is bounded, continuous and positive everywhere,*

- *$\int p_0(x) \log p_0(x)\, dx < \infty$.*

- *$-\int p_0(x) \log \inf_{||y||<\delta} p_0(x-y)\, dx < \infty$, for some $\delta > 0$.*

- *$\int_{||y||<||x||^\eta} \chi(x-y) \geq \underline{\chi}(x)$ for large $||x||$ and a function $\underline{\chi}$ that is decreasing as its arguments moves away from zero and satisfies*

$$-\int p_0(x) \log \underline{\chi}(2x|x|^\eta(\,dx < \infty,$$

for some $\eta \in (0,1)$.

Then $p_0 \in KL(\Pi)$ for the prior Π on $p_f = \int h^{-d}\chi((\cdot - \mu)/h)\, df(\mu, h)$ induced by a prior on F with full support on $\mathcal{M}(\mathbb{R}^d \times (0, \infty))$; and also for the prior on $p_{F,h} = \int h^{-d}\chi((\cdot - \mu)/hu)\, dF(\mu)$ induced by a product prior on (F, h) with full support on $\mathcal{M}(\mathbb{R}^d) \times (0, \infty)$.

In this case the second claim is relevant for stick-breaking mixtures. Here we see that if the stick-breaking weights are distributed according to a distribution fully supported on $[0, 1]$ we can apply this theorem.

Chapter 7

Discussion and future research

7.1 Discussion

In the computation of the posterior, we assumed that the base measure α was atomless. If not, then if you have twice the same observation, you do not know if they come from the same θ_i or not. One can try to correct for this. However, the description would become pretty complicated very quickly, so we choose to restrict ourselves to easier to handle models. In the chapter of consistency, we restricted ourselves to the case where the relative stick-breaking distributions are all the same. This has the advantage that the expression we encounter all become a lot easier to handle. Moreover, as Theorem 6.2.4 on the posterior consistency of the Pitman-Yor process shows, we cannot expect to get a theorem that states that every choice of prior will be consistent for every true distribution. We also restricted to distributions of the relative stick-breaking weights which admitted a density f with a particular nice form, namely it could be written as $v^{\alpha-1}(1-v)^{\beta-1}g(v)$ with g continuous and bounded away from zero. Both the fact that it is continuous and that it is bounded away from zero are probably needed. If one lets go of the continuity requirement, one can consider $f \propto \frac{1}{(x-\frac{1}{2})^2}$. However, then the argument we used to control some errors would not work, and they can grow beyond any bound. Similarly, if we consider the function given by $f(v) = 3\mathbb{1}_{v<\frac{1}{3}}$, then the posterior likely will fail to be consistent for the Bernoulli random variables. Thus we will need both arguments. One objection against this counterexample might be that one could also consider functions with zeros of finite order. In that case the constructed proof might

still work. However, we need to replace some approximation arguments.

7.2 Possible future research

In the chapter on consistency, we only considered priors in which all the relative stick-breaking distributions where the same. For consistency in the mixture case this is clearly not needed, as we can inspect from one of the first conditions given in the proofs. It is actually enough if the expected values of $\mathbb{E}[(1 - V_j)]$ are decreasing quickly enough, where quick enough is something like $e^{W(n)-W(n+1)}$ where W is the Lambert W function (inverse map of $x \mapsto x \log x$). There is room for improvement here, as this is only based on considering the first moments of $(1 - V_j)$. Note as well that in case everything has the same distribution the moments of $(1 - V_j)$ will not be decreasing but constant, so there is a clear gap in the requirements. It would be interesting to note if this gap is just an artifact of the proof, or if there is a real fundamental thing going on when all the relative stick-breaking distributions are the same.

Something we alluded to in the text, a Bernstein-von Mises like theorem related theorem will probably exists. Bernstein-von Mises theorems state some form of asymptotic normality of the posterior distribution, and the tools to research this have been already obtained. With Lemma 6.2.9 and the form of the moments of the posterior we can start computing this. The only possible problem would be that the $\sqrt{n}$ factor blows up some errors which possibly do not converge to zero when blowing up by $\sqrt{n}$. We are pretty sure this will not happen, as both Lemma 6.2.13 and Lemma 6.2.12 give strong indications that the convergence error terms appearing naturally have bounded influence. If we look at the general case instead of $\beta(a, b)$ random variables, we also need to approximate the ratios with Lemma A.2, however, again we have a strong form of convergence. These arguments are probably strong enough to give the convergence of all the moments to the standard normal moment.

People are not just interested in consistency. Another concept people are interested in is contraction rates, which, roughly, give how quick you converge to the true distribution. Recall the definition of consistency Definition 6.1.1

Definition 7.2.1. *The posterior distribution $\Pi_n(\cdot|X^{(n)})$ is said to be **(weakly) consistent** at $\theta_0 \in \Theta$ if, for all open neighborhoods U of θ_0, $\Pi_n(U^c|X^{(n)}) \to 0$*

*in $\mathbb{P}_{\theta_0}^n$ probability, as $n \to \infty$. The posterior is said to be **strongly consistent** at $\theta_0 \in \Theta$ if this convergence is in the almost-sure sense.*

Roughly speaking, a *contraction rate* of a posterior measures how quick the posterior converges to the right hypothesis. A contraction rate is an upper bound, so they are not unique. A contraction rate is a strengthening of the result of consistency.

Definition 7.2.2. *A sequence ϵ_n is a **posterior contraction rate** at the parameter θ_0 with respect to the semimetric d if $\Pi_n \left(\theta : d(\theta, \theta_0) \geq M_n \epsilon_n | X^{(n)} \right) \to 0$ in $P_{\theta_0}^{(n)}$ probability, for every $M_n \to \infty$. If all experiments share the same probability space and the convergence to zero takes place almost surely $[P_{\theta_0}^{(\infty)}]$, then ϵ_n is said to be a posterior contraction rate in the strong sense.*

The Dirichlet process mixtures has optimal contraction rates up to logarithmic factors. If you look at the proof of this, most of the arguments carry naturally to the case of stick-breaking processes. In fact, in case of the proof of [1, Theorem 9.9], there is one lemma, [1, Lemma 9.14] which does not carry over directly to the general setting. So if one would find an alternative for this lemma you can immediately get results on contraction rates for stick-breaking processes.

As a fourth point, the computation of the posterior distribution can face some problems with numerical stability if one implements this naively. For practical results, a quick and stable algorithm would be very valuable. Even if one might be able to do draws from the posterior, a quick algorithm to approximate the moments of W_{I_k} would be valuable.

Appendices

Appendix A

Stirling approximation and bounds

Lemma A.1 (Stirling approximation 1). *Let V be a random variable with density f such that there exists $\alpha, \beta > 0$ and a continuous function g bounded away from zero such that $f(v) = v^{\alpha-1}(1-v)^{\beta-1}g(v)$. Then*

$$\lim_{x,y\to\infty} \frac{\mathbb{E}[V^x(1-V)^y]}{g(\frac{x+a}{x+y+a+b})Be(x+a, y+b)} = 1$$

Proof. Let $\epsilon > 0$. Because g is continuous and $[0,1]$ is compact, g has upper and lower bounds, call them L and l respectively. Because g is positive it is bounded between $0 < l$ and L. There exists an $\delta > 0$ such that $|g(v)-g(w)| < \frac{\epsilon}{3}$ for all v such that $|v - w| < \delta$. We can do this for all $w \in [0,1]$. This yields an open cover for the interval $[0,1]$ which is compact. Hence we can extract a finite subcover. Over this finite subcover, we can extract the largest diameter of the opens, which we will call D. We now look at a few steps needed for this estimation.

We first recall the variance of the $Be(x,y)$ distribution. This is $\frac{xy}{(x+y)^2(x+y+1)}$, and the expected value is $\mu = \frac{x}{x+y}$. Now we can apply Chebychev to compute the mass in an small area around the mean. This yields

$$\mathbb{P}(|X - \mu| > k) < \frac{\sigma^2}{k^2}.$$

Note that

$$\mathbb{P}(|X - \frac{x}{x+y}| > k) = \frac{\int_0^{\mu-k} v^{x-1}(1-v)^{y-1}dv + \int_{\mu+k}^1 v^{x-1}(1-v)^{y-1}\,\mathrm{d}\,v}{\int_0^1 v^{x-1}(1-v)^{y-1}\,\mathrm{d}\,v}.$$

Thus Chebyshevs inequality implies

$$\frac{\int_0^{\mu-k} v^{x-1}(1-v)^{y-1}dv + \int_{\mu+k}^1 v^{x-1}(1-v)^{y-1}\,d\,v}{\int_0^1 v^{x-1}(1-v)^{y-1}\,d\,v} < \frac{\sigma^2}{k^2}$$

and

$$1 - \frac{\sigma^2}{k^2} \leq \frac{\int_{\mu-k}^{\mu+k} v^{x-1}(1-v)^{y-1}\,d\,v}{\int_0^1 v^{x-1}(1-v)^{y-1}\,d\,v} \leq 1.$$

If we want to compute the integral $\int_0^1 v^{x+a-1}(1-v)^{y+a-1}g(v)\,d\,v$ we can split into the integral over $[0, \frac{x+a}{x+y+a+b} - D]$, $[\frac{x+a}{x+y+a+b} + D, 1]$ and the area between $(\frac{x+a}{x+y+a+b} - D, \frac{x+a}{x+y+a+b} + D)$.

Hence we can estimate the integrals of the first and second intervals between 0 and $L \int_I v^{x+a-1}(1-v)^{y+b-1}\,d\,v$, where I is the first or second interval. On the third interval we can estimate $g(v)$ by $g(\frac{x+a}{x+y+a+b}) \pm \frac{\epsilon}{3}$.

$$(g(\frac{x+a}{x+y+a+b} - \frac{\epsilon}{3}) \int_{\frac{x+a}{x+y+a+b}-D}^{\frac{x+a}{x+y+a+b}+D} v^{x+a-1}(1-v)^{y+b-1}\,d\,v$$

$$\leq \int_0^1 v^{x+a-1}(1-v)^{y+b-1}g(v)\,d\,v \leq$$

$$L(\int_0^{\frac{x+a}{x+y+a+b}-D} v^{x+a-1}(1-v)^{y+b-1}\,d\,v + \int_{\frac{x+a}{x+y+a+b}+D}^1 v^{x+a-1}(1-v)^{y+b-1}\,d\,v) +$$

$$(g(\frac{x+a}{x+y+a+b}) + \frac{\epsilon}{3}) \int_{\frac{x+a}{x+y+a+b}-D}^{\frac{x+a}{x+y+a+b}+D} v^{x+a-1}(1-v)^{y+b-1}\,d\,v$$

If we divide now by $g(\frac{x+a}{x+y+a+b}) \int_0^1 v^{x+a-1}(1-v)^{y+b-1} \, d\, v$ we get

$$\frac{(g(\frac{x+a}{x+y+a+b}) - \frac{\epsilon}{3}) \int_{\frac{x+a}{x+y+a+b}-D}^{\frac{x+a}{x+y+a+b}+D} v^{x+a-1}(1-v)^{y+b-1} \, d\, v}{g(\frac{x+a}{x+y+a+b}) \int_0^1 v^{x+a-1}(1-v)^{y+b-1} \, d\, v}$$

$$\leq \frac{\int_0^1 v^{x+a-1}(1-v)^{y+b-1} g(v) \, d\, v}{g(\frac{x+a}{x+y+a+b}) \int_0^1 v^{x+a-1}(1-v)^{y+b-1} \, d\, v} \leq$$

$$\frac{L(\int_0^{\frac{x+a}{x+y+a+b}-D} v^{x+a-1}(1-v)^{y+b-1} \, d\, v + \int_{\frac{x+a}{x+y+a+b}+D}^{1} v^{x+a-1}(1-v)^{y+b-1} \, d\, v)}{g(\frac{x+a}{x+y+a+b}) \int_0^1 v^{x+a-1}(1-v)^{y+b-1} \, d\, v} +$$

$$\frac{(g(\frac{x+a}{x+y+a+b}) + \frac{\epsilon}{3}) \int_{\frac{x+a}{x+y+a+b}-D}^{\frac{x+a}{x+y+a+b}+D} v^{x+a-1}(1-v)^{y+b-1}}{g(\frac{x+a}{x+y+a+b}) \int_0^1 v^{x+a-1}(1-v)^{y+b-1} \, d\, v} \, d\, v$$

Using Chebychev we can now bound every integral in here, namely

$$(1 - \frac{\sigma_{x,y}^2}{D^2}) \leq \frac{\int_{\frac{x+a}{x+y+a+b}-D}^{\frac{x+a}{x+y+a+b}+D} v^{x+a-1}(1-v)^{y+b-1} \, d\, v}{\int_0^1 v^{x+a-1}(1-v)^{y+b-1} \, d\, v} \leq 1$$

This means we can bound the terms of the $g(\frac{x+a}{x+y+a+b}) \pm \frac{\epsilon}{3}$. The first term can be estimated below by

$$(1 - \frac{\epsilon}{3l})(1 - \frac{\sigma_{x,y}^2}{D^2}).$$

The last term can be estimated above by

$$1 + \frac{\epsilon}{3l}.$$

The term with the bounds of L can be estimated above by

$$\frac{L\sigma_{x,y}^2}{D^2}.$$

This means we got to pick x, y so large that

$$|1 - (1 - \frac{\epsilon}{3l})(1 - \frac{\sigma_{x,y}^2}{D^2})| < \epsilon$$

and
$$|1 - (1 + \frac{\epsilon}{3l} + \frac{L\sigma_{x,y}^2}{D^2})| < \epsilon.$$
If we rewrite these expressions a bit we get that we are searching for a bound on x, y such that

$$\sigma_{x,y}^2 < D^2(1 - \frac{1-\epsilon}{1-\frac{\epsilon}{3l}}).$$

and
$$\sigma_{x,y}^2 < \frac{D^2}{L} \frac{(3l+1)\epsilon}{3l}$$
Now if $x >$ or $y > z$ then $\sigma_{x,y}^2 < \frac{1}{z}$. So pick

$$z = \max(\frac{L}{D^2} \frac{3l}{(3l+1)\epsilon}, \frac{1}{D^2(1 - \frac{1-\epsilon}{1-\frac{\epsilon}{3l}})}).$$

Then for all $\min(x, y) > z$ we have that indeed

$$|1 - \frac{\mathbb{E}[V^x(1-v)^y]}{g(\frac{x+a}{x+y+a+b}) \int_0^1 v^{x+a-1}(1-v)^{y+b-1} \, dv}| < \epsilon.$$

$\square$

We want a stronger version of Stirling approximation, where we control the error of estimating the two ratios. This shows up in the proof of the theorem which shows consistency in case there is an infinite discrete support.

Lemma A.2 (Stirling with control on error terms). *Let $k > 0$. Let V be a random variable with density f such that there exists $\alpha, \beta > 0$ and a twice continuous differentiable function g bounded away from zero such that $f(v) = v^{\alpha-1}(1-v)^{\beta-1}g(v)$. Then for all $x, y > 0$ there exists a real number $r_{s,y}$ such that*

$$\frac{\mathbb{E}[V^x(1-V)^{y+k}]}{\mathbb{E}[V^x(1-V)^y]} = (1 + r_{x,y,k}) \prod_{i=0}^{k-1} \frac{y+b+i}{x+y+a+b-1+i}$$

with for all positive sequences K_n with $\frac{K_n}{n} \to 0$ and every sequence of positive real numbers p_n converging to $q > 0$

$$\lim_{n \to \infty} \sup_{x \le n}(1 + |r_{x,np_n,k}|)^{K_n} = 1. \qquad \forall k$$

Proof. First observe that the variance of a $\beta(a,b)$ distributed random variable is $\frac{ab}{(a+b)^2(a+b+1)}$. Then we can estimate

$$\int_0^1 v^{x+a-1}(1-v)^{y+b-1}g(v)\,\mathrm{d}\,v$$

by applying Taylors formula to g around $\mu = \frac{x+a-1}{x+y+a+b-1}$. If we do this we get

$$g(v) = g(\mu) + g'(\mu)(v-\mu) + g''(\xi)(v-\mu)^2.$$

Because g'' is continuous, it has finite bounds $l \leq g''(v) \leq L$, so we can give bounds. If we integrate these we get

$$\int_0^1 v^{x+a-1}(1-v)^{y+b-1}(g(\mu) + g'(\mu)(v-\mu) + l(v-\mu)^2)\,\mathrm{d}\,v$$

$$\leq \int_0^1 v^{x+a-1}(1-v)^{y+b-1}g(v)\,\mathrm{d}\,v \leq$$

$$\int_0^1 v^{x+a-1}(1-v)^{y+b-1}(g(\mu) + g'(\mu)(v-\mu) + L(v-\mu)^2)\,\mathrm{d}\,v.$$

Recognizing the expressions for the variance and expectation of a $\mathrm{Beta}(a,b)$ distributed random variable, we see that this integral is

$$g(\mu)\int_0^1 v^{x+a-1}(1-v)^{y+b-1}\,\mathrm{d}\,v+$$

$$l\frac{(x+a)(y+b)}{(x+y+a+b)^2(a+b+x+y+1)}\int_0^1 v^{x+a-1}(1-v)^{y+b-1}\,\mathrm{d}\,v$$

$$\leq \int_0^1 v^{x+a-1}(1-v)^{y+b-1}g(v)\,\mathrm{d}\,v \leq$$

$$g(\mu)\int_0^1 v^{x+a-1}(1-v)^{y+b-1}\,\mathrm{d}\,v+$$

$$L\frac{(x+a)(y+b)}{(x+y+a+b)^2(a+b+x+y+1)}\int_0^1 v^{x+a-1}(1-v)^{y+b-1}\,\mathrm{d}\,v.$$

Now if $y \sim n$ and $x \leq n$ this error is of order at most $\frac{1}{n}$. Hence we can rewrite this as

$$(1+O(\frac{1}{n}))g(\frac{x+a-1}{y+b-1})\int_0^1 v^{x+a-1}(1-v)^{y+b-1}\,\mathrm{d}\,v.$$

Note that $(1 + O(\frac{1}{n}))^{K_n} \to 1$ for all $\frac{K_n}{n} \to 0$ by Lemma 6.2.11. Now compute

$$\frac{g(\frac{x+a-1}{x+y+a+b-1})}{g(\frac{x+a-1}{x+y+a+b-2})}.$$

This is by differentiability of g of order $\frac{1}{n}$ and hence we get a term of the form $(1 + O(\frac{1}{n}))$. Hence we can apply Lemma 6.2.11 to get that $\left(\frac{g(\frac{x+a-1}{x+y+a+b-1})}{g(\frac{x+a-1}{x+y+a+b-2})} \right)^{K_n}$ converges to 1. We can do this for both terms in

$$\frac{\mathbb{E}[V^x(1-V)^{y+k}]}{\mathbb{E}[V^x(1-V)^y]}.$$

This means that this converges to

$$\frac{\text{Beta}(x+a-1, y+b+k-1)}{\text{Beta}(x+a-1, y+b-1)} = \prod_{i=0}^{k-1} \frac{y+b+i}{x+y+a+b-1+i}$$

and

$$\lim_{n\to\infty} \sup_{x\le n} (1 + |r_{x,np_n,k}|)^{K_n} = 1. \qquad \forall k$$

$\square$

If we inspect the previous proof, we can conclude that the next lemma also holds:

Lemma A.3 (Stirling with control on error terms). *Let $k > 0$. Let V be a random variable with density f such that there exists $\alpha, \beta > 0$ and a twice continuous differentiable function g bounded away from zero such that $f(v) = v^{\alpha-1}(1-v)^{\beta-1}g(v)$. Then for all $x, y > 0$ there exists a real number $r_{s,y}$ such that*

$$\frac{\mathbb{E}[V^{x+k}(1-V)^y]}{\mathbb{E}[V^x(1-V)^y]} = (1 + r_{x,y,k}) \prod_{i=0}^{k-1} \frac{x+a+i}{x+y+a+b-1+i}$$

with for all positive sequences K_n with $\frac{K_n}{n} \to 0$ and every sequence of positive real numbers p_n converging to $q > 0$

$$\lim_{n\to\infty} \sup_{x\le n} (1 + |r_{x,np_n,k}|)^{K_n} = 1. \qquad \forall k$$

Lemma A.4 (Bounds on estimation)**.** *Suppose that V has a density f such that $f(v) = v^{\alpha-1}(1-v)^{\beta-1}g(v)$ for $v \in (0,1)$ and g a continuous function bounded away from 0, $\alpha, \beta > 0$. There exists $0 < c < C < \infty$ such that for all $s \geq 0$ and $j \geq 1$ we can bound the expectation $\mathbb{E}[V^s(1-V)^j]$ by*

$$\frac{c}{j^{s+\alpha}} \leq \mathbb{E}[V^s(1-V)^j] \leq \frac{C}{j^{s+\alpha}}.$$

Proof. We start by observing that g is bounded between l and L. This means directly that

$$l\int_0^1 v^{s+a-1}(1-v)^{j+b-1}\,\mathrm{d}\,v < \mathbb{E}[V^s(1-V)^k] < L\int_0^1 v^{s+a-1}(1-v)^{j+b-1}\,\mathrm{d}\,v.$$

So if we can find uniform bounds for $s \in [0, S]$ for $\int_0^1 v^{s+a-1}(1-v)^{j+b-1}\,\mathrm{d}\,v$ we are done. For this observe we can use the known expression for $\text{Beta}(s + a, j + b)$ in terms of the Γ function. This is

$$\text{Beta}(s+a, j+b) = \frac{\Gamma(s+a)\Gamma(j+b)}{\Gamma(s+a+j+b)}.$$

Here $\Gamma(s+a)$ is bounded away from zero and infinity for all $s \in [0, S]$, so we can can absorb the maximal and minimal contribution of this term into the constants. This leaves us to analyze

$$\frac{\Gamma(j+b)}{\Gamma(j+s+a+b)}.$$

For this we can apply the Stirling approximation for the Γ function. This yields an approximation with bounded error terms. We can again absorb these factors into the constants, so this yields

$$\frac{\Gamma(j+b)}{\Gamma(j+a+b+s)} \approx \frac{(\frac{j+b}{e})^{j+b}\sqrt{2\pi(j+b)}}{(\frac{s+a+j+b}{e})^{s+a+j+b}\sqrt{2\pi(j+a+sb)}}$$

$$= e^{a+s}\frac{1}{(1+\frac{s+a}{j+b})^{j+b}}\Big(\frac{j}{s+a+j+b}\Big)^{s+a}\Big(\frac{j+b}{s+a+j+b}\Big)^{\frac{1}{2}}\frac{1}{j^{s+a}}.$$

The first four terms are bounded away from zero and infinity, so we can again absorb these into the error terms. This yields that for all $S > 0$ there exists constants $c, C > 0$ such that for all $s \in [0, S]$, all $j > 0$ we have

$$\frac{c}{j^{s+a}} \leq \mathbb{E}[V^s(1-V)^j] \leq \frac{C}{j^{s+a}}.$$

$\square$

Lemma A.5. *With the same notation as the previous lemma*

$$\mathbb{E}[V^s(1-V)^j] = g(0)\frac{\Gamma(s+\alpha)}{j^{s+\alpha}}(1+r_s(j))$$

where $\sup_{0 \le s \le S} r_s(j) \to 0$ *as* $j \to \infty$, *for any* $S < \infty$.

Proof. The proof idea is the same as in Lemma A.1. Let $\epsilon > 0$. Pick $\delta > 0$ such that $|g(v) - g(0)| < \epsilon$ for all $\delta > 0$. Then we can estimate

$$\int_0^1 v^{x+a-1}(1-v)^{j+b-1}g(v)\,\mathrm{d}\,v$$

by

$$(g(0) - \epsilon)\int_0^\delta v^{x+a-1}(1-v)^{j+b-1}\,\mathrm{d}\,v$$

$$\le \int_0^1 v^{x+a-1}(1-v)^{j+b-1}g(v)\,\mathrm{d}\,v \le$$

$$(g(0) + \epsilon)\int_0^\delta v^{x+a-1}(1-v)^{j+b-1}\,\mathrm{d}\,v + L\int_\delta^1 v^{x+a-1}(1-v)^{j+b-1}\,\mathrm{d}\,v.$$

Picking j large enough so that the mean + standard deviation of a beta distributed random variable are less than δ, which is to state, we pick j large enough so that for all $\in [0, S]$ we have

$$\frac{x+a}{x+j+a+b} + \sqrt{\frac{(x+a)(j+b)}{(x+j+a+b)^2(x+j+a+b+1)}} \le \delta$$

For this observe we can bound the mean plus k standard deviation this expression by

$$\frac{S+a}{S+j+a+b} + k\sqrt{\frac{(S+a)(j+b)}{(j+a+b)^3}}.$$

which decreases to zero as $j \to \infty$. Thus there exists $J > 0$ such that for all $j > 0$ this inequality holds. Using the same bounds as before we find that we can get an uniform error bound by approximating

$$\frac{\mathbb{E}[V^x(1-V)^j]}{g(0)\int_0^1 v^{x+a-1}(1-v)^{j+b-1}\,\mathrm{d}\,v}.$$

Next we approximate this Beta integral using Stirlings approximation for Beta integrals. This yields that $\int_0^1 v^{x+a-1}(1-v)^{j+b-1}\,\mathrm{d}\,v$ with uniform error bounds for $x \in [0, S]$. Combining these two errors gives the uniform bounds as claimed. $\qquad\square$

Index

Bibliography

[1] Subhashis Ghosal, Aad van der Vaart, *Fundamentals of nonparametric Bayesian inference*, Cambridge university press, Cambridge, 1st edition, 2017.

[2] Jim Pitman, *Random Discrete Distributions Invariant under size-biased permutation* Advances in Applied Probability, 1992.

[3] Jim Pitman, Marc Yor. *The two parameter Poisson-Dirichlet distribution derived from a stable subordinator* The Annals of probability, 1997.

[4] Aad van der Vaart, *Asymptotic statistics*, Volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge university press, Cambridge, 1998.