Conditional Tail Sampling

for General (Marked) Point Processes

Baeho Kim
baehokim@korea.ac.kr

Professor of Finance & Business Analytics
- Korea University Business School

Visiting Scholar
- CDAR, UC Berkeley

(Joint work with **Alex Shkolnik**)

Motivation

The problems of **event counting** within a given time horizon

- They naturally arise in a wide range of disciplines

  (e.g.), epidemiology, insurance claims, corporate defaults, ...

Motivation (cont.)

A **rare** event often refers to infrequently observable events

- It may have widespread impacts, potentially leading to the instability of **an entire system**

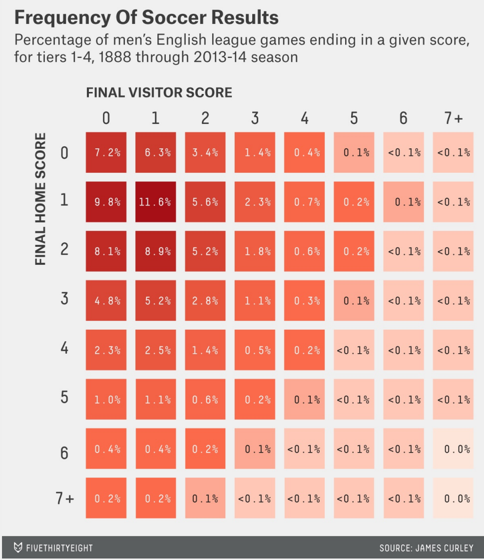
Rare events often involve **extreme losses** that fall far from the mean

- An accurate estimation of the distributional tail behavior is challenging but critical
- Inaccurate estimates can lead to suboptimal resource allocation or missed opportunities

A **rare** event is an event occurring with a **small** probability

- (Q1) What would be the likelihood that **a given (large) number of events** is observed by some horizon?

- (Q2) What would be the **expected consequence** under such circumstances?



FINAL HOME SCORE \ FINAL VISITOR SCORE	0	1	2	3	4	5	6	7+
0	7.2%	6.3%	3.4%	1.4%	0.4%	0.1%	<0.1%	<0.1%
1	9.8%	11.6%	5.6%	2.3%	0.7%	0.2%	0.1%	<0.1%
2	8.1%	8.9%	5.2%	1.8%	0.6%	0.2%	<0.1%	<0.1%
3	4.8%	5.2%	2.8%	1.1%	0.3%	0.1%	<0.1%	<0.1%
4	2.3%	2.5%	1.4%	0.5%	0.2%	<0.1%	<0.1%	<0.1%
5	1.0%	1.1%	0.6%	0.2%	0.1%	<0.1%	<0.1%	<0.1%
6	0.4%	0.4%	0.2%	0.1%	<0.1%	<0.1%	<0.1%	0.0%
7+	0.2%	0.2%	0.1%	<0.1%	<0.1%	<0.1%	<0.1%	0.0%

Monte Carlo (MC) Simulation

- Consider a counting process N
  - N_T counts the number of events on $[0, T]$ for some $T > 0$

- Tail probability of our interest is written as,

$$\varepsilon \triangleq \varepsilon_k(T) = \mathbf{P}(N_T \geq k) = \mathbf{E}\left(1_{\{N_T \geq k\}}\right)$$

  for some k sufficiently large (relative to T)

- The mathematical model is often too complicated to be solved by analytical methods for many real-world problems
  - When the underlying models are complex, **Monte Carlo (MC)** is the standard and useful method of assigning a numerical value to ε

- An advantage of MC methods is that they often generalize to related (and practically more important) problems
  - The Bayes' formula allows one to investigate the expected behavior of any random variable X conditional on a tail scenario

$$\mathbf{E}\left(X \,|\, N_T \geq k\right) = \frac{\mathbf{E}(X \cdot 1_{\{N_T \geq k\}})}{\varepsilon}$$

  - Just like ε, one can estimate $\mathbf{E}(X \cdot 1_{\{N_T \geq k\}})$ via MC sampling
  - No knowledge of the law of X conditional on the tail event is required

- For **rare event** problems, however, this MC simulation approach should not be implemented naively

- A **plain MC (pMC)** method is inaccurate and inefficient for small ε
  - The plain MC scheme needs $1/\varepsilon$ replications on average to get a single occurrence
  - The further we move into the tail, the smaller ε becomes
  - The pMC scheme generates very few nonzero samples, leading to inefficiencies

  $\Rightarrow$ It is not adequate when the ε is too small to observe the sufficient number of events that hits the rare event threshold

Drawbacks of plain MC (cont.)

- Let φ denote an unbiased estimator of ε $\quad \left(\varepsilon = \mathbf{E}(\varphi)\right)$
  - A standard measure to quantify the (in)efficiency is the **relative error**

$$\nu = \frac{\sqrt{\mathbf{Var}(\varphi)}}{\varepsilon}$$

  - The pMC estimator, $\varphi = 1_{\{N_T \geq k\}}$, yields

$$\nu = \sqrt{1/\varepsilon - 1} \to \infty \quad \text{as} \quad \varepsilon \downarrow 0$$

  - It becomes even more notorious for estimating $\mathbf{E}\,(X \,|\, N_T \geq k)$

- In the rare event regime when the denominator vanishes, our hope is to get a **bounded relative error (BRE)** in the limit
  - This has motivated the design of variance reduction methods which supply alternative estimators φ to reduce $\mathbf{Var}(\varphi)$
  - **Important sampling** aims to choose a 'good' distribution for variance reduction based on a measure-change argument

- A bit of intuition ($\mathbf{P}$ vs. $\mathbf{P}^\star$)
  - Let $\mathbf{P}$ be the reference probability measure for pMC
  - We construct a **(conditional) tail sampling** measure $\mathbf{P}^\star$ (distinct from $\mathbf{P}$) with

$$\mathbf{P}^\star (N_T \geq k) = 1$$

  - We draw samples of "adjusted" φ under $\mathbf{P}^\star$
    (⇒ A measure-change theory is needed for correcting the distortion)

- A bit of intuition ($\mathbf{P}$ vs. $\mathbf{P}^\star$)
  - Let $\mathbf{P}$ be the reference probability measure for pMC
  - We construct a **(conditional) tail sampling** measure $\mathbf{P}^\star$ (distinct from $\mathbf{P}$) with

$$\mathbf{P}^\star (N_T \geq k) = 1$$

  - We draw samples of “adjusted” φ under $\mathbf{P}^\star$
    (⇒ A measure-change theory is needed for correcting the distortion)

- A measure-change argument ($\mathbf{P} \to \mathbf{P}^\star$)
  - Estimate the expectation by **twisting** the probability measure

$$\begin{aligned} \varepsilon = \mathbf{P}(N_T \geq k) = \mathbf{E}\left(1_{\{N_T \geq k\}}\right) &= \int_\Omega 1_{\{N_T \geq k\}} d\mathbf{P} \\ &= \int_\Omega \underbrace{1_{\{N_T \geq k\}}}_{=1} \cdot \underbrace{\frac{d\mathbf{P}}{d\mathbf{P}^\star}}_{=\varphi} d\mathbf{P}^\star = \mathbf{E}^\star(\varphi) \end{aligned}$$

  - Can we find such a pair of $(\mathbf{P}^\star, \varphi)$? (⇒ Yes)
  - Can we achieve $\nu^\star = \frac{\sqrt{\mathbf{Var}^\star(\varphi)}}{\varepsilon} < \infty$ as $\varepsilon \downarrow 0$? (⇒ Yes, in many cases)

MAIN CONTRIBUTIONS

This study develops a novel, easy to simulate and fast MC estimator of rare event probabilities via **(conditional) Tail Sampling (cTS)** schemes

- It accommodates any model specification provided it can be simulated

Our algorithm provides meaningful efficiency gains by ensuring each simulated path hits the rare event with probability one

- It guarantees that none of the simulated paths will be wasted
- Our approach facilitates a reduction in the sampling error that often contaminates event time simulation estimators

The cTS approach possesses attractive properties for simulation

- Our method does not require the computation of any optimal or tuning parameter(s) and eliminates the need for numerical inversion procedures
- They can be extended to estimate conditional expectations on the tail event

$$\mathbf{E}\,(X\,|\,N_T \geq k) = \frac{\mathbf{E}(X \cdot 1_{\{N_T \geq k\}})}{\mathbf{E}(1_{\{N_T \geq k\}})} = \frac{\mathbf{E}^{\star}(\hat{X} \cdot \varphi)}{\mathbf{E}^{\star}(\varphi)}$$

We test our algorithms on a wide spectrum of applications using empirically motivated reduced-form models

- Our findings illustrate the superior performance of the proposed cTS scheme over plain MC

Overview

- Part I: Tail Probability Estimation

$$\varepsilon = \mathbf{P}\left(N_T \geq k\right) \quad \text{for large } k \text{ (relative to } T\text{)}$$

- Part II: Conditional Expectations on the Tail

$$\mathbf{E}\left(X \middle| N_T \geq k\right) = \frac{\mathbf{E}\left(X \cdot 1_{\{N_T \geq k\}}\right)}{\varepsilon}$$

- Part III: Applications to Finance & Insurance
  - Expected aggregate loss conditional on systemic credit events
  - (Ultra-short) Term structure of credit spreads on defaultable securities
  - Expected maximum drawdown conditional on catastrophic scenarios

Part I:
Tail Probability Estimation

Basic setup

- Fix a measurable space (Ω, Σ)
- We construct both $\mathbf{P}$ and $\mathbf{P}^\star$ on (Ω, Σ)
  - $\mathbf{P}$: the reference measure for plain MC (pMC)
  - $\mathbf{P}^\star$: the conditional tail-sampling measure for cTS

- Introduce a sequence of *ordered* stopping times $\{\tau_\ell\}_{\ell \geq 0}$ such that

$$0 = \tau_0 < \tau_1 < \tau_2 < \cdots, \text{ where } \lim_{\ell \to \infty} \tau_\ell = \infty$$

- Define the counting process

$$N_t = \sum_{\ell \geq 1} 1_{\{\tau_\ell \leq t\}}$$

- For integer $\ell \geq 1$, define $\theta_\ell = \tau_\ell - \tau_{\ell-1}$ (> 0) as the inter-arrival times

Basic setup (cont.)

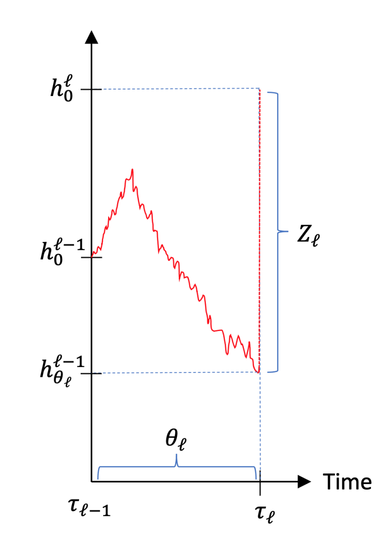
- Construct a sequence of nonnegative processes $\{h^\ell\}_{\ell \geq 1}$
  - Each h^ℓ is activated between event arrival times
- The sequence of variables $\{Z_\ell\}_{\ell \geq 1}$ specifies initial conditions as

$$h_0^\ell = \begin{cases} Z_\ell & \text{if } \ell = 1 \\ h_{\theta_\ell}^{\ell-1} + Z_\ell & \text{if } \ell \geq 2 \end{cases}$$

- Define λ as

$$\lambda_{t+\tau_{\ell-1}} = h_t^\ell \quad \text{for} \quad t \in [0, \theta_\ell)$$

- We take $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ with $\mathcal{F}_\infty \subseteq \Sigma$ to be the right-continuous (and completed) filtration generated by the pair (N, λ)



Time-change scheme (Meyer 1971)

- A complete probability space $(\Omega, \Sigma, \mathbf{P})$
- Introduce an i.i.d. sequence of standard exponential r.v.'s $\{\mathcal{E}_\ell\}_{\ell \geq 1}$
- Define the random (hitting) time θ_ℓ by

$$\theta_\ell = \inf\left\{t > 0 : \int_0^t h_s^\ell \, \mathrm{d}s \geq \mathcal{E}_\ell\right\}$$

- We view the h^ℓ as the (conditional) inter-arrival rate of the ℓ^{th} event of N; i.e., we refer to each process $\{h^\ell\}_{\ell \geq 1}$ an inter-arrival intensity of N
- The event counting process N admits λ as its intensity
  - The intensity represents the conditional mean arrival rate at each time for small $\Delta > 0$

$$\lambda_t = \lim_{\Delta \downarrow 0} \frac{\mathbf{E}\left(N_{t+\Delta} - N_t \mid \mathcal{F}_t\right)}{\Delta}$$

Our (conditional) tail sampling scheme (under $\mathbf{P}^\star$)

- We construct the cTS measure $\mathbf{P}^\star$ specific to the tail event $\{N_T \geq k\}$
  1. For some $\gamma > 0$, we construct $\mathbf{P}_\gamma$ as the probability measure on (Ω, Σ) under which N adopts the following values as its intensity:

     $$\begin{cases} \gamma & \text{for } \; t \in [0, \tau_k) \\ \lambda_t & \text{for } \; t \geq \tau_k \end{cases}$$

  2. For a fixed $T > 0$, we construct $\mathbf{P}^\star_\gamma$ as

     $$\mathbf{P}^\star_\gamma(\mathscr{A}) \triangleq \mathbf{P}_\gamma(\mathscr{A} \mid N_T \geq k) \;\text{ for all }\; \mathscr{A} \in \Sigma$$

  3. Under the Portmanteau theorem (for convergence of measures), we take the limiting measure

     $$\mathbf{P}^\star_\gamma \Rightarrow \mathbf{P}^\star \;\text{ as }\; \gamma \downarrow 0$$

Our (conditional) tail sampling scheme (under $\mathbf{P}^\star$)

- We construct the cTS measure $\mathbf{P}^\star$ specific to the tail event $\{N_T \geq k\}$
  1. For some $\gamma > 0$, we construct $\mathbf{P}_\gamma$ as the probability measure on (Ω, Σ) under which N adopts the following values as its intensity:

$$\begin{cases} \gamma & \text{for } t \in [0, \tau_k) \\ \lambda_t & \text{for } t \geq \tau_k \end{cases}$$

  2. For a fixed $T > 0$, we construct $\mathbf{P}_\gamma^\star$ as

$$\mathbf{P}_\gamma^\star(\mathscr{A}) \triangleq \mathbf{P}_\gamma(\mathscr{A} \mid N_T \geq k) \text{ for all } \mathscr{A} \in \Sigma$$

  3. Under the Portmanteau theorem (for convergence of measures), we take the limiting measure

$$\mathbf{P}_\gamma^\star \Rightarrow \mathbf{P}^\star \text{ as } \gamma \downarrow 0$$

- Primary properties of $\mathbf{P}^\star$
  - The change of measure is only absolutely continuous ($\mathbf{P}^\star \ll \mathbf{P}$); not equivalent
  - It concentrates all probability mass on $\{N_T \geq k\} \Rightarrow \mathbf{P}^\star(N_T \geq k) = 1$
  - The sequence $\{\tau_1, \ldots, \tau_k\}$ forms the **uniform order statistics** on $[0, T]$

- Let $\{u_\ell\}_{\ell=1}^k$ be a collection of i.i.d. uniform order statistics on $[0, T]$
- For $\ell = 1, \dots, k$, we redefine the spacing $\theta_\ell = u_\ell - u_{\ell-1}$ (by setting $\tau_\ell = u_\ell$)
- Define $\eta_\ell = h^\ell_{\theta_\ell} \exp\left(-\int_0^{\theta_\ell} h^\ell_s ds\right)$ and

$$\varphi_k(T) = \frac{T^k}{k!} \prod_{\ell=1}^k \eta_\ell$$

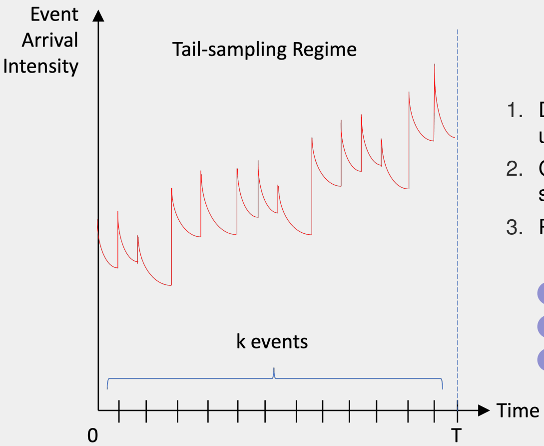
Theorem (Conditional Tail Sampling)

For any $T > 0$ and integer $k \geq 1$, we have

$$\varepsilon = \mathbf{P}(N_T \geq k) = \mathbf{E}^\star\big(\varphi_k(T)\big).$$

- For the sake of notational simplicity, we will use $\mathbf{E}$ and $\mathbf{E}^\star$ interchangeably

Simulation Algorithm (Sketch)



1. Draw an ordered sample of $\{u_1, \dots, u_k\}$ uniformly from $[0, T]$
2. Compute $\varphi_k(T)$ from a conditional sample path of $\{h_t^\ell\}_{\ell=1}^k$ on $t \in [0, \theta_\ell)$
3. Return the sample mean of $\varphi_k(T)$

- Exact Bridge Transform
- Conditional Point Sampling
- Doubly Stochastic Poisson Processes

$$\varphi_k(T) = \frac{T^k}{k!} \prod_{\ell=1}^{k} h_{\theta_\ell}^\ell \exp\left(-\int_0^{\theta_\ell} h_s^\ell ds\right)$$

The $T/k \to 0$ asymptotics ‣ Asymptotic optimality

Theorem (Bounded relative error)

Suppose that we have $\frac{T(x)}{k(x)} \to 0$ *as* $x \to \infty$. *The cTS estimator* $\varphi(x)$ *of* $\varepsilon(x)$ *achieves its asymptotic bounded relative error; i.e.,*

$$\limsup_{x\to\infty} \frac{\sqrt{\mathbf{Var}^\star\left(\varphi(x)\right)}}{\varepsilon(x)} < \infty\,,$$

if the following conditions hold:

1. *(Upper Bound):* $\limsup_{x\to\infty} \mathbf{E}^\star\left(\prod_{\ell=1}^{k(x)}\left(\frac{\eta_\ell}{\mathbf{E}^\star(\eta_\ell)}\right)^2\right) < \infty$
2. *(Lower Bound):* $\liminf_{x\to\infty} \mathbf{E}^\star\left(\prod_{\ell=1}^{k(x)}\frac{\eta_\ell}{\mathbf{E}^\star(\eta_\ell)}\right) > 0$

(Note): $\eta_\ell = h^\ell_{\theta_\ell}\exp\left(-\int_0^{\theta_\ell} h^\ell_s ds\right) \xrightarrow{p} h^\ell_0$ *as* $T/k \to 0$ *(under mild regularity conditions)*

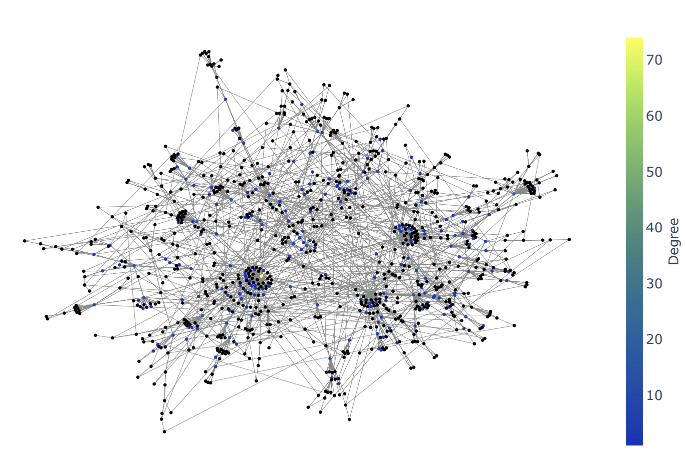
Example #1: Tail probability estimation

Epidemiologic Network

- This example deals with a risk analysis model of how diseases can be transmitted within a networked population
- Epidemiologic networks can take various forms, including social networks, contact networks, and more complex models that incorporate factors such as disease incubation periods and transmission probabilities
- The network can be used to study the dynamics of disease transmission, identify sources of infection, and develop strategies for disease control and prevention
- Several natural and human-made systems, including the World Wide Web, citation networks, and some social networks, contain few nodes (called hubs) with unusually high degree as compared to the other nodes

A **realistic** network structure of the Barabási-Albert model

- Our network analysis adopts the Barabási-Albert (BA) model
- It reflects the scale-free power-law of degree distribution by addressing the preferential attachment feature existing in real-world networks



A bottom-up formulation of self-exciting intensity specification

- For given $n \in \mathbb{N}$ agents, each infection indicator process $N^i \in \{0, 1\}$ admits its intensity with $\omega_i > 0$

$$\lambda^i = \left(\omega_i x^0 + x^i\right)\left(1 - N^i\right)$$

- λ^i denotes the intensity of the infection process of the i^{th} agent
  - $x^0 \Rightarrow$ the systematic risk factor as the common source of indirect transmission
  - $\{x^i\}_{i=1}^n \Rightarrow$ a set of idiosyncratic factors as the drivers of direct contagion

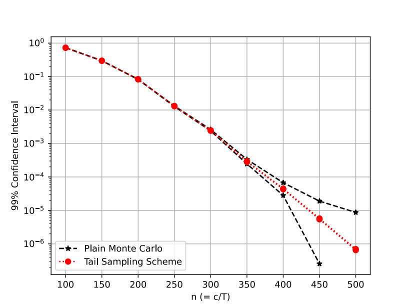
  ▸ Model specification

- The intensity of the infection counting process $N_t = \sum_{i=1}^n N_t^i$ is obtained by

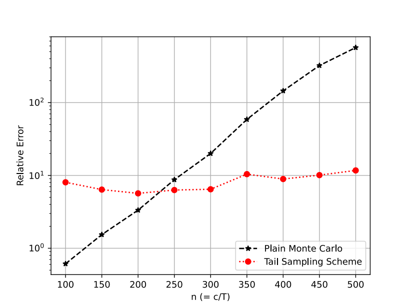
$$\lambda_t = \sum_{i=1}^{n} \lambda_t^i \quad \text{for} \quad t \geq 0$$

Example #1: Tail probability estimation (cont.)

- Let n be the rarity parameter for estimating $\mathbf{P}\left(N_{T_n} \geq k_n\right)$ with relative errors
  - We set $T_n = c/n$ and $k_n = \mu n$ for $c = 300$ and $\mu = 0.1$
  - We increase the size of the network and shrink the investigation horizon, while keeping the threshold at 10% of the population
  - We allow 60 seconds of CPU time for each estimation



(a) Estimated $\mathbf{P}\left(N_{T_n} \geq k_n\right)$



(b) Relative Errors (Log scale)

Part II:
Conditional Expectations on the Tail

- We focus on the extreme values of a distribution, where the probability is low but the consequences can be significant
  - The **conditional expectation on the tail events** can facilitate an understanding of how a random variable behaves in such circumstances

- The Bayes' formula allows one to investigate the expected behavior of any random variable X conditional on a tail scenario

$$\mathbf{E}\left(X \,|\, N_T \geq k\right) = \frac{\mathbf{E}(X \cdot 1_{\{N_T \geq k\}})}{\varepsilon}$$

  - No knowledge of the law of X conditional on the tail event is required

Marked Point Process

- Consider a sequence of random quantities (e.g., random losses) $\pi_\ell \geq 0$ as a **mark** associated with each arrival time τ_ℓ for $\ell \geq 1$
- The loss process L is defined as

$$L_t = \sum_{\ell=1}^{N_t} \pi_\ell \ ,$$

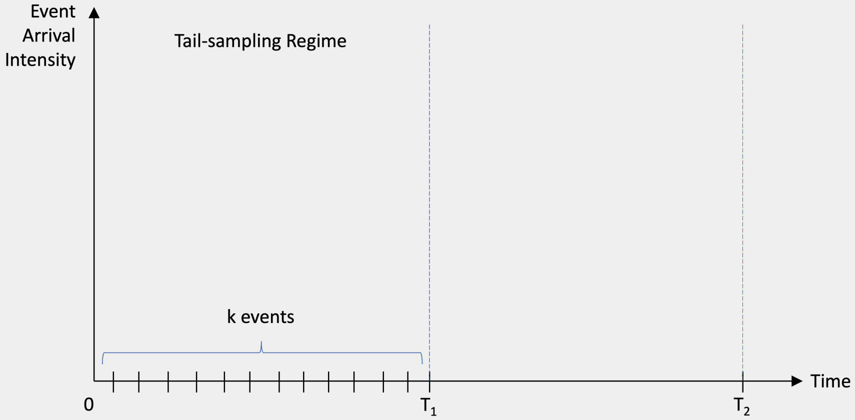
  where the jump times of N and L coincide, and the ℓ-th jump size of L is π_ℓ
- N can be described as a special example of L, where $\pi_\ell = 1$ for all $\ell \geq 1$

An extended cTS scheme

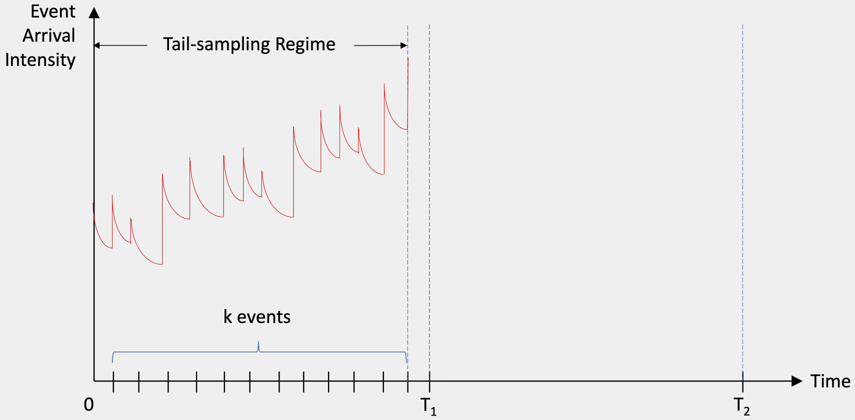
- Fix some (T_1, T_2) such that $0 < T_1 \leq T_2$ to estimate $\mathbf{E}\left(L_{T_2} \,\middle|\, N_{T_1} \geq k\right)$
- Now we extend the tail sampling scheme specific to the tail event $\{N_{T_1} \geq k\}$
- Let $\left\{u_\ell^{T_1}\right\}_{\ell=1}^{k}$ be a collection of i.i.d. uniform order statistics on $[0, T_1]$
- Also consider an i.i.d. sequence of standard exponential random variables $\{\mathcal{E}_\ell\}_{\ell \geq k+1}$
- We redefine the spacing θ_ℓ as

$$\theta_\ell = \begin{cases} u_\ell^{T_1} - u_{\ell-1}^{T_1} & \text{for } \ell = 1, \ldots, k \\ \inf\left\{t > 0 : \int_0^t h_s^\ell \, \mathrm{d}s \geq \mathcal{E}_\ell\right\} & \text{for } \ell \geq k+1 \end{cases}$$

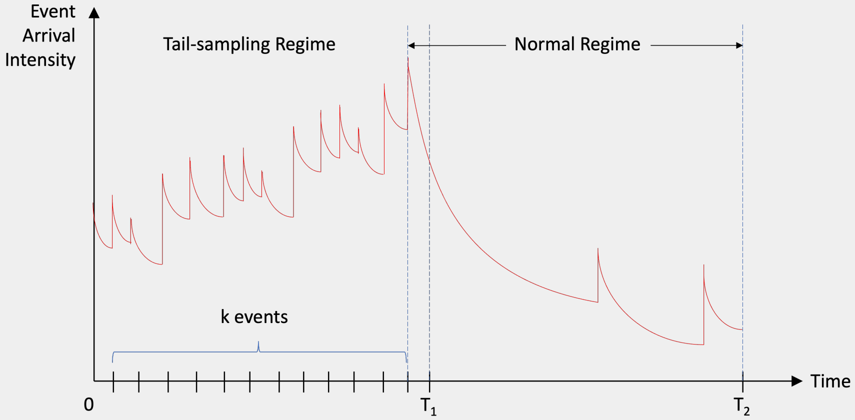
Conditional Tail Sampling Algorithm



Conditional Tail Sampling Algorithm



Conditional Tail Sampling Algorithm



Theorems for conditional expectations

Theorem (Extended conditional tail sampling)

For any integer $k \geq 1$, the following identities hold for $(0 <)\ T_1 \leq T_2$:

$$\mathbf{E}\left(L_{T_2} \cdot 1_{\{N_{T_1} \geq k\}}\right) = \mathbf{E}^\star\left(L_{T_2} \cdot \varphi_k\big(T_1\big)\right); \qquad \mathbf{E}\left(L_{T_2} \mid N_{T_1} \geq k\right) = \frac{\mathbf{E}^\star\left(L_{T_2} \cdot \varphi_k\big(T_1\big)\right)}{\mathbf{E}^\star\left(\varphi_k\big(T_1\big)\right)}.$$

Theorem (Relative error bound)

Let $\varphi \triangleq \varphi_k(T_1)$. Then, we have

$$\frac{\sqrt{\mathbf{Var}\left(L_{T_2} \middle| N_{T_1} \geq k\right)}}{\mathbf{E}\left(L_{T_2} \middle| N_{T_1} \geq k\right)} \leq \frac{\sqrt{\mathbf{E}\left(\varphi^2\right)}}{\mathbf{E}\left(\varphi\right)} \cdot \frac{\sqrt{\mathbf{E}\left(L_{T_2}^2 \cdot \varphi^2\right)}}{\mathbf{E}\left(L_{T_2} \cdot \varphi\right)},$$

if $\mathbf{Cov}\left(L_{T_2}\varphi, \varphi\right) \geq 0$ holds.

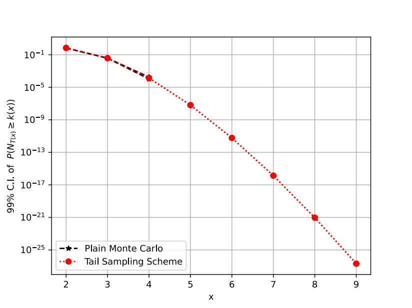
Default Clustering in a Stochastic Network

- This example examines the potential (expected) consequences that a financial system or market may face in extreme or tail-risk scenarios in a shorter horizon
- This analysis focuses on understanding how the credit risk at the system-level evolves when the financial system experiences severe stress or crisis events at the beginning
- We are interested in estimating both

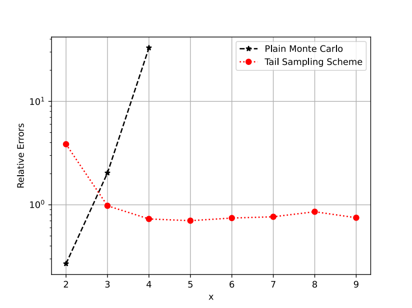
$$\mathbf{P}\left(N_{T(x)} \geq k(x)\right) \quad \text{and} \quad \mathbf{E}\left(N_T \,\middle|\, N_{T(x)} \geq k(x)\right),$$

where $T(x) \to 0$ and $k(x) \to \infty$ as $x \to \infty$ ▸ Model specification

- Estimated $\mathbf{P}\left(N_{T(x)} \geq k(x)\right)$ and their relative errors
  - We take $T(x) = T/x$ with $T = 5$ and $k(x) = 5x$

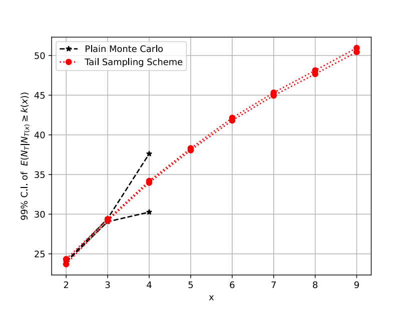


(a) $\mathbf{P}\left(N_{T(x)} \geq k(x)\right)$

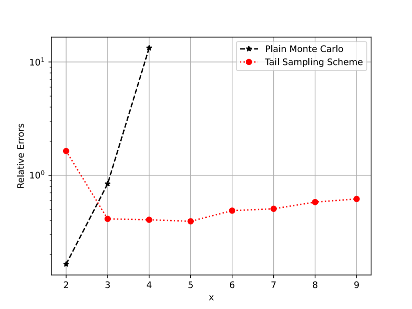


(b) Relative Errors (Log scale)

- Estimated $\mathbf{E}\left(N_T \,\middle|\, N_{T(x)} \geq k(x)\right)$ and their relative errors
  - We take $T(x) = T/x$ with $T = 5$ and $k(x) = 5x$



(a) $\mathbf{E}\left(N_T \,\middle|\, N_{T(x)} \geq k(x)\right)$



(b) Relative Errors (Log scale)

Part III:
Applications to Financial Examples

Example #3: Defaultable Security Pricing

(Short) Term structure of credit spreads for risky zero coupon bonds

- To estimate the term structure of credit spreads by determining fair compensation for bearing credit risk across various maturities
  - Our focus is on the short-term regime with a small value of $T \downarrow 0$

Practical relevance with “small” T

- When a depository institution establishes a **daily interest facility (DIF)**, the central bank adjusts the DIF rate to account for the **overnight credit spread** between unsecured and collateralized overnight lending
- The growing prevalence of **blockchain technology** has created a need for an ultra-short tenor interest rate curve that can be estimated **at an intraday level** to enable immediate settlement of transactions in the real-time interbank money market

Example #3: Defaultable Security Pricing (cont.)

An illustrative model specification (τ: Default arrival time)

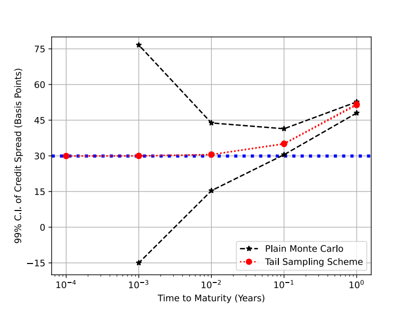
- Short-rate process: $dr_t = \kappa(y - r_t)dt + \sigma\sqrt{r_t}dW_t^r$
- A state process: $dx_t = a(b - x_t)dt + c\sqrt{x_t}dW_t^x, \quad W^r \perp W^x$
- Default intensity process: $\lambda_t = \left(\rho r_t + \sqrt{1-\rho^2}x_t\right) \cdot 1_{\{\tau > t\}}, \quad \rho \in (0,1)$
- Recovery process: $R_t = \frac{\lambda_0}{\lambda_0 + \lambda_{t-}} \in (0,1)$
- Loss process: $L_t = (1 - R_\tau) \cdot 1_{\{\tau \le t\}} = \underbrace{\frac{\lambda_{\tau-}}{\lambda_0 + \lambda_{\tau-}}}_{=\text{ mark}} \cdot 1_{\{\tau \le t\}}$
- (Our interest): Term structure of credit spreads as $T \downarrow 0$

  $\Rightarrow$ No closed-form solution is available for defaultable security pricing

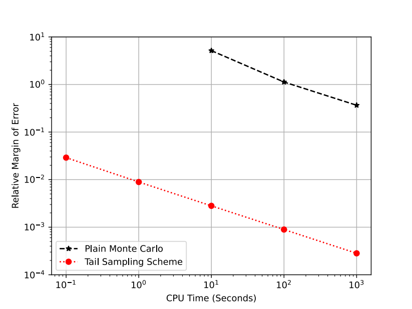
▸ Pricing details

Example #3: Defaultable Security Pricing (cont.)

- (Left panel): A fixed simulation time budget of 60 seconds
  - The true value of the short-horizon limit of credit spread is indicated by a horizontal thick dotted line ▸ Credit triangle
- (Right panel): Set $T = 1/252$ across different simulation time budgets
  - The relative margin of error is defined as the ratio of deviation around the point estimate of $s(T)$ at the 99% confidence interval



(a) Estimated $s(T)$



(b) Relative Margin of Error (Log scale)

Example #4: Maximum drawdown

An example of insurance risk analysis

- Let (N, L) denote the claim counting and the associated loss processes
- A reserve process is defined for some $\alpha > 0$ as

$$R_t = R_0 + \alpha t - L_t$$

- The drawdown process is expressed as

$$D_t = \sup_{s \in [0,t]} R_s - R_t$$

- For some fixed $T > 0$, the maximum drawdown is given by

$$D_T^\star = \sup_{t \in [0,T]} D_t$$

by measuring the largest reserve drop from its peak to trough in $[0, T]$

- (Our interest): The estimation of $\mathbf{E}\left(D_T^\star \middle| N_T \geq k\right)$
- The conditional expectation of maximum drawdown can be expressed as

$$\mathbf{E}\left(D_T^\star \middle| N_T \geq k\right) = \frac{\mathbf{E}\left(D_T^\star \cdot 1_{\{N_T \geq k\}}\right)}{\mathbf{E}\left(1_{\{N_T \geq k\}}\right)} \qquad \text{(Plain Monte Carlo)}$$

- It can be rewritten under the tail sampling scheme as

$$\mathbf{E}\left(D_T^\star \middle| N_T \geq k\right) = \frac{\mathbf{E}\left(D_T^\star \cdot \varphi_k(T)\right)}{\mathbf{E}\left(\varphi_k(T)\right)} \qquad \text{(Conditional Tail Sampling)}$$

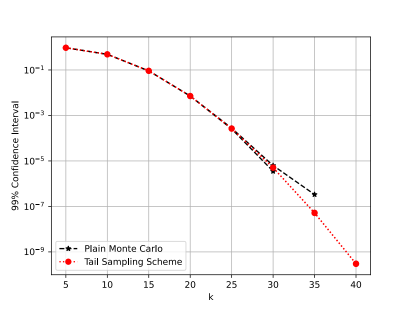
  where we have $\left\{\tau_\ell \overset{d}{=} u_\ell\right\}_{\ell=1}^{k}$ under the cTS scheme
- $D_T^\star$ can be expressed as a function of $\{\tau_\ell\}_{\ell=1}^{N_T}$ given the intensity trajectory

details

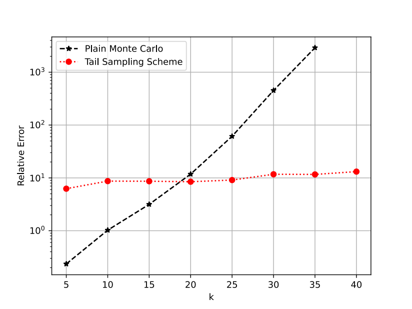
EXAMPLE #4: MAXIMUM DRAWDOWN (CONT.)

Estimated Tail Probabilities: $\mathbf{P}(N_T \geq k)$ for $5 \leq k \leq 40$

- We allow 60 seconds of CPU time for each estimation
- The cTS scheme shows an efficient variance reduction under the stochastic regime-changing intensity dynamics ▸ details



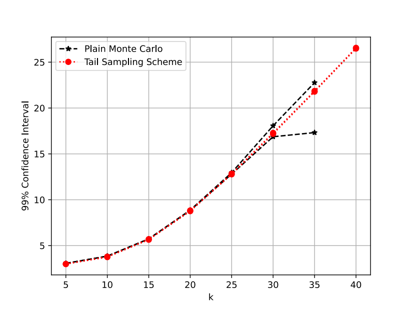
(a) Estimated $\mathbf{P}(N_T \geq k)$



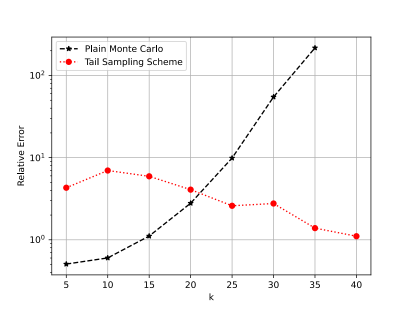
(b) Relative Error (Log scale)

Estimated Conditional Expectations: $\mathbf{E}\left(D_T^\star \mid N_T \geq k\right)$ for $5 \leq k \leq 40$

- We allow 60 seconds of CPU time for each estimation
- Our proposed cTS scheme is computationally more efficient than the benchmark pMC method as k increases



(a) Estimated $\mathbf{E}\left(D_T^\star \mid N_T \geq k\right)$



(b) Relative Error (Log scale)

Conclusion

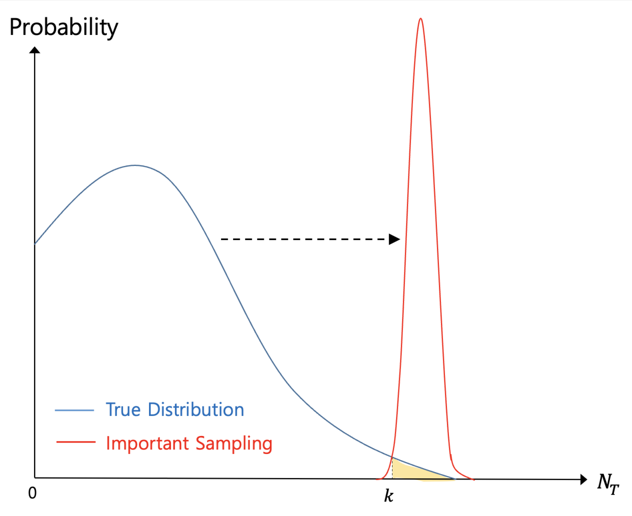
- This study develops a novel, easy to simulate and fast MC estimator of rare event probabilities via **conditional Tail Sampling (cTS)**
  - It accommodates any model specification provided it can be simulated
- Our algorithms provide meaningful efficiency gains by ensuring each simulated path hits the rare event with probability one
  - It ensures that none of the simulated paths will be wasted
- The limiting measure possesses attractive properties for simulation
  - Our approach facilitates a substantial reduction in the sampling error
- We test our algorithms on a wide spectrum of applications using empirically motivated reduced-form models
  - Our findings illustrate the superior performance of the proposed cTS scheme over plain MC
- **Our proposed methodology has potential for application in a wide range of real-world problems!**

Thank you!

Appendix

Appendix: Importance sampling

- The intuition behind **importance sampling** is to shift the sampling process from a difficult-to-sample distribution to a more manageable distribution



Exact bridge transform

- The unbiased estimator of $\varphi_k(T)$ can be exactly sampled efficiently when the sequence $\{h_0^\ell\}_{\ell \geq 1}$ satisfies the Markov property

$$
\begin{aligned}
\mathbf{P}\left(N_T \geq k\right) &= \mathbf{E}\left(\varphi_k(T)\right) \\
&= \frac{T^k}{k!}\mathbf{E}\left(\prod_{\ell=1}^{k} h_{\theta_\ell}^\ell \mathbf{E}\left(\exp\left(-\int_0^{\theta_\ell} h_s^\ell ds\right)\middle| \theta_\ell, h_0^\ell, h_{\theta_\ell}^\ell\right)\right)
\end{aligned}
$$

- This implies that an unbiased estimator of $\mathbf{P}\left(N_T \geq k\right)$ is available by sampling $\{u_1, \ldots, u_k\}$ when exact samples of $h_{\theta_\ell}^\ell$ can be simulated conditional on h_0^ℓ for $\ell = 1, \ldots, k$
- ... and, in many cases, the bridge transform

$$
\mathbf{E}\left(\exp\left(-\int_0^{\theta_\ell} h_s^\ell ds\right)\middle| \theta_\ell, h_0^\ell, h_{\theta_\ell}^\ell\right)
$$

can be evaluated without bias; e.g., see (Broadie & Kaya 2006) ▸ Return

Extension: Conditional Point Sampling

Corollary (Conditional Point Sampling)

For any integer $k \geq 1$, we have

$$\mathbf{P}(N_T = k) = \mathbf{E}^{\star}\left(\varphi_k(T) \cdot e^{-\int_0^{\theta_{k+1}} h_s^{k+1} ds}\right),$$

where $\theta_{k+1} = T - u_k$.

Return

Doubly stochastic Poisson processes

Time-change argument (Meyer 1971)

- General statement
  - N maybe be identified with a time-changed standard Poisson process
  - Given the filtration $\mathcal{G} = (\mathcal{F}_{A_t})_{t\geq 0}$, there exists a $\mathcal{G}$-adapted Poisson process C of unit rate such that

$$N_t = C_{A_t}$$

    where $A_t = \int_0^t \lambda_s ds$
- Doubly stochastic Poisson processes
  - No arrival time τ_ℓ may affect the dynamics of the intensity λ
  - $N_t = C_{A_t}$ holds in distribution for a standard Poisson process C independent of A

$$\mathbf{P}\left(C_{A_T} \geq k\right) = \boxed{\mathbf{E}\left(\varphi_k(A_T)\right)} \quad (\lambda = 1 \text{ for } C)$$

$$= \frac{1}{k!}\,\mathbf{E}\left(A_T^k\, e^{-\beta_k A_T}\right) \quad \left(\beta_k = \frac{u_k}{A_T} \sim \text{Beta}(k,1)\right)$$

$$= \frac{1}{(k-1)!}\,\mathbf{E}\Bigg(\underbrace{\frac{1}{k}\,A_T^k\,\mathbf{E}\left(\frac{1}{k}e^{-\beta_k A_T}\,\middle|\,A_T\right)}_{=\gamma(k,A_T)}\Bigg)$$

$$= \frac{1}{(k-1)!}\,\mathbf{E}\left((k-1)!\left(1 - e^{-A_T}\sum_{\ell=0}^{k-1}\frac{A_T^\ell}{\ell!}\right)\right)$$

$$= \boxed{1 - \sum_{\ell=0}^{k-1}\frac{\mathbf{E}\left(A_T^\ell e^{-A_T}\right)}{\ell!}}$$

$$= \mathbf{P}\left(N_T \geq k\right)$$

Return

Appendix: Asymptotic optimality

The $T/k \to 0$ asymptotics ▸ BRE

Assumption

There exists a function $f(x) > 0$ *with* $\frac{1}{f(x)} \log k(x) \to 0$ *as* $x \to \infty$ *such that*

$$\liminf_{x\to\infty} \frac{1}{f(x)} \log \varepsilon(x) \geq -1 ,$$

which is consistent with large deviations theory for rare events.

Assumption

The function $f(x)$ *defined above satisfies*

$$\limsup_{x\to\infty} \frac{1}{f(x)} \log \mathbf{P}\left(\mathscr{A}(x)\right) \leq -2 ,$$

where the event $\mathscr{A}(x)$ *is given by* $\mathscr{A}(x) = \{\varphi(x) \geq M(x)\}$ *with*

$$M(x) \triangleq e^{-f(x)} \prod_{\ell=1}^{k(x)} \frac{k(x)}{e \cdot \ell} \approx \frac{e^{-f(x)}}{\sqrt{2\pi k(x)}} < 1 \qquad \textit{(Stirling's approx.)}$$

The $T/k \to 0$ asymptotics (cont.) ▸ BRE

Theorem (Asymptotic optimality condition)

The cTS estimator $\varphi(x)$ is an asymptotically optimal estimator of $\varepsilon(x)$, if $\mathbf{P}\big(\mathscr{A}(x)\big) \downarrow 0$ as $x \to \infty$ holds.

Corollary (Approximate cTS with asymptotic optimality)

Define $\widetilde{\varphi}(x) \triangleq \min\{\varphi(x), M(x)\}$ and let $\widetilde{\varepsilon}(x) \triangleq \mathbf{E}\left(\widetilde{\varphi}(x)\right)$. Then, the following statements are true:

(i) *$\widetilde{\varphi}(x)$ is an asymptotically optimal estimator of $\widetilde{\varepsilon}(x)$.*

(ii) *We have $0 < \widetilde{\varepsilon}(x) \leq \varepsilon(x)$ for all x.*

(iii) *We have $\left|\widetilde{\varepsilon}(x) - \varepsilon(x)\right| \leq \big(1 - M(x)\big) \cdot \mathbf{P}\left(\mathscr{A}(x)\right)$ for all x.*

Example #1: Tail probability estimation (cont.)

- The systematic factor x^0 evolves with some $\kappa_0 > 0$ and $y_0 > 0$ by satisfying

$$dx_t^0 = \kappa_0 \left(y_0 - x_t^0 \right) dt + dJ_t$$

by driving the innovation of systematic factor dynamics

  - $J_t = \sum_{j=1}^{n} \delta_{0j} N_t^j$ captures the indirect feedback mechanism by driving the innovation of systematic factor dynamics
  - $\delta_{0j} \geq 0$ addresses the instant contribution of individual j's infection to the systematic risk factor

- The idiosyncratic factor process x^i follows

$$dx_t^i = \kappa_i \left(y_i - x_t^i \right) dt + \sum_{j=1}^{n} \delta_{ij} dN_t^j$$

  - The vector $\delta_i = (\delta_{i1}, \ldots, \delta_{in}) \geq 0$ represents i's sensitivity to events in the system for $i = 1, \ldots, n$

return

Example #1: Tail probability estimation (cont.)

- The construction of this model involves processes $h^{i\ell}$ which specify the conditional rate of arrival of the ℓth event at the ith component
- Letting $\mathcal{S}_\ell = \{i : N^i_{\tau_{\ell-1}} = 0\}$ denote the components that "survive" by time $\tau_{\ell-1}$,

$$h^\ell_t = \sum_{i \in \mathcal{S}_\ell} h^{i\ell}_t$$

  specifies the inter-arrival intensity of N which defines θ_ℓ under $\mathbf{P}$

- The distribution of the component that generates the ℓth event is

$$\begin{aligned} \mathbf{P}(\tau_\ell = \xi_i \,|\, \mathcal{F}_{\tau_\ell -}) &= \mathbf{P}^\star(\tau_\ell = \xi_i \,|\, \mathcal{F}_{\tau_\ell -}) \\ &= \frac{\lambda^i_{\tau_\ell -}}{\lambda_{\tau_\ell -}} = \frac{h^{i\ell}_{\theta_\ell}}{h^\ell_{\theta_\ell}} \qquad 1 \le i \le n \end{aligned}$$

return

Default Clustering in a Stochastic Network

- Suppose that there are $m = 100$ defaultable entities in the system
  - A policymaker should be concerned about failure of an abnormally large fraction of the total population in the system
- A bottom-up formulation
  - Consider a systematic risk factor $x^0 \geq 0$ and a set of idiosyncratic factor processes $\{x^i\}_{i=1}^m$ so that each default indicator process N^i admits

$$\lambda^i = \left(\omega_i x^0 + x^i\right)(1 - N^i)$$

    as its intensity
  - Here, $\omega_i > 0$ is the systematic factor loading of the i^{th} name in the system

▸ return

Example #2: Systemic Credit Loss (cont.)

- We assume that η^0 is the strong solution of the SDE given by

$$dx_t^0 = \kappa_0\Big(\theta_0 - x_t^0\Big)dt + \sigma_0\sqrt{x_t^0}dW_t^0$$

- We further assume that η^i is governed by the SDE under the statistical probability measure $\mathbf{P}$

$$dx_t^i = \kappa_i\left(\theta_i - x_t^i\right)dt + \sigma_i\sqrt{x_t^i}dW_t^i + \sum_{j=1}^{m}\delta_{ij}dN_t^j\,,$$

  - $(W^0, W^1, \ldots, W^m)$ is a vector of mutually independent Brownian motions
  - The Feller conditions are respected to ensure $x^0 > 0$ and $x^i > 0$ almost surely
  - $\left(\delta_{i1}, \ldots, \delta_{im}\right)$ represents name i's sensitivity to defaults in the system
  - The jump sensitivity are constructed by drawing each δ_{ij} from $[0, 1/m]$ uniformly

return

Example #3: Defaultable Security Pricing (cont.)

- The default-free bond price with unit face value: $V_0(T) = \mathbf{E}\left(e^{-\int_0^T r_s ds}\right)$
- The defaultable bond price with unit face value:

$$\begin{aligned} V_\lambda(T) &= \mathbf{E}\left(e^{-\int_0^T r_s ds}\, 1_{\{\tau > T\}} + R_\tau e^{-\int_0^\tau r_s ds}\, 1_{\{\tau \le T\}}\right) \\ &= \underbrace{\mathbf{E}\left(e^{-\int_0^T r_s ds}\right)}_{=V_0(T)} - \mathbf{E}\left(\underbrace{\left(e^{-\int_0^T r_s ds} - R_\tau e^{-\int_0^\tau r_s ds}\right)}_{:=X_T(\tau)}\, 1_{\{\tau \le T\}}\right) \\ &= V_0(T) - \mathbf{E}\Big(X_T(u_1)\, \varphi_1(T)\Big) \end{aligned}$$

- The credit spread is given by

$$\begin{aligned} s(T) &= -\frac{\log V_\lambda(T)}{T} + \frac{\log V_0(T)}{T} \\ &= -\frac{1}{T}\log\left(1 - \frac{\mathbf{E}\left(X_T(\tau)\cdot 1_{\{\tau \le T\}}\right)}{V_0(T)}\right) = -\frac{1}{T}\log\left(1 - \frac{\mathbf{E}^\star\left(X_T(u_1)\cdot \varphi_1(T)\right)}{V_0(T)}\right) \end{aligned}$$

▸ return

Example #3: Defaultable Security Pricing (cont.)

Theorem (Plain Monte Carlo)

The pMC estimator of $\mathbf{E}\left(X_T(\tau)\, 1_{\{\tau > T\}}\right)$ *has unbounded relative error as* $T \downarrow 0$.

Theorem (Conditional Tail Sampling)

The tail-sampling estimator of $\mathbf{E}^{\star}\left(X_T(u_1)\, \varphi_1(T)\right)$ *has bounded relative error as* $T \downarrow 0$.

- The short-horizon limit of the credit spread is given by

$$\lim_{T \downarrow 0} s(T) = (1 - R_0)\lambda_0 \ ,$$

which is also known as the *credit triangle* formula

▸ return

Example #4: Maximum drawdown (cont.)

- We adopt a Markov regime-switching model to model the dynamics of the stochastic claim intensity process λ
  - We presume that there are two claim regimes in that the state process $s_t \in \{0, 1\}$ for $t \geq 0$ follows the continuous-time Markov chain with $s_0 = 0$
  - The time-t intensity process takes the form of $\lambda_t \triangleq \lambda_{s_t} \in \{\lambda_0, \lambda_1\}$, where the time until the next regime-shift from state i to j is exponentially distributed with rate $v_{ij} > 0$ for $i \neq j$
  - For numerical analysis, we specify the baseline parameter set as

$$(\lambda_0, \lambda_1) = (1.5, 3.0),\ T = 5.0,\ R_0 = 25,\ \alpha = 3.0,\ (v_{01}, v_{10}) = (0.5, 1.0)$$

    and $\{\pi_1, \pi_2, \ldots\}$ are uniformly drawn from $[0.5, 1.5]$ independently
- Notice that the Markov regime-switching intensity process λ and the claim-counting process N satisfy a *doubly stochastic* property
- Due to its deterministic nature of the reserve process between two consecutive claim times with $\alpha > 0$, it is sufficient to check the running maximum of R_t and D_t for $t \in [0, T]$ just before each claim arrival time

▸ return