

Het stochastisch modelleren van een limietorderboek

Bachelorscriptie door:
Sophie Houtkooper 6901891

Onder begeleiding van:
Prof. dr. ir. C.W. (Kees) Oosterlee

22-1-2024



**Universiteit
Utrecht**

Inhoudsopgave

1	Inleiding	2
1.1	Limietorderboek	2
1.2	Algoritmisch en hoogfrequente handel	4
1.3	Modelleringsmethodes	5
1.4	Onderzoek	7
2	Sprongprocessen	7
2.1	Het poissonproces	7
2.1.1	Het samengestelde poissonproces	9
2.1.2	De karakteristieke functie en laplacetransformatie voor het poissonproces	10
2.1.3	Stochastische integratie met het poissonproces	11
2.2	Het hawkesproces	13
2.2.1	De karakteristieke functie en laplacetransformatie voor het hawkesproces	15
3	Uitvoeringskans van een limietorder	16
3.1	Geboorte- en sterfteproces met poisson	17
3.1.1	Transities	17
3.1.2	Laplacetransformatie eerste-doorgangstijd	18
3.1.3	Uitvoeringskans limietorder op de beste aanbodprijs	21
3.1.4	Activiteit LOB modelleren met poisson	25
3.2	Geboorte- en sterfteproces met hawkes	25
3.2.1	Kansdichtheid eerste-doorgangstijd	26
3.2.2	Activiteit LOB modelleren met hawkes	27
4	Conclusie	27
5	Bibliografie	29

1 Inleiding

1.1 Limietorderboek

Op de vraag-/aanbodmarkt kan een deelnemer een order, of ook wel opdracht, plaatsen om aan te geven dat hij/zij een bepaalde hoeveelheid van iets wil kopen/verkopen. Wanneer het zo is dat een deelnemer een specifieke prijs wil geven/ontvangen voor deze bepaalde hoeveelheid, dan noemen we dit een limietorder. Alle limietorders die worden geplaatst, zitten verzameld in het limietorderboek (LOB), zie figuur 1. Wanneer een aanbodorder op een hoog prijslevel ligt, is dit natuurlijk voor een verkoper erg aantrekkelijk. We definiëren daarom de beste aanbodprijs (p_B), die gelijk is aan de hoogste prijs waarvoor er een aanbodorder is geplaatst. Andersom hebben we ook de beste vraagprijs (p_A), die gelijk is aan de laagste prijs waarvoor een vraagorder is geplaatst. Nu vinden we de zogenaamde bid-ask spread (p_S), als het verschil tussen deze twee prijzen:

De beste vraagprijs op tijd t , is

$$p_A(t) \equiv \inf\{p = 1, \dots, n : X_p(t) > 0\} \wedge (n + 1). \quad (1.1)$$

De beste aanbodprijs op tijd t , is

$$p_B(t) \equiv \sup\{p = 1, \dots, n : X_p(t) < 0\} \wedge 0. \quad (1.2)$$

Hier is $|X_p(t)|$ het aantal uitstaande limietorders op moment t op prijslevel p , met $1 \leq p \leq n$. Om vraag en aanbod uit elkaar te houden geven we de prijslevels van aanbodorders weer met een negatieve

waarde. Dus, als $X_p(t) < 0$, dan zijn er $-X_p(t)$ aanbodorders op het prijslevel p en als $X_p(t) > 0$, dan zijn er $X_p(t)$ vraagorders op het prijslevel p .

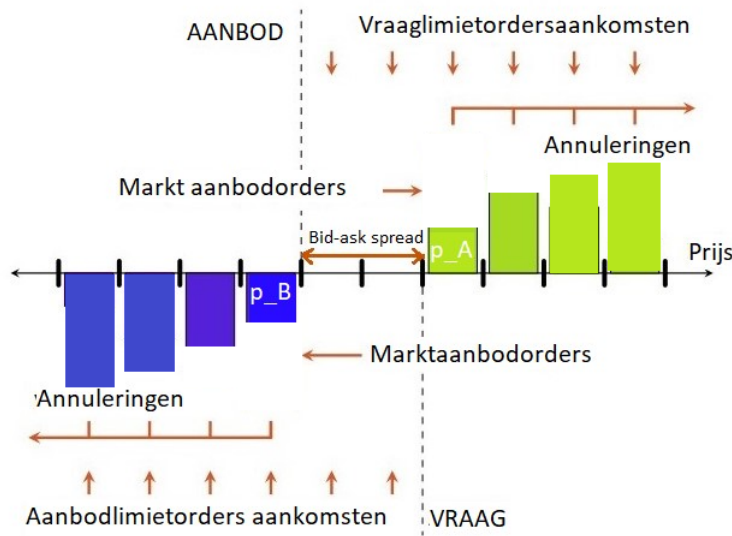
De bid-ask spread op tijd t , is

$$p_S(t) \equiv p_A(t) - p_B(t). \quad (1.3)$$

Ook definiëren we de middenprijs, wat het gemiddelde is van de beste vraag- en aanbodprijs

$$p_M(t) \equiv \frac{p_B(t) + p_A(t)}{2}. \quad (1.4)$$

Deze middenprijs zullen we gebruiken om in de gaten te houden of de beste vraag- en aanbodprijs aan het veranderen zijn in het limietorderboek. De bid-ask spread is nuttig om te weten, omdat we zullen zien dat het aankomen van een limietorder in het LOB afhankelijk is van deze afstand tussen de beste prijzen.



Figuur 1: Weergave van de vraag- en aanbodzijde van het LOB.

De kleinste afstand die tussen twee prijzen kan liggen in een LOB, noemen we de *ticksize*. Dit betekent dat als we een ticksize hebben van 0.001 dan kunnen er limietorders aan het LOB worden toegevoegd voor een bedrag van bijvoorbeeld 1.004 en 1.003, maar niet van 1.0035. We gebruiken de ticksize als eenheid voor de bid-ask spread. Een bid-ask spread ter grootte van 1 ticksize betekent dus ook dat er geen limietorder tussen de beste vraag- en aanbodprijs geplaatst kan worden, omdat de minimale afstand tussen deze beste prijzen ligt.

We hebben dus een orderboek met limietorders, die blijven daar echter niet voor altijd. Een limietorder wordt na enige tijd in het LOB geannuleerd of uitgevoerd. Dit tweede gebeurt wanneer er een marktorder is, die op hetzelfde prijslevel aankomt als deze betreffende limietorder. Een handelaar biedt een marktorder aan wanneer hij/zij direct wil dat deze wordt uitgevoerd tegen een zo goed mogelijke prijs, oftewel de beste vraag-/aanbodprijs. Een marktorder noemen we hierom een agressieve liquiditeitsnemer en een limietorder is een passieve liquiditeitsnemer. Het is ook vaak zo dat er meerdere limietorders in het LOB op hetzelfde prijslevel liggen. Hoe wordt er dan bepaald welke limietorder als eerste wordt uitgevoerd wanneer er een matchende marktorder is? Op de meeste markten gaat dit volgens de 'first in, first out'-regel (Fifo), wat betekent dat de limietorder die al het langst in het LOB

zit wordt uitgevoerd. Maar het gebeurt ook dat er op ordergrootte wordt bepaald welke limietorder de hoogste prioriteit heeft, bijvoorbeeld van groot naar klein. Of door middel van *pro-rata matching*, waarbij de inkomende marktorder wordt gekoppeld aan een limietorder van ongeveer dezelfde grootte. In dit onderzoek gebruiken we de Fifo-wachtrij.

Limietorders hebben dus niet allemaal dezelfde groottes en ze zijn over het algemeen niet precies gelijk aan de groottes van inkomende marktorders. We geven de grootte van een inkomende marktorder weer met q_m en de grootte van een limietorder met de hoogste prioriteit, volgens de gekozen wachtrij-regel, met q_l . Dan zijn er twee situaties die kunnen voorkomen:

Als $q_m > q_l$ dan wordt de limietorder volledig uitgevoerd en wordt de niet uitgevoerde rest ($q_m - q_l$) van de marktorder gematcht aan de volgende limietorder volgens de gekozen wachtrij-regel.

Als $q_l > q_m$ dan is de limietorder deels uitgevoerd en het restant $q_l - q_m$ blijft in het orderboek tot hij uitgevoerd of geannuleerd wordt.

De zogenaamde 'diepte' van een orderboek wordt bepaald door de vraagorders die op een prijslevel boven de beste vraagprijs liggen of aanbodorders die op een prijslevel onder de beste aanbodprijs liggen. Voor handelaren is een prijs dieper in het orderboek natuurlijk beter, maar de kans dat een order dan wordt uitgevoerd is lager. Ga je voor uitvoeringszekerheid, tegen een minder goede prijs, of voor het risico van niet-uitvoering tegen een betere prijs? De keuze op welk prijslevel je een limietorder plaatst, is makkelijker te maken wanneer we de uitvoeringskansen van een limietorder op verschillende prijslevels zouden weten.

1.2 Algoritmisch en hoogfrequente handel

Wat we met algoritmisch handelen bedoelen is gebruik maken van computeralgoritmes en wiskundige modellen om orders uit te voeren. Deze algoritmes maken keuzes, op korte termijn, gebaseerd op input parameters, die worden ingediend door handelaren en marktcondities zoals de volatiliteit¹, prijs, tijd enzovoorts. Algoritmische handelssystemen kunnen onder andere: het detecteren van winstgevende mogelijkheden, het optimaal uitvoeren van orders, detecteren en exploiteren van strategieën van een concurrent en het verbergen van de eigen intenties. De Nederlandse Autoriteit voor Financiële Markten (AFM) maakt onderscheid tussen de twee verschillende termen algoritmisch handelen en algoritmisch uitvoeren [10].

Algoritmisch uitvoeren betekent het uitvoeren van orders gebaseerd op vooraf bepaalde investeringsstrategieën of besluiten.

Algoritmisch handelen betekent een specifieke investeringsstrategie automatiseren. Het uitvoeren is een onderdeel van het algoritme. In tegenstelling tot algoritmisch uitvoeren, neemt het systeem de investeringsbesluiten zelf, gebaseerd op de input van de markt en beperkingen in het model.

Hoogfrequente handel (HFT) is een onderdeel van algoritmisch handelen waarbij de besluiten in een enorm korte tijdsspanne gemaakt worden (vaak in millisecondes). Handelssystemen proberen door extreem snelle signaalverwerking hun tegenstanders voor te zijn.

De door de AMF beschouwde voordelen van het sterk stijgende deel van de handelsactiviteit door algoritmische systemen, zijn:

- 1) Algoritmisch handelen heeft voor een verhoogde snelheid en snellere overdracht van risico's bij hoogfrequente handel gezorgd.
- 2) Door algoritmisch handelen en groeiende competitie tussen marktdeelnemers, is de bid-ask spread significant afgenomen voor een aantal financiële instrumenten. Hierdoor kunnen deze onderdelen nu goedkoper en makkelijker worden gekocht en verkocht, wat de marktliquiditeit heeft verhoogd.
- 3) Een algoritmisch handelssysteem handelt op basis van hoe het geprogrammeerd is in plaats van op basis van emoties, dus kunnen fouten door menselijke emoties worden voorkomen.

Maar er zijn ook risico's:

¹De mate van schommelingen van de prijzen van een financieel instrument, gedurende een bepaalde periode.

- 1) Algoritmische systemen kunnen zorgen voor verstoring van de markt wat de stabiliteit en efficiëntie van de markt aantast. Bij een *flash crash*² bijvoorbeeld zal een algoritmisch handelssysteem ervoor zorgen dat een activa scherp daalt voor een korte tijd.
- 2) Algoritmische systemen kunnen leren hoe ze de markt moeten manipuleren, zelfs als dit niet de intentie was van de ontwerper.
- 3) Algoritmisch handelen kan de kwaliteit van de markt aantasten, per ongeluk of expres. Deze kwaliteit wordt gemeten door onder andere de bid-ask spread, orderboekdiepte en de efficiëntie van de markt.

1.3 Modelleringsmethodes

Zoals we al benoemd hadden is het weten van de uitvoeringskans, van limietorders, nuttig voor het bepalen op welk prijslevel je een limietorder het beste kunt plaatsen. Er zijn meerdere manieren om een limietorderboek wiskundig te modelleren en zo de uitvoeringskansen van limietorders te berekenen. Een eenvoudige manier is door middel van een optimalisatieprobleem, waarbij er vanuit wordt gegaan dat de uitvoeringskans exponentieel is en afhangt van de afstand tussen de prijs van een limietorder en de beste prijs, en dus afneemt hoe verder deze prijs bij de beste prijs vandaan ligt. De benadering is makkelijk te volgen, maar houdt geen rekening met invloedrijke factoren zoals de toestand van een orderboek en de tijd dat gebeurtenissen kosten.

Andere manieren om de dynamiek van een orderboek te omschrijven zijn met machine learning, econometrische en stochastische modellen. Er zijn al veel van dit soort modellen ontwikkeld, elk met eigen voor- en nadelen. We bespreken er een aantal om vervolgens te besluiten op welke we zouden willen voortbouwen.

Met machine learning kunnen modellen gemaakt worden die goede voorspellingen doen, vooral wanneer omstandigheden veel en snel veranderen blijven de modellen accuraat. Een nadeel van machine learning is wel dat er een enorme hoeveelheid data nodig is als "training" voor het model en veel computerpower nodig is. Er zijn twee onderzoeken die machine learning gebruiken om onderzoek te doen rondom het LOB die we nu gaan bekijken.

Als eerste zien we het onderzoek van Arroyo, Cartea, Moreno-Pino en Zohren [3]. Zij hebben machine learning gecombineerd met overlevingsanalyse door tijdsafhankelijke eigenschappen van het LOB af te beelden op de verdeling van de uitvoeringstijd van een limietorder. Met de statistische methode genaamd overlevingsanalyse kun je een schatting maken van hoe lang het duurt voordat een gebeurtenis plaatsvindt. In dit geval is de overlevingstijd gelijk aan de tijd dat een order in het orderboek blijft voordat het wordt uitgevoerd. Het onderzoek geeft een methode om de uitvoeringstijden van limietorders, op meerdere levels in het LOB, te schatten. Ze geven een statistische analyse van de uitvoeringskansen van orders met verschillende wachtrijafspraken en handelsactiviteiten. Volgens Arroyo et al. volgt uit juiste scoreregels dat hun methode significant beter is dan andere onderzoeken waarbij overlevingsanalyse wordt gebruikt.

Een volgend bekend onderzoek dat ook machine learning methodes met overlevingsanalyse combineert om uitvoeringskansen en tijden te schatten, is het onderzoek van Maglaras, Moallemi en Wang [14]. Ze gebruiken een concept genaamd *Recurrent Neural Network*³ om te bepalen wat de verdeling is van de uitvoeringstijd conditioneel op de huidige markttoestand. Deze toestand kan bijvoorbeeld zijn dat het limietorderboek uit balans is. Er is gebruik gemaakt van data van NASDAQ⁴ als "training" voor het model. Hierdoor leidt dit onderzoek tot enorme kostenafname tijdens het toepassen van een handelsstrategie.

Waar machine learning goed is voor complexe en/of non-lineaire relaties tussen inputvariabelen mo-

²Plotselinge sterke daling van de aandelenkoersen op de financiële markt, is na enkele minuten vaak weer hersteld.

³Een type kunstmatig neurale netwerk dat is ontworpen voor het verwerken van sequentiële gegevens en het werken met herhalende patronen.

⁴National Association of Securities Dealers Automated Quotations, is een Amerikaanse effectenbeurs. Het is een van de grootste en meest bekende beurzen ter wereld.

dellere, kunnen econometrische modellen dit niet. Econometrische modellen analyseren historische data door met regressie de relaties tussen de statistieken te modelleren, en zo het gedrag van een orderboek te voorspellen. Deze relaties zijn goed gedefinieerd en daardoor zijn de modellen vaak goed te begrijpen. Wel zorgt dit ervoor dat de complexiteit van het LOB niet goed kan worden weergegeven. We werken hieronder kort twee onderzoeken uit waar econometrische modellen worden gebruikt voor het benaderen van een LOB.

Als eerste bekijken we het onderzoek van Cho en Nelling [5] waarbij er vanuit wordt gegaan dat marktorders aankomen volgens een poissonproces en de tijd tot uitvoering Weibull verdeeld is. Een annulering behandelen ze als gecensureerd, wat betekent dat de exacte reden of timing van de annulering niet volledig wordt waargenomen of opgenomen in het model. Door annuleringen als gecensureerd te behandelen, kan het model nog steeds proberen orderboekgedrag te modelleren zonder gedetailleerd in te gaan op individuele annuleringen. Het kan de complexiteit verminderen en de modeltraining vereenvoudigen, vooral als het doel is om bredere trends in marktgedrag te begrijpen. Limietorders die niet volledig worden uitgevoerd, behandelen ze alsof de grootte van de uitvoering even groot is als de ordergrootte. De data die ze gebruiken als "training", is van de beurs van New York en dit heeft geleid tot een aantal bevindingen: 1) Hoe langer een limietorder in het orderboek blijft, des te kleiner de kans dat het nog uitgevoerd wordt; 2) Een vraagorder wordt minder snel uitgevoerd dan een aanbodorder; 3) Voor grotere orders en orders die ver van de beste prijs liggen is de uitvoeringskans lager; en 4) In tijden van hoge prijsvolatiliteit en een grote bid-ask spread neemt de uitvoeringskans van orders toe.

Een volgend onderzoek wat een econometrisch model heeft gebruikt is dat van Lo, McKinlay en Zhang [12]. Ze baseren het model op overlevingsanalyse en de verdeling van uitvoeringstijden wordt bepaald conditioneel op bijvoorbeeld ordergrootte, marktvolatiliteit of limietprijs. Net zoals bij Cho en Nelling wordt een annulering gezien als gecensureerde observatie. De data die ze gebruiken als "training", is van *S&P 500*⁵ en dit heeft geleid tot de verdeling van de tijd tot de eerste uitvoering, namelijk een generaliserende gamma-verdeling met een versnelde faaltijd. De versnelde faaltijd wijst op het feit dat bepaalde variabelen of factoren de snelheid van de uitvoering kunnen beïnvloeden. Met andere woorden, het geeft aan dat sommige belangrijke variabelen het tijdstip van uitvoering kunnen versnellen of vertragen. Er volgt in dit geval uit, dat de ordergrootte weinig invloed heeft op de uitvoeringskans.

De (aankomst)tijden van limietorders, marktorders en annuleringen zijn heel willekeurig. Omdat dit direct kan worden opgenomen in een stochastisch model is dit een vierde handige manier om het limietorderboek te modelleren. Toch is ook een stochastisch model niet perfect. Zo zijn stochastische modellen afhankelijk van complexe wiskunde en ook maken ze gebruik van sterke, versimpelde aannames over de orderboekdynamiek.

Deze versimpelde aannames maakt dat de modellen vaak eenvoudig geschat kunnen worden op basis van de gegevens. Hier maakt het onderzoek van Cont, Stoikov en Talreja [6] gebruik van. Er wordt aangenomen dat de orders allemaal even groot zijn en aankomen volgens een poissonproces (PP) (zie sectie 2.1). Er volgen meerdere conditionele kansen uit het onderzoek, met behulp van matrixberekeningen en laplacetransformaties (zie sectie 3.1.2). Dit onderzoek maakt gebruik van data van de Tokio beurs om zo parameters te schatten. Bij dit onderzoek is er nog veel ruimte voor uitbreidingen zoals een breder scala aan empirisch waargenomen eigenschappen, toestandsafhankelijke intensiteiten, heterogeniteit⁶ van ordergroottes en een realistischer verdeling voor de tussenaankomsttijden.

Een ander stochastisch onderzoek dat het LOB bestudeert is dat van de Universiteit van Oxford [1]. Zij maken gebruik van het hawkesproces in plaats van het poissonproces om de dynamiek van het LOB vast te leggen. Ze gebruiken een schaallimietstelling en het onderzoek van Hambly et al. [?] om een stelsel differentiaalvergelijkingen te vinden die het volledige LOB modelleert. Het model simuleert goed het prijsproces van het LOB, maar overschat de volatiliteit.

⁵Standard & Poor's 500, is een van de meest gevolgde aandelenindices in de Verenigde Staten en wereldwijd.

⁶Als orders niet uniform of constant zijn in termen van hun grootte.

1.4 Onderzoek

We zien dus nu allerlei manieren om een LOB te benaderen en we zien twee stochastische processen die bruikbaar zijn bij het analyseren van een LOB, maar elk met eigen voor- en nadelen. We kiezen er voor om het onderzoek van Cont et al. te gaan bestuderen en dus het LOB te modelleren met een geboorte- en sterfteproces. Maar door het onderzoek van Oxford zijn we geïnteresseerd geraakt in wat het hawkesproces voor voordelen heeft op het poissonproces. Hieruit volgen de onderzoeks- en deelvraag van deze scriptie:

Kunnen we de dynamiek van een limietorderboek accuraat modelleren met een stochastisch proces?

En

Wat zijn de voor- of nadelen van het hawkesproces in plaats van het poissonproces in het kader van een geboorte- en sterfteproces, voor het berekenen van de uitvoeringskansen van een limietorder?

We bestuderen eerst in hoofdstuk 2 het poissonproces en kort het hawkesproces door middel van het onderzoek van Luis A. Souto Arias [2] en dat van Errais et al. [9]. Hierna bestuderen we in hoofdstuk 3 hoe we net zoals in het onderzoek van Cont et al. [6] het geboorte- en sterfteproces kunnen gebruiken voor het berekenen van de uitvoeringskansen en bekijken we hoe we op een vergelijkbare manier het hawkesproces kunnen gebruiken, zoals in het onderzoek van Gao, Zhou en Zhu [11]. Uiteindelijk in hoofdstuk 4 vergelijken we deze twee methodes uit deze onderzoeken en geven we antwoord op de onderzoeksvraag.

2 Sprongprocessen

2.1 Het poissonproces

Dit stuk theorie komt uit het boek *Stochastic Calculus for finance II* [15] hoofdstuk 11. Voor uitgebreide bewijzen van de stellingen, zie het boek.

Laat τ een exponentiële kansvariabele zijn met kansdichtheid

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0, \\ 0, & t < 0, \end{cases} \quad (2.1)$$

waar λ een positieve constante is. Door middel van partiële integratie vinden we de verwachte waarde van τ :

$$\mathbb{E}[\tau] = \frac{1}{\lambda}. \quad (2.2)$$

We vinden de cumulatieve distributiefunctie (CDF) als volgt

$$F(t) = \mathbb{P}(\tau \leq t) = \int_0^t \lambda e^{-\lambda u} du = -e^{-\lambda u} \Big|_0^t = 1 - e^{-\lambda t}, t \geq 0. \quad (2.3)$$

Hieruit volgt $\mathbb{P}(\tau > t) = e^{-\lambda t}$, $t \geq 0$.

Een exponentiële verdeling is geheugenloos. Dit betekent dat na s tijdseenheden de kans dat we nog eens t tijdseenheden moeten wachten op een gebeurtenis (bijvoorbeeld de annulering van een limietorder) hetzelfde is als de kans dat we t tijdseenheden moeten wachten wanneer we beginnen op tijd 0. We zien voor een gebeurtenis waarvan de tijd waarop het plaatsvindt exponentieel verdeeld is met een verwachte waarde $1/\lambda$ dat:

$$\mathbb{P}(\tau > t + s | \tau > s) = \frac{\mathbb{P}(\tau > t + s \cap \tau > s)}{\mathbb{P}(\tau > s)} = \frac{\mathbb{P}(\tau > t + s)}{\mathbb{P}(\tau > s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathbb{P}(\tau > t). \quad (2.4)$$

Neem een reeks onafhankelijke exponentiële kansvariabelen τ_1, τ_2, \dots met dezelfde verwachte waarde $1/\lambda$. We noemen deze kansvariabelen de tussenaankomsttijden van de sprongen. Dus de eerste sprong in een poissonproces is op tijd τ_1 en de tweede sprong is hierna τ_2 tijdseenheden later en zo verder. De tijden waarop de sprongen plaatsvinden, ook wel de aankomsttijden, definiëren we als

$$S_n = \sum_{k=1}^n \tau_k. \quad (2.5)$$

Poissonprocessen tellen het aantal sprongen dat voorkomt op (of voor) een tijd t . Dit kan dus goed gebruikt worden om bepaalde tijdsafhankelijke processen te modelleren, zoals het aankomen en vertrekken van orders uit een LOB. We geven het poissonproces weer als volgt:

$$N(t) = \begin{cases} 0 & \text{als } 0 \leq t < S_1 \\ 1 & \text{als } S_1 \leq t < S_2 \\ \vdots & \\ n & \text{als } S_n \leq t < S_{n+1} \\ \vdots & \end{cases} \quad (2.6)$$

We zien dat de functie rechts-continu is want $N(t) = \lim_{s \downarrow t} N(s)$. We noteren de filtratie van informatie die volgt uit $N(t)$ als $\mathcal{F}(t)$, zie definitie 2.4. De verwachte tijd tussen de sprongen is $1/\lambda$, dat betekent dat er gemiddeld λ sprongen per tijdseenheid plaatsvinden. We zeggen ook wel dat het poissonproces $N(t)$ een intensiteit λ heeft.

Met behulp van inductie vinden we dat de hiervoor genoemde kansvariabele S_n voor $n \geq 1$ Gamma verdeeld is. Het heeft de volgende kansdichtheid

$$g_n(s) = \frac{(\lambda s)^{n-1}}{(n-1)!} \lambda e^{-\lambda s}, s \geq 0. \quad (2.7)$$

Voor het poissonproces $N(t)$ met intensiteit λ vinden we door middel van het feit dat $\mathbb{P}(N(t) \geq k) = \mathbb{P}(S_k \leq t) \forall k \geq 1$, dat

$$\mathbb{P}(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, k = 0, 1, \dots \quad (2.8)$$

We gaan nu kijken naar het increment $N(t+s) - N(s)$. We weten dat dit onafhankelijk is van $\mathcal{F}(s)$ want de tijd tussen s en de eerste sprong na s , oftewel τ_{s+1} , heeft een exponentiële verdeling met verwachte waarde $1/\lambda$. Dus het is onafhankelijk van wat er gebeurd is tot op tijd s . Hetzelfde geldt voor elke sprong die hierna volgt. De verdeling van $N(t+s) - N(s)$ is dus hetzelfde als de verdeling van $N(t)$. In beide gevallen wordt er namelijk geteld hoeveel sprongen voorkomen in een tijdsinterval van lengte t . We noemen een variabele stationair wanneer de verdeling enkel afhankelijk is van het verschil tussen de tijdstippen. We vinden voor incrementen nu de verdeling als volgt: Laat $N(t)$ een poissonproces zijn met intensiteit $\lambda > 0$ en $0 = t_0 < t_1 < \dots < t_n$ gegeven. Dan zijn de incrementen $N(t_1) - N(t_0), \dots, N(t_n) - N(t_{n-1})$ stationair en onafhankelijk en

$$\mathbb{P}(N(t_{j+1}) - N(t_j) = k) = \frac{\lambda^k (t_{j+1} - t_j)^k}{k!} e^{-\lambda(t_{j+1} - t_j)}, k = 0, 1, \dots \quad (2.9)$$

We hebben dus in (2.9) gevonden dat een poisson-increment $N(t) - N(s)$ een verdeling

$$\mathbb{P}(N(t) - N(s) = k) = \frac{\lambda^k (t - s)^k}{k!} e^{-\lambda(t-s)}, k = 0, 1, \dots \quad (2.10)$$

heeft. Dit is dus de kans dat er een k aantal sprongen plaatsvindt in de volgende $t - s$ tijdseenheden. Door middel van de bekende exponentiële machtsreeks $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ vinden we nu dat de verwachte waarde van zo een increment gelijk is aan

$$\mathbb{E}[N(t) - N(s)] = \lambda(t - s). \quad (2.11)$$

Dus het gemiddelde aantal sprongen tussen tijden s en t is $\lambda(t - s)$. Wanneer we het tweede moment van het increment bepalen, zien we

$$\mathbb{E}[(N(t) - N(s))^2] = \lambda^2(t - s)^2 + \lambda(t - s). \quad (2.12)$$

Hieruit volgt direct dat

$$\text{Var}[N(t) - N(s)] = \lambda(t - s). \quad (2.13)$$

We definiëren het gecompenseerde poissonproces

$$M_P(t) = N(t) - \lambda t. \quad (2.14)$$

Waarbij $N(t)$ een poissonproces is met intensiteit λ . We zien dat $M_P(t)$ een martingaal is, want $\mathbb{E}[M_P(t)|\mathcal{F}(s)] = M_P(s)$.

2.1.1 Het samengestelde poissonproces

Bij een poissonproces of gecompenseerd poissonproces is de spronggrootte 1, maar voor financiële modellen is het belangrijk dat we sprongen van willekeurige groottes modelleren. Hierom introduceren we het samengestelde poissonproces.

We nemen een reeks identiek verdeelde kansvariabelen Y_1, Y_2, \dots met allemaal dezelfde verwachte waarde $\mathbb{E}[Y_i] = \kappa$, en nemen aan dat deze variabelen onafhankelijk zijn van elkaar en van het poissonproces $N(t)$ met intensiteit λ . We definiëren nu het samengesteld poissonproces als

$$Q(t) = \sum_{i=1}^{N(t)} Y_i, t \geq 0. \quad (2.15)$$

We zien dat de sprongen in $Q(t)$ op hetzelfde moment zijn als in $N(t)$, maar de grootte van de sprong is niet altijd meer 1. De eerste sprong in $Q(t)$ is van grootte Y_1 de tweede sprong Y_2 enzovoort.

We zien dat de incrementen van het samengesteld poissonproces net zoals van het simpele poissonproces onafhankelijk zijn en dat $Q(s) = \sum_{i=1}^{N(s)} Y_i$ voor $0 \leq s < t$ onafhankelijk is van $Q(t) - Q(s) = \sum_{i=N(s)+1}^{N(t)} Y_i$. Omdat we bij (2.10) zagen dat $N(t) - N(s)$ dezelfde verdeling heeft als $N(t - s)$, volgt dat $Q(t) - Q(s)$ ook dezelfde verdeling heeft als $Q(t - s)$. Gemiddeld zijn er λt sprongen in het tijdsinterval $[0, t]$ en de gemiddelde spronggrootte is κ . Het aantal sprongen en de spronggrootte zijn onafhankelijk, dus is de verwachte waarde van $Q(t)$:

$$\mathbb{E}[Q(t)] = \kappa \lambda t.$$

Net zoals bij een simpel poissonproces is het gecompenseerde samengesteld poissonproces

$$Q(t) - \kappa \lambda t,$$

een martingaal.

Neem een eindige set van niet-nul getallen y_1, \dots, y_m en laat $r(y_1), \dots, r(y_m)$ positieve getallen zijn die opgeteld 1 zijn. Laat Y_1, Y_2, \dots een reeks onafhankelijke, identiek verdeelde kansvariabele zijn met $\mathbb{P}(Y_i = y_m) = r(y_m)$, $m = 1, \dots, M$. Laat $N(t)$ een poissonproces en $Q(t)$ zoals in vergelijking (2.15).

Voor $m = 1, \dots, M$ definiëren we $N_m(t)$ als het aantal sprongen in Q tot en met tijd t van de grootte y_m . Deze processen N_1, \dots, N_M zijn onafhankelijke poissonprocessen en hebben elk een intensiteit $\lambda = r(y_m), m = 1, \dots, M$. Dan

$$N(t) = \sum_{m=1}^M N_m(t) \text{ en } Q(t) = \sum_{m=1}^M y_m N_m(t).$$

2.1.2 De karakteristieke functie en laplacetransformatie voor het poissonproces

Voor de verdeling van $Q(t_j - t_{j-1})$ is niet gemakkelijk een functie te vinden, omdat de kansdichtheid hiervan vrij ingewikkeld is. Wel is de formule van de moment-generende functie goed te vinden.

Definitie 2.1. *De moment-genererende functie van een kansvariabele Y_i voor een reële s is*

$$\varphi_Y(s) = \mathbb{E}[e^{sY_i}]. \quad (2.16)$$

We nemen $Q(t)$ zoals gedefinieerd in (2.15) en vinden nu met behulp van bovenstaande definitie de moment-genererende functie voor $Q(t)$:

$$\varphi_{Q(t)}(s) = \exp\{\lambda t(\varphi_Y(s) - 1)\}. \quad (2.17)$$

Als de kansvariabelen Y_i niet volledig willekeurig zijn maar altijd een constante waarde y hebben dan geldt $Q(t) = yN(t)$ en $\varphi_y(s) = e^{sy}$. Dus y maal een poissonproces heeft de moment-genererende functie

$$\varphi_{yN(t)}(s) = \exp\{\lambda t(e^{sy} - 1)\}. \quad (2.18)$$

Wanneer $y = 1$ vinden we dan de moment-genererende functie van een simpel poissonproces

$$\varphi_{N(t)}(s) = \exp\{\lambda t(e^s - 1)\}. \quad (2.19)$$

Vanuit deze moment-genererende functie kunnen we nu gemakkelijk de zogenaamde karakteristieke functie vinden.

Definitie 2.2. *De karakteristieke functie van een kansvariabele Y_i voor een reële s is*

$$\psi_Y(s) = \mathbb{E}[e^{isY_i}] = \int_{-\infty}^{\infty} e^{ist} f_Y(t) dt. \quad (2.20)$$

Door in vergelijking (2.19), s te vervangen met is vinden we de karakteristieke functie van een simpel poissonproces

$$\psi_{N(t)}(s) = \exp\{\lambda t(e^{is} - 1)\}. \quad (2.21)$$

De karakteristieke functie van een kansvariabele kunnen we gebruiken voor het vinden van de kansdichtheid van deze kansvariabele.

Een andere manier om de kansdichtheid van een kansvariabele te vinden is door middel van de laplacetransformatie. Wat een integraal transformatie is die een functie met een reële variabele (t) omzet naar een functie met een complexe variabele (s).

Definitie 2.3. *De laplacetransformatie van een functie $f_Y : \mathbb{R} \rightarrow \mathbb{R}$ is als volgt*

$$\hat{f}_Y(s) = \int_{-\infty}^{\infty} e^{-st} f_Y(t) dt, \quad (2.22)$$

waar $s = \gamma + ib \in \mathbb{C}$ voor $\gamma, b \in \mathbb{R}$ en i het imaginaire getal ($i^2 = -1$). We zeggen dat als een variabele Y_i de kansdichtheidsfunctie f_Y heeft, dan is \hat{f}_Y de laplacetransformatie van Y_i .

Merk op dat, door s in de moment-genererende functie zoals gedefinieerd in vergelijking (2.16) te vervangen met $-s$ verkrijgen we de laplacetransformatie zoals gedefinieerd in vergelijking (2.22).

Twee nuttige eigenschappen van de laplacetransformatie, die we gaan gebruiken bij het berekenen van de uitvoeringskans van een limietorder in secties 3.1.2 en 3.1.3, zijn:

i) Voor twee onafhankelijke variabelen X en Y met elk hun laplacetransformaties goed gedefinieerd, zien we dat

$$\hat{f}_{X+Y}(s) = \mathbb{E}[e^{-s(X+Y)}] = \mathbb{E}[e^{-sX}]\mathbb{E}[e^{-sY}] = \hat{f}_X(s)\hat{f}_Y(s). \quad (2.23)$$

ii) De laplacetransformatie heeft een inverse. Deze wordt gegeven door het zogenaamde Bromwich contour integraal

$$f(t) = \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} e^{st} \hat{f}(s) ds, \quad (2.24)$$

als $f(t)$ continu is op t en als $\int_{-\infty}^{\infty} |\hat{f}(\gamma + ib)| db < \infty$ voor $\gamma, b \in \mathbb{R}$ en i het imaginaire getal.

2.1.3 Stochastische integratie met het poissonproces

Er zal in dit gedeelte altijd een enkele filtratie, geassocieerd zijn aan een brownse beweging (BM), poissonproces of samengesteld poissonproces. We bedoelen het volgende met zo een filtratie.

Definitie 2.4. *Laat $(\Omega, \mathcal{F}, \mathbb{P})$ een kansruimte zijn, en laat $\mathcal{F}(t), t \geq 0$ een filtratie op deze ruimte zijn. We zeggen dat een brownse beweging W een BM geassocieerd met deze filtratie is als $W(t) \mathcal{F}(t)$ -meetbaar is voor alle t en voor alle $u > t$ de incrementen $W(u) - W(t)$ onafhankelijk zijn van $\mathcal{F}(t)$. Hetzelfde is te zeggen voor een (samengesteld-)poissonproces dat $\mathcal{F}(t)$ -meetbaar is en onafhankelijke incrementen heeft.*

We definiëren de stochastische integraal

$$\int_0^t \Phi(s) dX(s),$$

met de integrator X die sprongen kan hebben. Alle integratoren die we gaan behandelen zullen rechts-continu zijn en van de vorm

$$X(t) = X(0) + I(t) + R(t) + J(t). \quad (2.25)$$

Het continue deel van (2.25) is gedefinieerd als

$$X^c(t) = X(0) + I(t) + R(t) = X(0) + \int_0^t \Gamma(s) dW(s) + \int_0^t \Theta(s) ds. \quad (2.26)$$

Het bestaat uit: $X(0)$ een niet-willekeurige initiële toestand; een Itô-integraal van een aangepast proces $\Gamma(s)$, met respect tot een BM relatief tot de filtratie; en een riemannintegraal van een aangepast proces $\Theta(t)$ met respect tot een BM relatief tot de filtratie.

Het laatste deel van (2.25), $J(t)$, is een aangepast, rechts-continu puur sprongproces met $J(0) = 0$. De links-continue variant van dit proces geven we weer met $J(t-)$ en dit is de waarde van J direct voor de sprong. We noemen $J(t)$ een puur sprongproces, omdat J niet springt op tijd nul, maar eindig veel sprongen heeft op een eindig tijdsinterval $(0, T]$, en constant is tussen sprongen. Een gecompenseerd poissonproces neemt af tussen sprongen in en is dus geen puur sprongproces.

We noemen een proces $X(t)$ zoals in (2.25)-(2.26) een sprongproces.

Definitie 2.5. *Laat $X(t)$ een sprongproces zijn van de vorm (2.25)-(2.26) en laat $\Phi(s)$ een aangepast proces zijn. Het stochastische integraal van Φ met respect tot X is gedefinieerd als*

$$\int_0^t \Phi(s) dX(s) = \int_0^t \Phi(s) \Gamma(s) dW(s) + \int_0^t \Phi(s) \Theta(s) ds + \sum_{0 < s \leq t} \Phi(s) \Delta J(s). \quad (2.27)$$

In differentiële notatie geeft dit

$$\Phi(t)dX(t) = \Phi(t)\Gamma(t)dW(t) + \Phi(t)\Theta(t)dt + \Phi(t)dJ(t) = \Phi(t)dI(t) + \Phi(t)dR(t) + \Phi(t)dJ(t). \quad (2.28)$$

We vinden voor $X(t) = M(t) = N(t) - \lambda t$, waar $N(t)$ een poissonproces is met intensiteit λ dus $M(t)$ een gecompenseerd poissonproces is zoals in (2.14), dat

$$\int_0^t \Phi(s)dM(s) = N(t), \quad (2.29)$$

voor een $\Phi(s) = \Delta N(s)$. Dit volgt uit het feit dat $I(t) = 0$, $X^c(t) = R(t) = \lambda t$, en $J(t) = N(t)$.

Voor een brownse beweging $W(t)$, definiëren we een stochastisch integraal als volgt $I(t) = \int_0^t \Gamma(s)dW(s)$, zodat $I(t)$ martingaal is. Dit volgt uit het feit dat voor elke n ,

$$I_n(t) = \int_0^t \Gamma_n(s)dW(s),$$

martingaal is. En dat $I(t)$ de limiet is van $I_n(t)$ als $n \rightarrow \infty$ en de limiet van martingalen zijn ook martingaal.

We willen dat een stochastisch integraal van een martingaal ook martingaal is, maar uit vergelijking (2.29) volgt dat dit niet altijd het geval is. We zien wel dat voor $X(t) = M(t) = N(t) - \lambda t$, waar $N(t)$ een poissonproces is met intensiteit λ dus $M(t)$ is een gecompenseerd poissonproces zoals in (2.14), dat

$$\int_0^t \Phi(s)dM(s) = \mathbb{1}_{\{S_1, \infty\}}(t) - \lambda(t \wedge S_1), \quad (2.30)$$

voor een $\Phi(s) = \mathbb{1}_{\{S_1, \infty\}}(s)$. Hier is de integraal van een martingaal dus wel martingaal. Het verschil tussen de vergelijkingen (2.29) en (2.30) is dat de ene rechts- en de ander links-continu is. We gebruiken de volgende stelling.

Stelling 2.1. *We nemen aan dat het sprongproces $X(s)$ van (2.25)-(2.26) martingaal is, de integrand $\Phi(s)$ is links-continu en aangepast, en*

$$\mathbb{E} \int_0^t \Gamma^2(s)\Phi^2(s)ds < \infty \text{ voor alle } t \geq 0.$$

Dan is het stochastische integraal $\int_0^t \Phi(s)dX(s)$ ook martingaal.

Voor een sprongproces $X(t)$ is de kwadratische variatie van X op $[0, T]$ gedefinieerd als

$$[X, X](T) = \lim_{\|\Pi\| \rightarrow 0} Q_\Pi(X),$$

waar $\Pi = \{t_0, t_1, \dots, t_n\}$ voor $0 = t_0 < t_1 < \dots < t_n = T$ en $\|\Pi\| = \max_j(t_{j+1} - t_j)$. En we definiëren

$$Q_\Pi(X) = \sum_{j=0}^{n-1} (X(t_{j+1}) - X(t_j))^2.$$

Wanneer de lengte van het langste subinterval naar nul gaat ($\|\Pi\| \rightarrow 0$) dan moet het aantal punten in Π naar oneindig gaan.

Voor sprongprocessen $X_1(t)$ en $X_2(t)$ definiëren we kruisvariatie als

$$[X_1, X_2](T) = \lim_{\|\Pi\| \rightarrow 0} C_\Pi(X_1, X_2),$$

waar

$$C_{\Pi}(X_1, X_2) = \sum_{j=0}^{n-1} (X_1(t_{j+1}) - X_1(t_j))(X_2(t_{j+1}) - X_2(t_j)).$$

Nu kunnen we de volgende stelling bekijken.

Stelling 2.2. *Laat $X_1(t) = X_1(0) + I_1(t) + R_1(t) + J_1(t)$ en $X_2(t) = X_2(0) + I_2(t) + R_2(t) + J_2(t)$ twee sprongprocessen zijn. Waar $I_i(t) = \int_0^t \Gamma_i(s) dW(s)$, $R_i(t) = \int_0^t \Theta_i(s) ds$ voor $i \in \{1, 2\}$, en $J_i(t)$ voor $i \in \{1, 2\}$ is een rechts-continu puur-sprongproces. Dan geldt,*

$$[X_1, X_2](T) = [X_1^c, X_2^c](T) + [J_1, J_2](T) = \int_0^T \Gamma_1(s) \Gamma_2(s) ds + \sum_{0 < s \leq t} \Delta J_1(s) \Delta J_2(s). \quad (2.31)$$

Let op voor een niet-kruisvariatie kun je gewoon $X_2 = X_1$ invullen.

In differentiële vorm zegt deze stelling eigenlijk dat $dX_1^c(t) dJ_2(t) = dX_2^c(t) dJ_1(t) = 0$, dus de kruisvariatie tussen een continu proces en een puur-sprongproces is nul. Hieruit volgt dat de kruisvariatie tussen een brownse beweging en een poissonproces nul is. In het algemeen geldt dat de kruisvariatie tussen twee processen nul is als een van de processen continu is en de ander geen Itô-integraal deel heeft.

Een ander gevolg van deze stelling is, voor aangepaste processen $\Phi_i(s)$, dat als

$$d\tilde{X}_1(t) = \Phi_1(t) dX_1(t) \text{ en } d\tilde{X}_2(t) = \Phi_2(t) dX_2(t),$$

dan

$$d\tilde{X}_1(t) d\tilde{X}_2(t) = \Phi_1(t) \Phi_2(t) dX_1(t) dX_2(t).$$

2.2 Het hawkesproces

Een belangrijke eigenschap van het poissonproces is dat het geheugenloos is. Dit betekent dat de verdeling van de volgende aankomsttijd enkel afhankelijk is van de huidige toestand en niet van de rest van het verleden. Hierom kan een poissonproces niet goed weergeven hoe bij een LOB samenklontering en onderlinge afhankelijkheid voorkomt. Een idee van een uitbreiding op het PP is dan het hawkesproces.

Het hawkesproces is net als het PP een tellingsproces, dat de hoeveelheid sprongen tot op een tijd t bijhoudt. We zien de positieve reeks van aankomsttijden $a_1 < a_2 < \dots$ op de kansruimte $(\Omega, \mathcal{F}, \mathbb{P})$. Dan krijgen we tellingsproces,

$$N(t) = \sum_{n=1}^{\infty} \mathbb{1}_{a_n \leq t}.$$

Verder definiëren we het geassocieerd puntproces L als

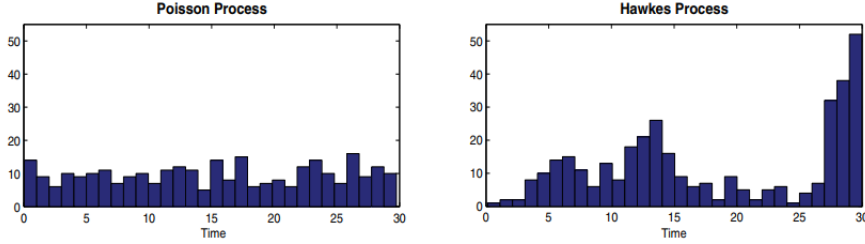
$$L(t) = \sum_{n=1}^{\infty} l_n \cdot \mathbb{1}_{a_n \leq t},$$

waar $\{l_n : n \geq 1\}$ een reeks onafhankelijke en identiek verdeelde niet-negatieve kansvariabelen zijn en voor elke $n \in \mathbb{N}$ is l_n , \mathcal{F}_{a_n} -meetbaar, ook wel de spronggroottes genoemd.

Definitie 2.6. *Een tellingsproces $N(t)$ is een hawkesproces als het intensiteit*

$$\lambda_t = \lambda^* + \int_{-\infty}^t \phi(t-s) dL(s) \quad (2.32)$$

heeft. Waar λ^ de basisintensiteit is en $\phi(\cdot)$ een positieve functie is die behoort tot de integreerbare functies.*



Figuur 2: Trajecten van de poisson- en hawkesprocessen. Een staaf geeft het aantal sprongen in een gegeven jaar. De aankomsten van sprongen zijn bij poisson ongeveer gelijk verdeeld over de tijd, de aankomsten van sprongen over de tijd zijn bij hawkes samengeklonterd. Voor het hawkesproces geldt $\lambda^* = 1$, $\beta = 1.5$, $\alpha = 2$ en elke l_n uit $\{l_n : n \geq 1\}$ heeft een uniforme verdeling ν op $\{0.4, 0.6, 0.8, 1\}$. Het poissonproces is een speciaal geval van het hawkesproces met $\alpha = 0$.

We noemen $\phi(\cdot)$ de geheugenkern omdat het de impact bepaalt van eerdere sprongen op huidige gebeurtenissen. Aangezien de geheugenkern positief is weten we dus dat met elke sprong de intensiteit toeneemt, het heeft dus een zelfopwekkend of samenklonterend effect. Hoe meer sprongen er voorkomen, des te groter de kans dat ze nog eens zullen voorkomen. Het hawkesproces is zo gedefinieerd dat deze zelfopwekkende eigenschap niet zorgt voor divergentie van de intensiteit.

Voor financiële markttoepassingen nemen we aan dat de geheugenkern een exponentiële functie is. Hierdoor wordt het hawkesproces dan een markovproces⁷ en specifiek een affine⁸ sprongdiffusie.

Propositie 2.1. *Beschouw het hawkesproces met exponentiële geheugenkern $\phi(t) = \alpha e^{-\beta t}$, voor $t, \alpha \in \mathbb{R}^+$ en $\beta \in \mathbb{R}^+ \setminus \{0\}$. Neem*

$$\int_{-\infty}^0 \alpha e^{\beta s} dN(s) := (\lambda_0 - \lambda^*) e^{-\beta t}, \quad (2.33)$$

het paar $(N(t), \lambda_t)$ is een markovproces. De intensiteit kan worden herschreven als

$$d\lambda_t = \beta(\lambda^* - \lambda_{t-})dt + \alpha dN(t), \quad (2.34)$$

met $t_- := \lim_{\epsilon \rightarrow 0^+} t - \epsilon$, het moment vlak voor t .

Bewijs. Gegeven vergelijking (2.33), door het differentiëren van vergelijking (2.32), vind je vergelijking (2.34). \square

Merk op dat als $\beta = 0$, dan is de intensiteit tussen gebeurtenissen (sprongen) in constant. En dan volgt dat $(N(t), L(t))$ een geboorteprocess⁹ is. En we zien als $\alpha = 0$ dan is $(N(t), L(t))$ een poissonproces waarvan de intensiteit deterministisch evolueert door de tijd en de tussenaankomsttijden zijn onafhankelijk. In figuur 2 zien we de trajecten van de poisson- en hawkesprocessen. We zien de samenklontering in het hawkesproces tegenover het vrij constant verdeelde poissonproces. De waarde van α bepaalt de frequentie van de samenklontering, β bepaalt de omvang van de samenklontering.

Net zoals bij het PP, definiëren we een gecompenseerde variant van het proces, namelijk het gecompenseerde hawkesproces:

$$M_H(t) := \sum_{j=1}^{N(t)} l_j - \eta \int_0^t \lambda_s ds, \quad (2.35)$$

⁷Een stochastisch proces, waarvoor geldt dat het verleden in een gegeven situatie irrelevant is om de toekomst te voorspellen.

⁸Het affine aspect betekent dat de intensiteiten van het model lineair afhankelijk zijn van de huidige staat van het systeem

⁹Een proces zoals we gaan zien in definitie 3.1 voor $\omega_i = 0 \forall i$

waar l_1 een kansvariabele voor spronggrootte is en $\eta_{l_1} = \mathbb{E}[e^{l_1} - 1]$.

2.2.1 De karakteristieke functie en laplacetransformatie voor het hawkesproces

We kunnen de karakteristieke functie van het hawkesproces vinden door een systeem van differentiaalvergelijkingen op te lossen. Dit systeem halen we uit het onderzoek van [2].

Propositie 2.2. *Laat λ_t de intensiteit van een hawkesproces zijn met een mate van samenklontering $\alpha \in \mathbb{R}^+$, vervaldatum $\beta \in \mathbb{R}^+ \setminus \{0\}$ en basisintensiteit $\lambda^* \in \mathbb{R}^+$. Laat l_1 de kansvariabele voor spronggrootte zijn met karakteristieke functie $\psi_{l_1}(\cdot)$ en $\eta_l = \mathbb{E}[e^{l_1} - 1]$. Dan wordt de karakteristieke functie van het paar $(\lambda_t, M(t))$ gegeven door*

$$\mathbb{E}[e^{iu\lambda_t + iv(M(t) - M(s))} | \mathcal{F}_s] = e^{A(u, v, s) + B(u, v, s)\lambda_s}, \quad (2.36)$$

waar de functies $A(u, v, \tau)$ en $B(u, v, \tau)$ voldoen aan het volgende systeem differentiaalvergelijkingen

$$\frac{\partial A(u, v, \tau)}{\partial \tau} = \beta \lambda^* B(u, v, \tau), \quad (2.37)$$

$$\frac{\partial B(u, v, \tau)}{\partial \tau} = e^{\alpha B(u, v, \tau)} \psi_{l_1}(v) - \beta B(u, v, \tau) - iv\eta_l - 1, \quad (2.38)$$

met grensvoorwaarden $A(u, v, t) = 0, B(u, v, t) = iu$. En waar $M(t) = M_H(t)$ het gecompenseerde hawkesproces is.

Bewijs. Voor een vergelijkbaar bewijs kun je propositie 2.2 uit [9] bekijken. De infinitesimale generator¹⁰ \mathcal{D} van het paar $(\lambda_t, M(t))$ wordt gegeven door

$$(\mathcal{D}g)(\lambda, M) = \beta(\lambda^* - \lambda)g_\lambda(\lambda, M) - \lambda\eta_l g_M(\lambda, M) + \lambda \int (g(\lambda + \alpha, M + x) - g(\lambda, M))d\nu(x), \quad (2.39)$$

waar g_z de partieel afgeleide van g naar z is en $\nu(\cdot)$ is de cumulatieve distributiefunctie(CDF) van de spronggrootte l_1 oftewel $\nu(x) = P(l_1 \leq x)$.

Dan definiëren we $\psi_{\lambda, M}(u, v, t, s) := \mathbb{E}[e^{iu\lambda_t + iv(M(t) - M(s))} | \mathcal{F}_s]$ voor $s \leq t$. We zien dat $\psi_{\lambda, M}(u, v, t, s)$ een martingaal is onder de filtratie gegenereerd door λ_t en $M(t)$. Dan volgt uit de eigenschappen van de infinitesimale generator dat

$$\begin{cases} \frac{\partial \psi_{\lambda, M}}{\partial s}(u, v, t, s) + (\mathcal{D}\psi_{\lambda, M})(u, v, t, s) = 0, \\ \psi_{\lambda, M}(u, v, t, t) = e^{iu\lambda_t + ivM(t)} \end{cases} \quad (2.40)$$

Omdat het paar $(\lambda_t, M^H(t))$ een sprong-diffusieproces is met affine eigenschappen, volgt uit [7] dat

$$\psi_{\lambda, M}(u, v, t, s) = e^{A(u, v, s) + B(u, v, s)\lambda_s}.$$

Dit in vergelijking 2.40 substitueren geeft het gewenste resultaat. □

In sommige gevallen kan het stelsel differentiaalvergelijkingen (2.37) en (2.38) analytisch worden opgelost, bijvoorbeeld wanneer $\beta = 0$. Maar in andere gevallen kan het stelsel differentiaalvergelijkingen (2.37) en (2.38) numeriek opgelost worden, bijvoorbeeld met het Runge-Kutta algoritme [4].

De karakteristieke functie van een kansvariabele kunnen we gebruiken voor het vinden van de kansdichtheid van deze kansvariabele.

¹⁰De infinitesimale generator, van een bepaald Markov proces, is een Fouriervermenigvuldiging operator die een hoop informatie over een proces bevat

Een andere manier om de kansdichtheid van een kansvariabele te vinden is door middel van de laplacetransformatie.

Uit het onderzoek van Gao, Zhou en Zhu [11] volgt onder andere de gezamenlijke laplacetransformatie voor het hawkesproces $(N(t), L(t))$ voor een $t > 0$.

Stelling 2.3. *Laat $\lambda^* \in \mathbb{R}^+$ de basisintensiteit zijn. Dan wordt voor elke $u, v \in \mathbb{C}$ de laplacetransformatie van het paar $(N(t), L(t))$ gegeven door*

$$\mathbb{E}[e^{-uN(t)-vL(t)}] = e^{\int_0^t \lambda^*(t-s)(H(s)-1)ds}, \quad (2.41)$$

waar de functie H de unieke oplossing van de integraal vergelijking

$$H(t) = e^{-u} \mathbb{E}[e^{-vl_1 + \int_0^t l_1 \phi(s)(H(t-s)-1)ds}], \quad (2.42)$$

met $|H(t)| \leq 1$ en waar l_1 de eerste waarde is van de reeks $\{l_n : n \geq 1\}$ die het puntproces L bepalen.

Voor het bewijs van deze stelling verwijs ik naar het bewijs van stelling 1 uit [11].

We zien dat deze stelling net zoals propositie 2.2 voldoet aan de transformatie van hawkes uit propositie 2.2 uit het onderzoek van Errais et al. [9].

Om in sectie 3.2.1 uiteindelijk de uitvoeringskans van een limietorder te kunnen berekenen met het hawkesproces hebben we de enkele laplacetransformatie van $L(t)$ nodig. Die vinden we door in stelling 2.3, $u = 0$ in te vullen.

Gevolg 2.1. *Voor elke $v \in \mathbb{C}$ waarvan het reële deel positief is geldt,*

$$\mathbb{E}[e^{-vL(t)}] = e^{\int_0^t \lambda^*(t-s)(H_L(s)-1)ds}, \quad (2.43)$$

waar de functie H_L de unieke oplossing van de integraal vergelijking

$$H_L(t) = \mathbb{E}[e^{-vl_1 + \int_0^t l_1 \phi(s)(H_L(t-s)-1)ds}], \quad (2.44)$$

met $|H_L(t)| \leq 1$.

Om in sectie 3.2.2 uiteindelijk te kunnen berekenen wat de kans is op een sprong in de volgende t' tijdseenheden, hebben we de enkele laplacetransformatie van $N(t)$ nodig. Die vinden we door in stelling 2.3, $v = 0$ in te vullen.

Gevolg 2.2. *Voor elke $u \in \mathbb{C}$ waarvan het reële deel positief is geldt,*

$$\mathbb{E}[e^{-uN(t)}] = e^{\int_0^t \lambda^*(t-s)(H_N(s)-1)ds}, \quad (2.45)$$

waar de functie H_N de unieke oplossing van de integraal vergelijking

$$H_N(t) = e^{-u} \mathbb{E}[e^{\int_0^t l_1 \phi(s)(H_N(t-s)-1)ds}], \quad (2.46)$$

met $|H_N(t)| \leq 1$.

3 Uitvoeringskans van een limietorder

We gaan in deze sectie de uitvoeringskans, of ook wel *fill probability* genaamd, berekenen. Dit is de kans dat een limietorder wordt uitgevoerd. Deze kans is afhankelijk van zowel interne (prijs en kwantiteit) als externe (marktconditie) factoren. Maar door snel veranderende omstandigheden van een orderboek, is de uitvoeringskans lastig in te schatten. Toch is een schatting hebben van wat de uitvoeringskans is, nuttig voor het maken van besluiten rondom orders plaatsen in het LOB.

3.1 Geboorte- en sterfteproces met poisson

In het onderzoek van Lokin[13], die we bestuderen, wordt gebruik gemaakt van het zogenaamde geboorte- en sterfteproces (BD-proces) op basis van het PP. We gaan het aankomen van een limietorder in het LOB zien als een 'geboorte' en het weggaan van een limietorder uit het LOB (door uitvoering of annulering), gaan we zien als een 'sterftegeval'.

Definitie 3.1. *Zij $P_{i,j}(t) = \mathbb{P}\{X(t+s) = j | X(s) = i\}$ en zij λ de geboorte-intensiteit en ω_i de sterfte-intensiteit. Het proces $X(t)$ is een geboorte- en sterfteproces als voor een kleine $\Delta t > 0$ de functie $P_{i,j}(\Delta t)$ voldoet aan:*

$$\begin{aligned} P_{i,i+1}(\Delta t) &= \lambda \Delta t + o(\Delta t), i \geq 0, \\ P_{i,i-1}(\Delta t) &= \omega_i \Delta t + o(\Delta t), i \geq 1, \\ P_{i,i}(\Delta t) &= 1 - (\lambda + \omega_i) \Delta t + o(\Delta t), i \geq 1. \end{aligned}$$

Het poissonproces gebruiken we in het BD-proces om de aankomsten van deze geboortes of sterfgevallen te modelleren als sprongen. De tussen-aankomsttijden van deze sprongen zijn, zoals we in sectie 2.1 gezien hebben, exponentieel verdeeld met een verwachte waarde van $\frac{1}{\lambda + \omega_i}$.

3.1.1 Transities

In het onderzoek van Lokin wordt het onderzoek van Cont et al. [6] bestudeerd, waarin we de toestand van het LOB weergeven als een continue-tijd markovproces

$$X(t) \equiv (X_1(t), \dots, X_n(t))_{t \geq 0}, \quad (3.1)$$

zoals we dat in sectie 2.1.3 hebben gezien. Waar $|X_p(t)|$ het aantal uitstaande limietorders zijn op moment t op prijslevel p , met $1 \leq p \leq n$. De prijslevels waarvoor een limietorder geplaatst kan worden, moeten in een zogenaamd prijsrooster $\{1, \dots, n\}$ liggen. De n is zo hoog gekozen dat de kans op een prijs $p' > n$ vrijwel nul is.

Zoals al eerder benoemt, geven we de prijslevels van aanbodorders weer met een negatieve waarde, om vraag en aanbod uit elkaar te houden. Dus, als $X_p(t) < 0$, dan zijn er een aantal van $-X_p(t)$ aanbodorders op het prijslevel p en als $X_p(t) > 0$, dan zijn er $X_p(t)$ vraagorders op het prijslevel p .

Er zijn drie gebeurtenissen, of sprongen, waarvan we de aankomsttijden bestuderen om een wiskundig goed model van het LOB te verkrijgen. Namelijk het aankomen van een limietorder bij het LOB, het aankomen van een marktorder bij het LOB of het annuleren van een limietorder. Met $X(t_j)$ geven we de toestand, oftewel het aantal orders in het LOB, weer op een moment t_j en met $X(t_{j+1})$ de toestand na een eerste sprong sinds tijd t_j . In dit onderzoek nemen we aan dat alle orders van dezelfde gemiddelde grootte zijn.

Het aankomen van een limietvraagorder op een prijslevel p_i zorgt voor de volgende transitie:

$$X_{p_i}(t_{j+1}) = X_{p_i}(t_j) + 1 \text{ met intensiteit } \lambda(p_i - p_B(t_j)), \text{ voor } p_i > p_B(t_j). \quad (3.2)$$

We zien dat de intensiteit afhankelijk is van de afstand tussen het prijslevel en de tegenovergestelde beste prijs, dit zullen we in sectie 3.1.3 terug zien komen door de intensiteit bid-ask spread afhankelijk te maken. Er geldt $p_i > p_B(t_j)$, omdat een vraagorder slechts op een prijslevel boven de beste aanbodprijs aan het LOB wordt toegevoegd, onder de beste aanbodprijs zou het direct op de markt worden verhandeld en dus geen limietorder zijn.

Op eenzelfde manier volgt dat het aankomen van een limietaanbodorder op een prijslevel $p_i < p_A(t_j)$ zorgt voor de transitie:

$$X_{p_i}(t_{j+1}) = X_{p_i}(t_j) - 1 \text{ met intensiteit } \lambda(p_A(t_j) - p_i), \text{ voor } p_i < p_A(t_j). \quad (3.3)$$

Omdat aanbodorders negatief worden weergegeven, zien we -1 staan, ook al betreft het wel een limietorder die in het LOB erbij komt.

Een marktorder haalt een uitstaande limietorder uit het LOB en komt altijd aan op de beste vraag- of aanbodprijs. Hierdoor zorgt het aankomen van een marktorder voor een van de volgende twee transities:

$$X_{p_A}(t_{j+1}) = X_{p_A}(t_j) - 1 \text{ met intensiteit } \mu, \quad (3.4)$$

$$X_{p_B}(t_{j+1}) = X_{p_B}(t_j) + 1 \text{ met intensiteit } \mu. \quad (3.5)$$

Hier zien we weer vanwege het negatief weergegeven van de aanbodzijde dat er -1 staat bij de vraagzijde en $+1$ bij de aanbodzijde.

Ook een annulering haalt een uitstaande limietorder uit het LOB. Dit kan slechts voorkomen op een prijslevel waar openstaande orders staan. Door annulering van een limietvraagorder krijgen we daarom de transitie:

$$X_{p_i}(t_{j+1}) = X_{p_i}(t_j) - 1 \text{ met intensiteit } \theta(p_i - p_B(t_j))|X_{p_i}(t_j)|, \text{ voor } p_i \geq p_A(t_j). \quad (3.6)$$

En door annulering van een limietaanbodorder zien we de transitie:

$$X_{p_i}(t_{j+1}) = X_{p_i}(t_j) + 1 \text{ met intensiteit } \theta(p_A(t_j) - p_i)|X_{p_i}(t_j)|, \text{ voor } p_i \leq p_B(t_j). \quad (3.7)$$

De intensiteit volgt uit het feit dat er een hogere kans op annulering is wanneer er meer uitstaande orders zijn op een bepaald prijslevel en dat de annuleringsintensiteit per order afhankelijk is van de afstand tussen het prijslevel en de tegenovergestelde beste prijs. We gaan hierom in sectie 3.1.3 terugzien dat de annuleringsintensiteit bid-ask spread- en toestandsafhankelijk worden weergegeven.

We zien dus dat elk van de drie soorten sprongen voorkomt met een andere intensiteit. De sterfte-intensiteit ω_i zoals in definitie 3.1 is dus een combinatie van de annuleringsintensiteit zoals in vergelijkingen (3.6) en (3.7) en de intensiteit μ van een aankomend marktorder.

3.1.2 Laplacetransformatie eerste-doorgangstijd

Het volgende stuk theorie over voortgezette breuken kom uit het boek van Lorentzen en Waadeland [8]. We leggen uit wat een voortgezette breuk is zodat we die daarna kunnen toepassen bij het vinden van de laplacetransformatie, die we nodig hebben voor het vinden van de uitvoeringskans.

We definiëren $\{f_n\}$ als de reeks

$$f_n = \frac{a_1}{1 + \frac{a_2}{1 + \frac{a_3}{\ddots + a_n}}}. \quad (3.8)$$

Hier gaan we ervan uit dat $\{a_n\} \neq 0$ voor alle n en het kan dat $f_n = \infty$. Verder is $\{f_n\}$ goed gedefinieerd in $\hat{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$, waarbij \mathbb{C} de verzameling van alle complexe getallen is. We noemen de volgende oneindige reeks een voortgezette breuk,

$$\mathcal{K}_{n=1}^{\infty} \frac{a_n}{1} = \frac{a_1}{1 + \frac{a_2}{1 + \frac{a_3}{1 + \dots}}}, \quad (3.9)$$

of ook wel geschreven als $f = \mathcal{K}_{n=1}^{\infty}(a_n/1) = \mathbf{K}(a_n/1)$. Op eenzelfde manier vinden we voor twee reeksen $\{a_n\}, \{b_n\} \in \mathbb{C}$ met $\{a_n\} \neq 0 \forall n$ de voortgezette breuk

$$\mathcal{K}_{n=1}^{\infty} \frac{a_n}{b_n} = \frac{a_1}{b_1 + \frac{a_2}{b_2 + \frac{a_3}{b_3 + \dots}}}, \quad (3.10)$$

oftewel $\mathbf{K}(a_n/b_n)$. We schrijven dit voor het gemak ook als

$$\mathcal{K}_{n=1}^{\infty} \frac{a_n}{b_n} = \frac{a_1}{b_1+} \frac{a_2}{b_2+} \frac{a_3}{b_3+} \cdots \quad (3.11)$$

Definitie 3.2. Een voortgezette breuk convergeert naar $f \in \hat{\mathbb{C}}$ als $\lim_{k \rightarrow \infty} f^{(k)} = f$, met $f^{(k)}$ de k -de benadering van de voortgezette breuk f .

Voor een voortgezette breuk $f = \frac{a_1}{b_1+} \frac{a_2}{b_2+} \frac{a_3}{b_3+} \cdots$, is de k -de benadering

$$f^{(k)} = \frac{a_1}{b_1+} \frac{a_2}{b_2+} \frac{a_3}{b_3+} \cdots \frac{a_k}{b_k} = \frac{A_k}{B_k}. \quad (3.12)$$

Waar

$$\begin{aligned} A_k &= b_k A_{k-1} + a_k A_{k-2} \text{ en} \\ B_k &= b_k B_{k-1} + a_k B_{k-2}, \end{aligned} \quad (3.13)$$

met $A_0 = 0, A_1 = a_1, B_0 = 1, B_1 = b_1$.

Om voortgezette breuken te analyseren, kun je de aangepaste Lentz-methode¹¹ gebruiken. Deze methode maakt gebruik van de volgende verhoudingen

$$C_k = \frac{A_k}{A_{k-1}} \text{ en } B_k = \frac{B_{k-1}}{B_k}. \quad (3.14)$$

Hieruit volgt, samen met (3.13), dat

$$C_k = b_k + \frac{a_k}{C_{k-1}}, \quad (3.15)$$

$$D_k = \frac{1}{b_k + a_k D_{k-1}}. \quad (3.16)$$

Door middel van iteratie kunnen we met

$$f^{(k)} = f^{(k-1)} C_k D_k, \quad (3.17)$$

$f^{(k)}$ evalueren. Wanneer $|f^{(k)} - f^{(k-1)}|$ klein genoeg is, oftewel

$$|C_k D_k - 1| < \epsilon, \quad (3.18)$$

voor een klein getal ϵ , stoppen we met itereren.

Nu we weten wat een voortgezette breuk is, kunnen we dit gaan gebruiken bij het vinden van de laplacetransformatie van een zogenaamde eerste-doorgangstijd. De tijd die het geboorte- en sterfproces nodig heeft om voor de eerste keer van een toestand i naar een toestand $i - 1$ te gaan, noemen we de eerste-doorgangstijd, en geven we weer als $\sigma_{i,i-1}$. Met toestand bedoelen we hier dus het aantal limietorders in het LOB. De eerste keer dat het geboorte- en sterfproces naar een toestand nul gaat, gegeven dat het in een toestand b begint, wordt weergegeven als

$$\sigma_b = \sigma_{b,b-1} + \sigma_{b-1,b-2} + \cdots + \sigma_{1,0}. \quad (3.19)$$

Waar alle verschillende $\sigma_{i,i-1}$ voor $i = 1, \dots, b$ onafhankelijk zijn. Laat $\hat{f}_{\sigma_{i,i-1}}$ de laplacetransformatie zijn van $\sigma_{i,i-1}$ voor $i = 1, \dots, b$ en \hat{f}_{σ_b} die van σ_b zoals in vergelijking (3.19). Dan volgt uit vergelijking (2.23) dat

$$\hat{f}_{\sigma_b}(s) = \prod_{i=1}^b \hat{f}_{\sigma_{i,i-1}}(s). \quad (3.20)$$

Dit klopt omdat alle termen in het rechterlid van de vergelijking (3.19) onafhankelijk zijn.

¹¹De aangepaste Lentz-methode is een numerieke iteratieve methode die wordt gebruikt om de convergentie van een reeks of een functie te bepalen

Propositie 3.1. *De laplacetransformatie $\hat{f}_{i,i-1}$ van de kansdichtheidsfunctie van de eerste-doorgangstijd $\sigma_{i,i-1}$ van een geboorte- en sterfteproces met constante geboorte-intensiteit λ en een sterfte-intensiteit ω_i in toestand i , wordt gegeven door*

$$\hat{f}_{\sigma_{i,i-1}}(s) = -\frac{1}{\lambda} \mathcal{K}_{k=i}^{\infty} \frac{-\lambda\omega_k}{\lambda + \omega_k + s}, \quad (3.21)$$

Bewijs. Omdat de aankomsten van de geboortes of sterftegevallen een poissonproces volgen, is de tijd dat we in een toestand i zijn, natuurlijk even lang als de tijd tussen de sprong die ons naar deze toestand i brengt en de sprong erna. De tijd tussen sprongen is, zoals we in sectie 2.1 gezien hebben, exponentieel verdeeld, in dit geval met een intensiteit $\lambda + \omega_i$. Herken dat hier, net zoals in sectie 3.1.1, de sterfte-intensiteit toestandsafhankelijk is en de geboorte-intensiteit niet. In de volgende sectie gaan we deze sterfte-intensiteit opsplitsen in een annuleringsintensiteit en aankomend marktorder intensiteit, maar voor nu geven we hem voor het gemak nog gecombineerd weer. Hieruit volgt dan, dat de tijd in een toestand i een kansdichtheidsfunctie

$$\mathbb{P}(\text{tijd in een toestand } i = t) = (\lambda + \omega_i)e^{-(\lambda + \omega_i)t}, \quad (3.22)$$

heeft. Na deze exponentieel verdeelde tijd in een toestand i , is er een sprong die ons naar ofwel een toestand $i + 1$, of een toestand $i - 1$, brengt. De kans dat een limietorder erbij komt (geboorte) is dan $\frac{\lambda}{\lambda + \omega_i}$. En de kans dat er een limietorder uit het LOB gaat (sterfte) is $\frac{\omega_i}{\lambda + \omega_i}$.

Met deze kansen en kansdichtheidsfunctie (3.22) kunnen we nu de kansdichtheidsfunctie van $\sigma_{i,i-1}$ berekenen. Laat $f_{\sigma_{i,i-1}}(t)$ de kansdichtheidsfunctie zijn van $\sigma_{i,i-1}$, dan geldt

$$\begin{aligned} f_{\sigma_{i,i-1}}(t) &= \frac{\omega_i}{\lambda + \omega_i} (\lambda + \omega_i)e^{-(\lambda + \omega_i)t} + \frac{\lambda}{\lambda + \omega_i} (\lambda + \omega_i)e^{-(\lambda + \omega_i)t} * f_{\sigma_{i+1,i}}(t) * f_{\sigma_{i,i-1}}(t) \\ &= \omega_i e^{-(\lambda + \omega_i)t} + \lambda e^{-(\lambda + \omega_i)t} * f_{\sigma_{i+1,i}}(t) * f_{\sigma_{i,i-1}}(t). \end{aligned} \quad (3.23)$$

Met het teken $*$ duiden we hier convolutie¹² aan. Er geldt dat de laplacetransformatie van een convolutie van twee functies gelijk is aan het product van de laplacetransformaties van deze functies, oftewel $\widehat{f * g}(s) = \hat{f}(s)\hat{g}(s)$.

Als volgende willen we de laplacetransformatie vinden van $f_{\sigma_{i,i-1}}(t)$. We werken eerst uit dat,

$$\begin{aligned} \int_0^{\infty} e^{-st} e^{-(\lambda + \omega_i)t} dt &= \int_0^{\infty} e^{-(\lambda + \omega_i + s)t} dt \\ &= \left[-\frac{1}{\lambda + \omega_i + s} e^{-(\lambda + \omega_i + s)t} \right]_0^{\infty} \\ &= \frac{1}{\lambda + \omega_i + s}. \end{aligned} \quad (3.24)$$

Dit gebruiken we om de laplacetransformatie, zie vergelijking (2.22), van vergelijking (3.23) te vinden. Samen met de laplacetransformatie eigenschap van een convolutie, krijgen we dan,

$$\hat{f}_{\sigma_{i,i-1}}(s) = \frac{\omega_i}{\lambda + \omega_i + s} + \frac{\lambda}{\lambda + \omega_i + s} \hat{f}_{\sigma_{i+1,i}}(s) \hat{f}_{\sigma_{i,i-1}}(s). \quad (3.25)$$

¹²De transformatie van een van beide functies door de andere, met als resultaat de oppervlakte van de overlap van beide functies, waarbij de tweede functie verschuift.

Dit omschrijven en gebruikmaken van de voortgezette breuk, geeft

$$\begin{aligned}
\hat{f}_{\sigma_{i,i-1}}(s) &= \frac{\omega_i}{\lambda + \omega_i + s - \lambda \hat{f}_{\sigma_{i+1,i}}(s)} \\
&= \frac{\omega_i}{\lambda + \omega_i + s - \frac{\lambda \omega_{i+1}}{\lambda + \omega_{i+1} + s - \lambda \hat{f}_{\sigma_{i+2,i+1}}(s)}} \\
&= \frac{\omega_i}{\lambda + \omega_i + s - \frac{\lambda \omega_{i+1}}{\lambda + \omega_{i+1} + s - \frac{\lambda \omega_{i+2}}{\lambda + \omega_{i+2} + s - \dots}}} \\
&= -\frac{1}{\lambda} \mathcal{K}_{k=i}^{\infty} \frac{-\lambda \omega_k}{\lambda + \omega_k + s}.
\end{aligned} \tag{3.26}$$

□

We combineren het resultaat uit deze propositie met vergelijking (3.20) en vinden de laplacetransformatie van de kansdichtheidsfunctie van de eerste-doorgangstijd σ_b :

$$\hat{f}_{\sigma_b}(s) = \prod_{i=1}^b \hat{f}_{\sigma_{i,i-1}}(s) = \prod_{i=1}^b \left(-\frac{1}{\lambda} \mathcal{K}_{k=i}^{\infty} \frac{-\lambda \omega_k}{\lambda + \omega_k + s} \right). \tag{3.27}$$

3.1.3 Uitvoeringskans limietorder op de beste aanbodprijs

We gebruiken vergelijking (3.27) voor het vinden van de uitvoeringskans van een limietorder die op de beste aanbodprijs geplaatst is. We bepalen deze kans conditioneel op de toestand van het LOB op $t = 0$, met deze 'toestand' bedoelen we het aantal uitstaande limietorders op elk prijsniveau op moment $t = 0$. We gebruiken de volgende notaties voor de hoeveelheid limietorders op de beste vraag- en aanbodprijs resp. $X_A = X_{p_A(\cdot)}(\cdot)$ en $X_B = |X_{p_B(\cdot)}(\cdot)|$.

We gaan de uitvoeringskans berekenen van een limietorder op de beste aanbodprijs, door de kans te berekenen dat de eerste-doorgangstijd van een toestand q_B naar 0, korter is dan de tijd voordat de middenprijs voor het eerst beweegt. We stellen hierbij dat q_B de initiële (op $t = 0$) hoeveelheid orders is op de beste aanbodprijs, dus $X_B(0) = q_B$. 'Voordat de middenprijs beweegt', stellen we omdat de beste vraag-/aanbodprijs dan veranderd, en we willen kijken naar de uitvoeringskans van een limietorder op de beste prijs. Verder nemen we aan dat die order, op de beste aanbodprijs, niet geannuleerd wordt. Laat NC_b (*Never Cancelled at best-bid price*) de gebeurtenis zijn dat een order, op de beste aanbodprijs op tijd t , niet geannuleerd wordt. Laat T de tijd zijn waarop de middenprijs voor het eerst beweegt. Dan geven we deze kans weer als volgt

$$\mathbb{P}[\sigma_{q_B} < T | X_A(0) = q_A, X_B(0) = q_B, p_S(0) = S, NC_b], \tag{3.28}$$

waarbij σ_{q_B} de eerste-doorgangstijd is van een toestand q_B tot 0. We stellen de initiële hoeveelheid orders op de beste vraagprijs, op een waarde q_A . We stellen de initiële hoeveelheid orders op de beste aanbodprijs, op een waarde q_B . En de initiële bid-ask spread zetten we op een waarde S , wat een grootte van S maal de ticksize betekent.

Eerste bekijken we een kort lemma wat we gebruiken bij het bewijs van de daarop volgende propositie, die ons de uitvoeringskans (3.28) gaat geven.

Lemma 3.1. *Laat f en F de kansdichtheid en CDF van een kansvariabele X , respectievelijk. De laplacetransformatie \hat{F} van de CDF F , wordt gegeven door*

$$\hat{F}(s) = \frac{1}{s} \hat{f}(s), \tag{3.29}$$

waar $\hat{f}(s)$ de laplacetransformatie is van de kansdichtheid f .

Bewijs. We weten dat $F' = f$,

$$\hat{F}(s) = \int_{-\infty}^{\infty} e^{-st} F(t) dt \text{ en } \hat{F}'(s) = \hat{f}(s) = \int_{-\infty}^{\infty} e^{-st} f(t) dt.$$

Door partieel integreren vinden we nu,

$$\begin{aligned} \hat{f}(s) &= \hat{F}'(s) = \int_{-\infty}^{\infty} e^{-st} F'(t) dt \\ &= [e^{-st} F(t)]_{-\infty}^{\infty} + s \int_{-\infty}^{\infty} e^{-st} F(t) dt \\ &= s \int_{-\infty}^{\infty} e^{-st} F(t) dt = s \hat{F}(s). \end{aligned} \tag{3.30}$$

□

De kans gegeven in vergelijking (3.28) kunnen we nu bepalen door middel van de volgende propositie.

Propositie 3.2. *Laat \tilde{X}_A en \tilde{X}_B , onafhankelijke geboorte- en sterfteprocessen met geboorte-intensiteiten $\lambda(S)$ en sterfte-intensiteiten $\mu + i\theta(S)$, $i \geq 1$, zo dat $\forall 0 \leq t \leq T$, $\tilde{X}_A = X_A(t)$ en $\tilde{X}_B = X_B(t)$. Laat σ_A en σ_B de eerste-doorgangstijden van \tilde{X}_A en \tilde{X}_B tot 0 zijn, resp. De laplacetransformatie van de kansdichtheidsfunctie van σ_{q_j} is dan*

$$\hat{f}_{\sigma_{q_j}}^S(s) = \prod_{i=1}^{q_j} \left(-\frac{1}{\lambda(S)} \mathcal{K}_{k=i}^{\infty} \frac{-\lambda(S)(\mu + k\theta(S))}{\lambda(S) + \mu + k\theta(S) + s} \right), \tag{3.31}$$

met $j \geq 1$. De laplacetransformatie van de kansdichtheidsfunctie van σ_{q_B} is

$$\hat{g}_{\sigma_{q_B}}^S(s) = \prod_{i=1}^{q_B} \frac{\mu + (i-1)\theta(S)}{\mu + (i-1)\theta(S) + s}, \tag{3.32}$$

waar $q_B \geq 0$. Laat

$$\Lambda_S = \sum_{\delta=1}^{S-1} \lambda(\delta), \tag{3.33}$$

dan is de kans (3.28) gegeven door de inverse van de laplacetransformatie van de CDF van σ_{q_B} en σ_A

$$\hat{F}_{\sigma_{q_B}, \sigma_A}^S(s) = \frac{1}{s} \hat{g}_{\sigma_{q_B}}^S(s) (\hat{f}_{\sigma_A}^S(2\Lambda_S - s) + \frac{2\Lambda_S}{2\Lambda_S - s} (1 - \hat{f}_{\sigma_A}^S(2\Lambda_S - s))), \tag{3.34}$$

bekeken op 0. Wanneer we de bid-ask spread nu voor het gemak de grootte van 1 ticksize maken, oftewel $S = 1$, zien we dat de kans (3.28) gelijk is aan de inverse van

$$\hat{F}_{\sigma_{q_B}, \sigma_A}^1(s) = \frac{1}{s} \hat{g}_{\sigma_{q_B}}^1(s) \hat{f}_{\sigma_A}^1(-s). \tag{3.35}$$

Bewijs. Om de laplacetransformatie van het geboorte- en sterfteproces $f_{\sigma_j}^S(s)$ te vinden, gebruiken we de in vergelijking (3.27) gegeven laplacetransformatie van de kansdichtheidsfunctie van σ_b , $\hat{f}_{\sigma_b}^S(s)$. Door het invullen van $\lambda = \lambda(S)$, $\omega_k = \mu + k\theta(S)$ en $b = q_j$, vinden we $\hat{f}_{\sigma_{q_j}}^S(s)$ zoals in vergelijking (3.31). Merk op dat we hier de intensiteiten uitbreiden zoals we dat in sectie 3.1.1 hebben gezien. De aankomsten van limietorders geven we weer met intensiteit λ en is afhankelijk van de bid-ask

spread. De aankomsten van marktorders geven we weer met intensiteit μ . De annuleringen geven we weer met intensiteit θ die bid-ask spread afhankelijk is en wordt vermenigvuldigd met het aantal orders (de toestand) op dat moment. De aankomsten van marktorders en annuleringen worden in een 'sterfte'-intensiteit ω gecombineerd omdat ze beide het aantal orders in het LOB laten afnemen.

Om de laplacetransformatie van het puur-sterfteproces¹³ $g_{\sigma_{q_B}}(s)$ te vinden, gaan we een vergelijkbare methode gebruiken als we bij het bewijs van propositie 3.1 hebben gedaan. We bestuderen nu een puur-sterfteproces in plaats van een geboorte- en sterfteproces, omdat volgens de Fifo-wachtrij een order die aankomt op hetzelfde of een lager prijsniveau dan de beste aanbodprijs geen invloed heeft op de uitvoeringskans, een geboorte heeft dus geen invloed. De kansdichtheidsfunctie van de eerste-doorgangstijd van toestand i naar $i-1$ voor $1 \leq i \leq q_B$ van het pure-sterfteproces is ook exponentieel verdeeld maar nu met een intensiteit ω_i , want enkel 'sterfte'. Dus volgt

$$g_{\sigma_{i,i-1}}(t) = \omega_i e^{-\omega_i t}. \quad (3.36)$$

Hiervan is de laplacetransformatie dan

$$\hat{g}_{\sigma_{i,i-1}}(s) = \int_0^\infty \omega_i e^{-\omega_i t} e^{-st} dt = \int_0^\infty \omega_i e^{-(\omega_i+s)t} dt = \frac{\omega_i}{\omega_i + s}. \quad (3.37)$$

Omdat we hebben gesteld dat de limietorder op de beste aanbodprijs niet geannuleerd wordt, is de annuleringsintensiteit in toestand i gelijk aan $(i-1)\theta(S)$ en dus $\omega_i = \mu + (i-1)\theta(S)$. Uit het feit dat $\sigma_{q_B} = \sigma_{q_B, q_B-1} + \dots + \sigma_{1,0}$ en de onafhankelijkheid van de eerste-doorgangstijden, volgt dat

$$\hat{g}_{\sigma_{q_B}}^S(s) = \prod_{i=1}^{q_B} \hat{g}_{\sigma_{i,i-1}}^S(s) = \prod_{i=1}^{q_B} \frac{\mu + (i-1)\theta(S)}{\mu + (i-1)\theta(S) + s}. \quad (3.38)$$

Als $S = 1$, dus de bid-ask spread is zo groot als één ticksize, dan kunnen er geen limietorders binnen de bid-ask spread geplaatst worden. Dan is de uitvoeringskans van een aanbodorder voordat de middenprijs beweegt gelijk aan

$$\mathbb{P}[\sigma_{q_B} < \sigma_A] = \mathbb{P}[\sigma_{q_B} - \sigma_A < 0]. \quad (3.39)$$

Waar σ_A de eerste-doorgangstijd is van geboorte- en sterfteproces \tilde{X}_A tot nul, $A = q_j$ voor een bepaalde waarde j . Want als σ_{q_B} eerder naar 0 gaat dan σ_A , betekent dit dat het puur-sterfteproces, dat kan worden geïnterpreteerd als de tijd die het kost voordat er een verandering plaatsvindt in het aantal beschikbare aanbodorders, eerder eindigt dan het geboorte- en sterfteproces, dat de tijd weergeeft voordat er een verandering plaatsvindt in het aantal beschikbare orders (zowel vraag als aanbod). Dus dit betekent dat de kans op uitvoering van een aanbodorder hoger is als het puur-sterfteproces eerder eindigt dan het geboorte- en sterfteproces. Als het puur-sterfteproces eerder eindigt, vindt de uitvoering van de aanbodorder plaats voordat er een significante verandering in de marktprijzen is opgetreden, zoals weergegeven door de beweging van de middenprijs. Uit het feit dat

$$\mathbb{P}[\sigma_{q_B} - \sigma_A < 0] = F_{\sigma_{q_B}, \sigma_A}^1(0), \quad (3.40)$$

volgt dat we de kans (3.28) kunnen berekenen door de inverse van $\hat{F}_{\sigma_{q_B}, \sigma_A}^1(s)$ te berekenen op 0. Hier is $\hat{F}_{\sigma_{q_B}, \sigma_A}^1(s)$ de gezamenlijke CDF van de eerste-doorgangstijden σ_{q_B} en σ_A . Uit lemma 3.1 en de onafhankelijkheid van de eerste doorgangstijden volgt dat deze gezamenlijke CDF als volgt is,

$$\hat{F}_{\sigma_{q_B}, \sigma_A}^1(s) = \frac{1}{s} \hat{f}_{\sigma_{q_B}, \sigma_A}^1(s) = \frac{1}{s} \hat{g}_{\sigma_{q_B}}^1(s) \hat{f}_{\sigma_A}^1(-s),$$

¹³Een puur-sterfteproces is een geboorte- en sterfteproces waar $\lambda = 0$.

waarbij $f_{\sigma_{q_B}, \sigma_A}$ de gezamenlijke kansdichtheidsfunctie van de eerste-doorgangstijden σ_{q_B} en σ_A is.

Als $S > 1$ dan is de uitvoeringskans van een aanbodorder, voordat de middenprijs beweegt gelijk aan

$$\mathbb{P}[\sigma_{q_B} < \sigma_A \wedge \tau_B \wedge \tau_A], \quad (3.41)$$

waar τ_B en τ_A de eerste tijden zijn dat een aanbod- of vraaglimietorder (resp.) in de bid-ask spread geplaatst worden. In het geval dat de bid-ask spread groter is dan één ticksize, kunnen limietorders dus in de bid-ask spread geplaatst worden en de middenprijs verandert hierdoor, vandaar dat het nu verwerkt zit in de uitvoeringskans die we zoeken.

We definiëren kansvariabele $Z = \tau_B \wedge \tau_A$ en we zien dat deze exponentieel verdeeld is met een intensiteit $2\Lambda_S$, waarbij $\Lambda_S = \sum_{\delta=1}^{S-1} \lambda(\delta)$. Omdat het gaat om een limietorder aankomst (geboorte), zien we λ , en dit gebeurt binnen de bid-ask spread dus loopt de sommatie tot $S - 1$.

Dan volgt de laplacetransformatie voor $\sigma_A \wedge \tau_B \wedge \tau_A = \sigma_A \wedge Z$,

$$\hat{f}_{\sigma_A \wedge Z}^S = \hat{f}_{\sigma_{q_B}}^1 (2\Lambda_S + s) + \frac{2\Lambda_S}{2\Lambda_S + s} (1 - \hat{f}_{\sigma_{q_B}}^1 (2\Lambda_S + s)), \quad (3.42)$$

waarbij $\hat{f}_{\sigma_{q_B}}^1$ is zoals in vergelijking (3.31) voor $j = B$. De kans in (3.41) is dus gelijk aan $F_{\sigma_{q_B}, \sigma_A \wedge Z}^S$ en we vinden de laplacetransformatie hiervan weer met behulp van lemma 3.1 als volgt,

$$\hat{F}_{\sigma_{q_B}, \sigma_A \wedge Z}^S(s) = \frac{1}{s} \hat{g}_{\sigma_{q_B}}^S(s) \hat{f}_{\sigma_A \wedge Z}^S(s) = \frac{1}{s} \hat{g}_{\sigma_{q_B}}^S(s) (\hat{f}_{\sigma_A}^S(2\Lambda_S - s) + \frac{2\Lambda_S}{2\Lambda_S - s} (1 - \hat{f}_{\sigma_A}^S(2\Lambda_S - s))). \quad (3.43)$$

De kans (3.28) kan nu berekend worden door de inverse van de laplacetransformatie van vergelijking (3.43) te berekenen op 0. \square

Met behulp van het Bromwich integraal (2.24), vinden we de inverse van vergelijking (3.34)

$$F_{\sigma_{q_B}, \sigma_A}^S(t) = \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} e^{st} \hat{F}_{\sigma_{q_B}, \sigma_A}^S(s) ds. \quad (3.44)$$

En dus de uitvoeringskans van een limietorder op de beste aanbodprijs voordat de middenprijs beweegt, gegeven dat de limietorder niet geannuleerd wordt, in een LOB met bid-ask spread S , is gelijk aan :

$$\begin{aligned} \mathbb{P}[\sigma_{q_B} < T | X_A(0) = q_A, X_B(0) = q_B, p_S(0) = S, NC_b] &= F_{\sigma_{q_B}, \sigma_A}^S(0) \\ &= \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} \hat{F}_{\sigma_{q_B}, \sigma_A}^S(s) ds \\ &= \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} \frac{1}{s} \hat{g}_{\sigma_{q_B}}^S(s) (\hat{f}_{\sigma_A}^S(2\Lambda_S - s) + \frac{2\Lambda_S}{2\Lambda_S - s} (1 - \hat{f}_{\sigma_A}^S(2\Lambda_S - s))) ds. \end{aligned} \quad (3.45)$$

En voor een bid-ask spread van 1 maal de ticksize:

$$\begin{aligned} \mathbb{P}[\sigma_{q_B} < T | X_A(0) = q_A, X_B(0) = q_B, p_S(0) = 1, NC_b] &= F_{\sigma_{q_B}, \sigma_A}^1(0) \\ &= \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} \hat{F}_{\sigma_{q_B}, \sigma_A}^1(s) ds \\ &= \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} \frac{1}{s} \hat{g}_{\sigma_{q_B}}^1(s) \hat{f}_{\sigma_A}^1(-s) ds. \end{aligned} \quad (3.46)$$

Hier is i het imaginaire getal.

3.1.4 Activiteit LOB modelleren met poisson

Om de activiteit van een LOB te modelleren, kan het interessant zijn om te berekenen wat de kans is dat er een sprong (komen of gaan van limietorder) voorkomt in de volgende t' tijdseenheden. Uit vergelijking (2.9) volgt de kansdichtheid van een increment van een simpel poissonproces. Dus voor een constante geboorte-intensiteit λ , toestand afhankelijke sterfte-intensiteit ω_j en een spronggrootte 1 is de kans op een sprong in de volgende t' tijdseenheden gelijk aan

$$\mathbb{P}(N(t+t') - N(t) = 1) = (\lambda + \omega_j)t'e^{-(\lambda+\omega_j)t'}, \quad (3.47)$$

en

$$\mathbb{P}(N(t+t') - N(t) \geq 1) = 1 - \mathbb{P}(N(t+t') - N(t) = 0) = 1 - e^{-(\lambda+\omega_j)t'}. \quad (3.48)$$

Hier is j de toestand van het LOB, oftewel het totaal aantal orders, op moment t .

Wanneer we ook rekening willen houden met de spronggrootte kijken we dus naar het samengestelde poissonproces, zie sectie 2.1.1. Uit vergelijking (2.18) volgt de moment-genererende functie voor het samengestelde poissonproces met vaste spronggrootte y . Hieruit volgt de laplacetransformatie van het samengestelde poissonproces met vaste spronggrootte y ,

$$\hat{f}_{yN(t)}(s) = e^{\lambda t(e^{-sy}-1)}.$$

Dan vinden we door inverteren met het Bromwich integraal (2.24), de kansdichtheid

$$f_{yN(t)}(x) = \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} e^{sx} e^{\lambda t(e^{-sy}-1)} ds.$$

Vanwege de geheugenloosheid van de exponentiële verdeling in het poissonproces zien we net zoals in sectie 2.1, dat de incrementen van een samengesteld poissonproces de volgende kansdichtheidsfunctie hebben,

$$\mathbb{P}(N(t+t') - N(t) = x) = \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} e^{sx} e^{\lambda((t+t')-t')(e^{-sy}-1)} ds = \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} e^{sx+\lambda((t+t')-t')(e^{-sy}-1)} ds.$$

Dus voor een constante geboorte-intensiteit λ , toestand afhankelijke sterfte-intensiteit ω_j en spronggrootte $y = \ell$, is de kans op een sprong in de volgende t' tijdseenheden gelijk aan

$$\mathbb{P}(N(t+t') - N(t) = 1 | \lambda = \lambda + \omega_j, y = \ell) = \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} e^{s+(\lambda+\omega_j)t'(e^{-s\ell}-1)} ds \quad (3.49)$$

en

$$\begin{aligned} \mathbb{P}(N(t+t') - N(t) \geq 1 | \lambda = \lambda + \omega_j, y = \ell) &= 1 - \mathbb{P}(N(t+t') - N(t) = 0 | \lambda = \lambda + \omega_j, y = \ell) \\ &= 1 - \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} e^{(\lambda+\omega_j)t'(e^{-s\ell}-1)} ds \end{aligned} \quad (3.50)$$

Hier is j de toestand van het LOB, oftewel het totaal aantal orders, op moment t , en i het imaginaire getal.

3.2 Geboorte- en sterfteproces met hawkes

In het onderzoek van Gao, Zhou en Zhu[11] zien we dat dark pool trading wordt benaderd als een geboorte- en sterfteproces met behulp van de laplacetransformatie van een hawkesproces. Wanneer we een vergelijkbare uitvoeringskans willen berekenen zoals we zojuist met het PP hebben gedaan, zien we het volgende in dit onderzoek:

De tijd tot de eerste complete uitvoering van een uitstaande order van grootte $x > 0$ in dark pool trading, wordt weer gegeven als σ_x . We zien dat dit hetzelfde is als de tijd die ons limietorderboek nodig heeft om van een toestand $x > 0$ naar nul te gaan. Oftewel, de eerste-doorgangstijd van een toestand $x > 0$ naar nul geven we op basis van een hawkesproces weer als volgt:

$$\sigma_x := \inf\{t \geq 0 : L(t) \geq x\}. \quad (3.51)$$

En hieruit volgt de kans

$$\mathbb{P}(\sigma_x \leq t) = \mathbb{P}(L(t) \geq x) = 1 - F_{L(t)}(x).$$

De laplacetransformatie van $L(t)$, zoals we die in gevolg 2.1 hebben gevonden, kunnen we nu inverteren om zo de uitvoeringskans van een limietorder op de beste aanbodprijs te vinden.

3.2.1 Kansdichtheid eerste-doorgangstijd

We inverteren de laplacetransformatie van $L(t)$ gevonden in gevolg 2.1 met behulp van het Bromwich integraal (zie vergelijking (2.24)). Zo vinden we de kansdichtheidsfunctie van $L(t)$ voor elke $v \in \mathbb{C}$ waarvan het reële deel positief is,

$$f_{L(t)}(x) = \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} e^{sx} e^{\int_0^s \lambda^*(s-s')(H_L(s')-1)ds'} ds, \quad (3.52)$$

waar de functie H_L de unieke oplossing van de integraal vergelijking

$$H_L(t) = \mathbb{E}[e^{-v t_1 + \int_0^t l_1 \phi(s)(H_L(t-s)-1)ds}], \quad (3.53)$$

met $|H(t)| \leq 1$.

Dan volgt hieruit de CDF van $L(t)$:

$$\begin{aligned} F_{L(t)}(x) &= \int_0^x f_{L(t)}(s'') ds'' = \int_0^x \left[\frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} e^{ss''} e^{\int_0^s \lambda^*(s-s')(H_L(s')-1)ds'} ds \right] ds'' \\ &= \frac{1}{2\pi i} \int_0^x \int_{\gamma-i\infty}^{\gamma+i\infty} e^{ss'' + \int_0^s \lambda^*(s-s')(H_L(s')-1)ds'} ds ds'' \\ &= \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} \int_0^x e^{ss'' + \int_0^s \lambda^*(s-s')(H_L(s')-1)ds'} ds'' ds \\ &= \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} \left[\frac{1}{s} e^{ss'' + \int_0^s \lambda^*(s-s')(H_L(s')-1)ds'} \right]_0^x ds \\ &= \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} \frac{1}{s} e^{sx + \int_0^s \lambda^*(s-s')(H_L(s')-1)ds'} - \frac{1}{s} e^{\int_0^s \lambda^*(s-s')(H_L(s')-1)ds'} ds \\ &= \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} \frac{1}{s} (e^{sx} - 1) e^{\int_0^s \lambda^*(s-s')(H_L(s')-1)ds'} ds. \end{aligned} \quad (3.54)$$

Dus

$$\mathbb{P}(\sigma_x \leq t) = 1 - \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} \frac{1}{s} (e^{sx} - 1) e^{\int_0^s \lambda^*(s-s')(H_L(s')-1)ds'} ds.$$

Wanneer we nu voor deze x een waarde kiezen wat gelijk is aan de hoeveelheid orders op bijvoorbeeld de beste prijs, kunnen we de kans berekenen dat deze toestand naar nul gaat voor een bepaalde tijd t . Vergelijkbaar met wat we in sectie 3.1.3 hebben gedaan met het PP.

3.2.2 Activiteit LOB modelleren met hawkes

We kunnen met hawkes ook kijken wat de kans is dat er in de volgende t' tijdseenheden een sprong voorkomt, gegeven dat we een initiële spronggrootte van $l_1 = \ell$ hebben. Zo'n sprong kan dus het komen of gaan van een limietorder zijn, oftewel we kunnen zo de activiteit van het LOB in de gaten houden.

We willen dus bepalen

$$\mathbb{P}(N(t+t') - N(t) = 1 | N(t) = 1, l_1 = \ell), \quad (3.55)$$

en

$$\mathbb{P}(N(t+t') - N(t) \geq 1 | N(t) = 1, l_1 = \ell) = 1 - \mathbb{P}(N(t+t') - N(t) = 0 | N(t) = 1, l_1 = \ell). \quad (3.56)$$

Om deze conditionele sprongkansen te kunnen berekenen gebruiken we de laplacetransformatie van $N(t)$ uit gevolg 2.2 en zien we:

$$\begin{aligned} & \mathbb{P}(N(t+t') - N(t) = 0 | N(t) = 1, l_1 = \ell) \\ &= \frac{\int_0^t \lambda^*(a_1) e^{-\int_0^{a_1} \lambda^*(s) ds} e^{-\int_{a_1}^t (\lambda^*(s) + \phi(s-a_1)\ell) ds} e^{-\int_t^{t+t'} (\lambda^*(s) + \phi(s-a_1)\ell) ds} da_1}{\int_0^t \lambda^*(a_1) e^{-\int_0^{a_1} \lambda^*(s) ds} e^{-\int_{a_1}^t (\lambda^*(s) + \phi(s-a_1)\ell) ds} da_1}, \end{aligned} \quad (3.57)$$

en

$$\begin{aligned} & \mathbb{P}(N(t+t') - N(t) = 1 | N(t) = 1, l_1 = \ell) \\ &= \left(\int_0^t \lambda^*(a_1) e^{-\int_0^{a_1} \lambda^*(s) ds} e^{-\int_{a_1}^t (\lambda^*(s) + \phi(s-a_1)\ell) ds} da_1 \right)^{-1} \\ & \cdot \int_0^t [\lambda^*(a_1) e^{-\int_0^{a_1} \lambda^*(s) ds} e^{-\int_{a_1}^{t+T} (\lambda^*(s) + \phi(s-a_1)\ell) ds} \\ & \cdot \int_t^{t+t'} (\lambda^*(s) + \phi(s-a_1)\ell) \mathbb{E}[e^{-\int_0^{t'+t-s} \ell \phi(s') ds'}] ds] da_1. \end{aligned} \quad (3.58)$$

Waar λ^* de basisintensiteit is, a_1 de eerste aankomsttijd en $\phi(\cdot)$ de geheugenkern.

Een uitvoerige afleiding van deze twee conditionele kansen is te vinden in de appendix van [11].

4 Conclusie

In sectie 3.1 hebben we het geboorte- en sterfproces uitgewerkt op basis van het PP, en gezien hoe we hiermee de uitvoeringskans van een limietorder op de beste aanbodprijs kunnen bepalen. Dit geboorte- en sterfproces berekende die uitvoeringskans door middel van de laplacetransformatie van een eerste-doorgangstijd. Maar zoals we in sectie 2.2.1 hebben gezien, heeft niet alleen het poissonproces zo een laplacetransformatie, maar ook het hawkesproces heeft dit.

Wat zijn de voor- of nadelen van het hawkesproces in plaats van het poissonproces in het kader van een geboorte- en sterfproces, voor het berekenen van de uitvoeringskans van een limietorder?

Een eerste verschil wat opvalt is dat het hawkesproces meer flexibiliteit biedt dan het poissonproces, dit komt door verschillende intensiteiten die de geheugeneigenschappen van een LOB regelen. We bedoelen met zo een zelfopwekkende-eigenschap dat een bepaalde gebeurtenis in het LOB invloed heeft op het tijdstip en de intensiteit van toekomstige vergelijkbare gebeurtenissen. Dit zien we in het LOB bijvoorbeeld doordat de plaatsing van een order de waarschijnlijkheid van toekomstige orderplaatsing vergroot.

Voorbeeld 4.1. *Stel er wordt in het LOB een aanbodorder uitgevoerd op een bepaalde prijs p . Dan zijn er twee zelfopwekkende gebeurtenissen die kunnen voorkomen:*

(i) Andere handelaren zien dat er een prijsbeweging heeft plaatsgevonden en verwachten nu verdere prijsbewegingen. Hierom plaatsen zij nu ook orders. De zelfopwekkende eigenschap van een LOB zorgt nu dus met één initiële transactie voor de toename van het aantal orders.

(ii) Marktmakers willen hun positie in het LOB niet kwijt en zullen na deze prijsbeweging dus extra orders bijplaatsen om zo hun plaats in het LOB terug te krijgen.

Deze zelfopwekkende eigenschap kan zo leiden tot samenklontering van transacties en andere activiteiten op bepaalde prijslevels en het kan de impact van individuele orders vergroten en zo leiden tot periodes van verhoogde activiteit. Het hawkesproces kan dus beter aansluiten op de feitelijke waarnemingen in marktgedrag waarin orderplaatsingen vaak samenklonteren.

Het poissonproces behandelt de plaatsing van orders als willekeurig en onafhankelijk met een constante intensiteit. Deze onafhankelijkheid leidt tot een eenvoudiger model om te gebruiken en te analyseren. Het kan dus zo zijn dat het hawkesproces meer rekenkracht vereist. Ook heeft het hawkesproces meer gegevens nodig om effectief de grotere hoeveelheid parameters te schatten. Door de flexibiliteit van het hawkesproces bestaat er ook nog het risico op overfitting: het model wordt te nauw afgestemd op de data en generaliseert hierdoor minder de nieuwe gegevens.

De keuze tussen het hawkesproces en het poissonproces hangt dus af van het evenwicht tussen modelcomplexiteit en de mate van realiteit die we in ons model willen opnemen. Verder heeft de keuze te maken met:

(i) Welke eigenschappen het LOB heeft: Het hawkesproces is beter als het LOB gedrag zelfopwekkende patronen, samenklontering van activiteiten en gevoeligheid voor eerdere gebeurtenissen vertoont. Wanneer het gedrag van het LOB meer willekeurig en onafhankelijk is kun je voor een eenvoudiger model gaan, zoals het poissonproces.

En (ii) welke eigenschappen je wilt onderzoeken: Het hawkesproces is beter als je onderzoek wilt doen naar de gevolgen van één transactie op toekomstige transacties en het poissonproces kan beter worden gebruikt om de veranderingen van het aantal actieve orders op verschillende prijslevels te bestuderen.

Beide processen zijn dus bruikbaar voor het modelleren van een LOB met een geboorte- en sterfep proces. Dit is slechts een enkel voorbeeld van een stochastisch proces om de dynamiek van een limietorderboek accuraat mee te modelleren. Het antwoord op onze onderzoeksvraag:

Kunnen we de dynamiek van een limietorderboek accuraat modelleren met een stochastisch proces?

is dus, ja!

5 Bibliografie

Referenties

- [1] Candidate Number: 1037839. Hawkes process-driven models for limit order book dynamics. Master's thesis, University of Oxford, Master of Science in Mathematical and Computational Finance, 2020. https://www.maths.ox.ac.uk/system/files/attachments/Hawkes%20Process-Driven%20Models%20for%20Limit%20Order%20Book%20Dynamics_0.pdf.
- [2] L.A. Souto Arias, P. Cirillo, and C.W. Oosterlee. A new self-exciting jump-diffusion process for option pricing. 2023. <https://arxiv.org/abs/2205.13321>.
- [3] A. Arroyo, A. Cartea, F. Moreno-Pino, and S. Zohren. Deep attentive survival analysis in limit order books: Estimating fill probabilities with convolutional-transformers, 2023. <https://doi.org/10.48550/arXiv.2306.05479>.
- [4] Vijeyata Chauhana and Pankaj Kumar Srivastavab. Computational techniques based on runge-kutta method of various order and type for solving differential equations. *International Journal of Mathematical, Engineering and Management Sciences Vol. 4, No. 2*, pages 375–386, 2019. <https://dx.doi.org/10.33889/IJMMS.2019.4.2-030>.
- [5] J. Cho and E. Nelling. The probability of limit-order execution. *Financial Analyst Journal* 56.5, 2000. <https://doi.org/10.2469/faj.v56.n5.2387>.
- [6] R. Cont, S. Stoikov, and R. Talreja. A stochastic model for order book dynamics. *Operations Research* 58.3, (0030-364X):549–563, 2010. <https://pubsonline.informs.org/doi/10.1287/opre.1090.0780>.
- [7] D. Duffie, J. Pan, and K. Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, pages 1343–1376, 2000. <https://onlinelibrary.wiley.com/doi/abs/10.1111/1468-0262.00164>.
- [8] L. Lorentzen en H. Waadeland. *Continued Fractions. 2nd ed. Vol. 1*. Number 978-90-78677-07-9. Atlantis Press, 2008.
- [9] E. Errais, K. Giesecke, and L. Goldberg. Affine point processes and portfolio credit risk. pages 642–665, 2010. <https://epubs.siam.org/doi/10.1137/090771272>.
- [10] The Dutch Authority for the Financial Markets (AFM). Afm market watch edition 8: Algorithmic trading. tech. rep., 2023. <https://www.afm.nl/en/sector/themas/beurzen-en-effecten/afm-market-watch>.
- [11] X. Gao, X. Zhou, and L. Zhu. Transform analysis for hawkes processes with applications in dark pool trading. 2017. <https://arxiv.org/abs/1710.01452>.
- [12] A.W. Lo, A. Craig MacKinlay, and J. Zhang. Econometric models of limit-order executions. *Journal of Financial Economics* 65.1, (0304-405X):31–71, 2002. <https://www.sciencedirect.com/science/article/abs/pii/S0304405X02001344?via%3Dihub>.
- [13] F.J.P. Lokin. Fill probabilities in a limit order book with applications to foreign exchange spot markets. Master's thesis, Delft University of Technology Faculty of EEMCS, 2023. <https://repository.tudelft.nl/islandora/object/uuid%3A55b62810-8dc4-453b-b7df-4092935d5d1e>.
- [14] C. Maglaras, C.C. Moallemi, and M. Wang. A deep learning approach to estimating fill probabilities in a limit order book. *Quantitative Finance* 22.11, (14697696):1989—2003, 2022. <https://www.tandfonline.com/doi/full/10.1080/14697688.2022.2124189>.

- [15] S.E. Shreve. *Stochastic Calculus for Finance II, Continuous-Time Models*. Springer Finance, 2004.