Importance Sampling for Heavy-tailed Processes

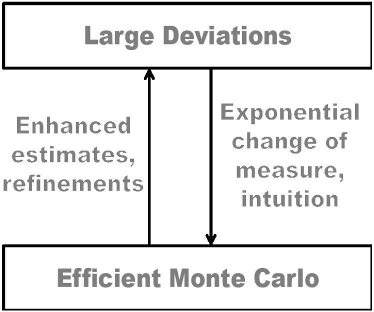
Jose Blanchet (based on work with Peter Glynn and Jingchen Liu)

Columbia University

June 2010

Efficient importance sampling (IS) and large deviations: Light tails

- S. R. S. Varadhan's Abel prize citation on large deviations theory: "It has greatly expanded our ability to use computers to simulate and analyze the occurrence of rare events."



Heavy tails: singular change of measure & "rogue" paths...

- Asmussen, Binswanger and Hojgaard '00

Heavy tails: singular change of measure & "rogue" paths...

- Asmussen, Binswanger and Hojgaard '00
- Direct IS strategy = *asymptotic* conditional distribution is singular **(likelihood ratio does not exist!)**

Heavy tails: singular change of measure & "rogue" paths...

- Asmussen, Binswanger and Hojgaard ‘00
- Direct IS strategy = *asymptotic* conditional distribution is singular **(likelihood ratio does not exist!)**
- **Contribution to the variance from some asymptotically negligible paths, "rogue paths", is typically substantial**

Illustrating the role of "rogue" paths

- X_1, X_2 i.i.d. Weibull, for $\beta \in (0,1)$ let

$$P(X_i > t) = \overline{F}(t) = \exp\left(-t^{\beta}\right), \ t > 0.$$

Illustrating the role of "rogue" paths

- X_1, X_2 i.i.d. Weibull, for $\beta \in (0,1)$ let
$$P(X_i > t) = \overline{F}(t) = \exp\left(-t^{\beta}\right), \ t > 0.$$
- **Estimate:** $P(X_1 + X_2 > b) \sim P(X_1 > b) + P(X_2 > b)$ as $b \nearrow \infty$.

Illustrating the role of "rogue" paths

- X_1, X_2 i.i.d. Weibull, for $\beta \in (0,1)$ let
$$P(X_i > t) = \overline{F}(t) = \exp\left(-t^{\beta}\right), \quad t > 0.$$
- **Estimate:** $P(X_1 + X_2 > b) \sim P(X_1 > b) + P(X_2 > b)$ as $b \nearrow \infty$.
- **"Natural" IS strategy:** Sample
$$(Y_1, Y_2) = \begin{cases} (X_1,\ X_2 | X_1 \ \&\ X_2 > b - X_1) & \text{with pr } 1/2 \\ (X_1 | X_2 \ \&\ X_1 > b - X_2,\ X_2) & \text{with pr } 1/2 \end{cases}$$

Illustrating the role of "rogue" paths

- X_1, X_2 i.i.d. Weibull, for $\beta \in (0,1)$ let

$$P\left(X_i > t\right) = \overline{F}\left(t\right) = \exp\left(-t^{\beta}\right),\ \ t > 0.$$

- **Estimate:** $P\left(X_1 + X_2 > b\right) \sim P\left(X_1 > b\right) + P\left(X_2 > b\right)$ as $b \nearrow \infty$.
- **"Natural" IS strategy:** Sample

$$\left(Y_1, Y_2\right) = \begin{cases} \left(X_1,\ X_2 | X_1\ \&\ X_2 > b - X_1\right) & \text{with pr } 1/2 \\ \left(X_1 | X_2\ \&\ X_1 > b - X_2,\ X_2\right) & \text{with pr } 1/2 \end{cases}$$

- **IS estimator**

$$\frac{f_{X_1}\left(y_1\right) f_{X_2}\left(y_2\right)}{f_{Y_1,Y_2}\left(y_1, y_2\right)} = \frac{2\overline{F}\left(b - y_1\right)\overline{F}\left(b - y_2\right) I\left(y_1 + y_2 > b\right)}{\overline{F}\left(b - y_1\right) + \overline{F}\left(b - y_2\right)}$$

Illustrating the role of "rogue" paths

- X_1, X_2 i.i.d. Weibull, for $\beta \in (0,1)$ let

$$P(X_i > t) = \overline{F}(t) = \exp\left(-t^{\beta}\right), \ t > 0.$$

- **Estimate:** $P(X_1 + X_2 > b) \sim P(X_1 > b) + P(X_2 > b)$ as $b \nearrow \infty$.
- **"Natural" IS strategy:** Sample

$$(Y_1, Y_2) = \begin{cases} (X_1, \ X_2 | X_1 \ \& \ X_2 > b - X_1) & \text{with pr } 1/2 \\ (X_1 | X_2 \ \& \ X_1 > b - X_2, \ X_2) & \text{with pr } 1/2 \end{cases}$$

- **IS estimator**

$$\frac{f_{X_1}(y_1) f_{X_2}(y_2)}{f_{Y_1,Y_2}(y_1, y_2)} = \frac{2\overline{F}(b - y_1)\overline{F}(b - y_2) I(y_1 + y_2 > b)}{\overline{F}(b - y_1) + \overline{F}(b - y_2)}$$

- **Second moment**

$$\int_{y_1 + y_2 > b} \left(\frac{f_{X_1} f_{X_2}}{f_{Y_1,Y_2}}\right)^2 f_{Y_1,Y_2} dy_1 dy_2 = \int_{y_1 + y_2 > b} \frac{f_{X_1} f_{X_2}}{f_{Y_1,Y_2}} f_{X_1} f_{X_2} dy_1 dy_2$$

Illustrating the role of "rogue" paths

- Second moment

$$\int_{y_1+y_2>b} \frac{f_{X_1} f_{X_2}}{f_{Y_1,Y_2}} f_{X_1} f_{X_2} \, dy_1 \, dy_2$$

Illustrating the role of "rogue" paths

- Second moment

$$\int_{y_1+y_2>b} \frac{f_{X_1} f_{X_2}}{f_{Y_1,Y_2}} f_{X_1} f_{X_2} dy_1 dy_2$$

- NOTE: $y_1 = b/2,\ y_2 = b/2$

$$\frac{1}{P\left(X_1 + X_2 > b\right)^2} \times \frac{f_{X_1}\left(b/2\right) f_{X_2}\left(b/2\right)}{f_{Y_1,Y_2}\left(b/2, b/2\right)} f_{X_1}\left(b/2\right) f_{X_2}\left(b/2\right)$$

$$= \frac{\overline{F}\left(b/2\right)^2 f_{X_1}\left(b/2\right)^2}{P\left(X_1 + X_2 > b\right)^2 \overline{F}\left(b/2\right)} \approx \frac{\exp\left(-3\left(b/2\right)^{\beta} + 2b^{\beta}\right)}{4}$$

Illustrating the role of "rogue" paths

- Second moment

$$\int_{y_1+y_2>b} \frac{f_{X_1} f_{X_2}}{f_{Y_1,Y_2}} f_{X_1} f_{X_2} dy_1 dy_2$$

- NOTE: $y_1 = b/2,\ y_2 = b/2$

$$\frac{1}{P\left(X_1 + X_2 > b\right)^2} \times \frac{f_{X_1}\left(b/2\right) f_{X_2}\left(b/2\right)}{f_{Y_1,Y_2}\left(b/2, b/2\right)} f_{X_1}\left(b/2\right) f_{X_2}\left(b/2\right)$$

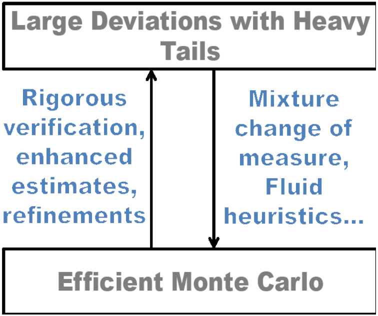
$$= \frac{\overline{F}\left(b/2\right)^2 f_{X_1}\left(b/2\right)^2}{P\left(X_1 + X_2 > b\right)^2 \overline{F}\left(b/2\right)} \approx \frac{\exp\left(-3\left(b/2\right)^{\beta} + 2b^{\beta}\right)}{4}$$

- **Conclusion:** Pick $\beta = 2/3$ then $3 < 2^{\beta+1}$

Squared Rel. Error: $\dfrac{Var\left(\mathrm{IS}\right)}{P\left(X_1 + X_2 > b\right)^2} \longrightarrow \infty$ as $b \nearrow \infty$.

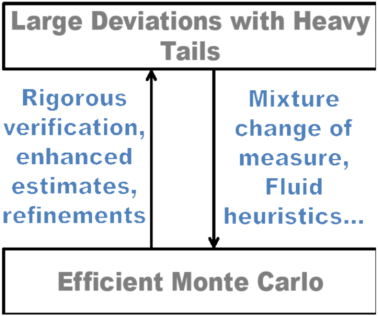
Contributions: Big Picture

- Efficient changes of measure for heavy tails



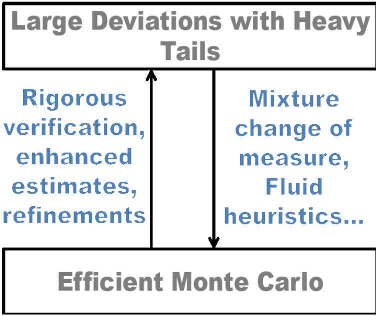
Contributions: Big Picture

- Efficient changes of measure for heavy tails
- Optimal complexity properties



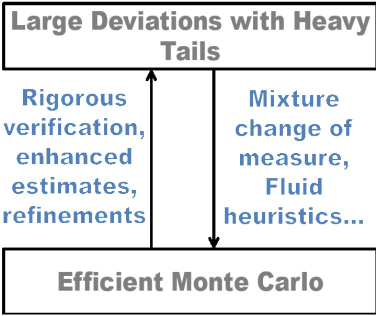
Contributions: Big Picture

- Efficient changes of measure for heavy tails
- Optimal complexity properties
- Lyapunov inequalities & construction



Contributions: Big Picture

- Efficient changes of measure for heavy tails
- Optimal complexity properties
- Lyapunov inequalities & construction
- Supporting conditional limit theorems & sampling



Ruin problems for random walks

- $X_1, X_2, \ldots$ i.i.d. & $EX_i = -\mu < 0$

Ruin problems for random walks

- $X_1, X_2, \ldots$ i.i.d. & $EX_i = -\mu < 0$
- $S_n = X_1 + \ldots + X_n, \quad \tau_b = \inf\{n \geq 0 : S_n > b\}$

Ruin problems for random walks

- $X_1, X_2, \ldots$ i.i.d. & $EX_i = -\mu < 0$
- $S_n = X_1 + \ldots + X_n, \quad \tau_b = \inf\{n \geq 0 : S_n > b\}$

- **Goal**: Design *efficient* (bounded rel. error) algorithm for

$$u(b) = P_0(\tau_b < \infty)$$

and

$$E[H(S_n : n \leq \tau_b), \tau_b < \infty]$$

Ruin problems for random walks

- $X_1, X_2, \ldots$ i.i.d. & $EX_i = -\mu < 0$
- $S_n = X_1 + \ldots + X_n, \quad \tau_b = \inf\{n \geq 0 : S_n > b\}$
- **Goal**: Design *efficient* (bounded rel. error) algorithm for

$$u(b) = P_0(\tau_b < \infty)$$

and

$$E[H(S_n : n \leq \tau_b), \tau_b < \infty]$$

- **Assumption:** X_i's suitably heavy tailed (Weibullian, power-law, lognormal type tails...)

Large deviations for heavy-tailed random walks

Theorem (Pakes-Veraberbeke)

Suppose X_i's and the integrated tail, $v(\cdot)$, are subexponential then

$$u(b) \sim v(b) := \frac{1}{\mu} \int_b^\infty P(X_i > t)\, dt$$

as $b \nearrow \infty$.

Simulation for heavy-tailed random walks

Theorem (B.-Glynn 07)

Suppose $X_i \in \mathcal{S}^*$ *let* W *satisfy*

$$P(W > t) = \min\{\frac{1}{\mu}\int_b^\infty P(X_i > t)\,dt, 1\}$$

and define $P^Q(\cdot)$,

$$P^Q(S_{n+1} \in dy | S_n = s) = \frac{f(y-s)\,v(b-y)}{w(b-s)}dy,$$

where $w(b-s) = E[v(b-s-X)]$. *Then,*

$$E_0^Q[\text{2nd moment IS estimator}] \leq cu(b)^2$$

and $E^Q\tau_b = O(b)$ *if* $E|X|^{2+\varepsilon} < \infty$.

Simulation for heavy-tailed random walks

- Replacing $v(b)$ by $u(b)$ in $P^{Q}(\ldots)$ gives zero variance

Simulation for heavy-tailed random walks

- Replacing $v(b)$ by $u(b)$ in $P^{Q}(\ldots)$ gives zero variance
- **Drawbacks?** Computing $w(b)$

Simulation for heavy-tailed random walks

- Replacing $v(b)$ by $u(b)$ in $P^{Q}(\ldots)$ gives zero variance
- **Drawbacks?** Computing $w(b)$
- What if EX^2? Termination time?

Large deviations for heavy-tailed random walks

Theorem (Asmussen-Kluppelberg)

(Simplified) If $P(X_i > t) \sim ct^{-\alpha}$ *for* $\alpha > 1$. *Then, conditional on* $\tau_b < \infty$, *we have that*

$$\left(\frac{S_{u\tau_b}}{\tau_b}, \frac{S_{\tau_b} - b}{b}, \frac{\tau_b}{b}\right) \Longrightarrow (-\mu u, Z_1, Z_2)$$

on $D(0,1) \times R \times R$ *as* $b \nearrow \infty$, *where* Z_1 *and* Z_2 *are Pareto with index* $\alpha - 1$.

- So, $E(\tau_b | \tau_b < \infty) = \infty$ if $\alpha \in (1,2)$ **&** $E(\tau_b | \tau_b < \infty) = O(b)$ **if** $\alpha > 2$.

Large deviations for heavy-tailed random walks

Theorem (Asmussen-Kluppelberg)

(Simplified) If $P(X_i > t) \sim ct^{-\alpha}$ *for* $\alpha > 1$. *Then, conditional on* $\tau_b < \infty$, *we have that*

$$\left(\frac{S_{u\tau_b}}{\tau_b}, \frac{S_{\tau_b} - b}{b}, \frac{\tau_b}{b}\right) \Longrightarrow (-\mu u, Z_1, Z_2)$$

on $D(0,1) \times R \times R$ *as* $b \nearrow \infty$, *where* Z_1 *and* Z_2 *are Pareto with index* $\alpha - 1$.

- So, $E(\tau_b | \tau_b < \infty) = \infty$ if $\alpha \in (1,2)$ **&** $E(\tau_b | \tau_b < \infty) = O(b)$ **if** $\alpha > 2$.
- *Are we doomed?*

Results

- Bounded relative error with mixture changes of measure (avoids computing $w(\cdot)$ in BG)

Results

- Bounded relative error with mixture changes of measure (avoids computing $w(\cdot)$ in BG)
- Can inforce $E^{Q}\tau_b = O(b)$ even if $E|X|^2 = \infty$

Results

- Bounded relative error with mixture changes of measure (avoids computing $w(\cdot)$ in BG)
- Can inforce $E^{Q}\tau_{b} = O(b)$ even if $E|X|^{2} = \infty$
- Verification technique via Lyapunov functions

Results

- Bounded relative error with mixture changes of measure (avoids computing $w(\cdot)$ in BG)
- Can inforce $E^{Q}\tau_{b} = O(b)$ even if $E|X|^{2} = \infty$
- Verification technique via Lyapunov functions
- Conditional central limit theorems (refine Asmussen & Kluppelberg)

Rest of the talk

- Walk you through the ideas behind the contributions via a *three step procedure*

Rest of the talk

- Walk you through the ideas behind the contributions via a *three step procedure*
- Step 1: Parametric family of changes of measure

Rest of the talk

- Walk you through the ideas behind the contributions via a *three step procedure*
- Step 1: Parametric family of changes of measure
- Step 2: Lyapunov function selection (fluid heuristics)

Rest of the talk

- Walk you through the ideas behind the contributions via a *three step procedure*
- Step 1: Parametric family of changes of measure
- Step 2: Lyapunov function selection (fluid heuristics)
- Step 3: Verfication (parameter selection)

Step 1: Parametric family of changes of measure

- $P(X_i > t) \sim ct^{-\alpha}$ as $t \nearrow \infty$ for $\alpha > 1$

Step 1: Parametric family of changes of measure

- $P(X_i > t) \sim ct^{-\alpha}$ as $t \nearrow \infty$ for $\alpha > 1$
- **Family of changes-of-measure:** s is current position, $(p(s)$ and $a \in (0,1))$

$$
\begin{aligned}
f_{X|s}(x|s) &= p(s)\frac{f_X(x)\,I(x > a(b-s))}{P(X > a(b-s))} \\
&\quad + (1-p(s))\frac{f_X(x)\,I(x \le a(b-s))}{P(X \le a(b-s))} \\
&\approx p(s)\frac{f_X(x)\,I(x > a(b-s))}{P(X > a(b-s))} + (1-p(s))\,f_X(x)
\end{aligned}
$$

Step 1: Parametric family of changes of measure

- $P\left(X_i > t\right) \sim ct^{-\alpha}$ as $t \nearrow \infty$ for $\alpha > 1$
- **Family of changes-of-measure:** s is current position, $\left(p\left(s\right) \text{ and } a \in \left(0,1\right)\right)$

$$
\begin{aligned}
f_{X|s}\left(x|\, s\right) &= p\left(s\right)\frac{f_X\left(x\right) I\left(x > a\left(b-s\right)\right)}{P\left(X > a\left(b-s\right)\right)} \\
&\quad + \left(1-p\left(s\right)\right)\frac{f_X\left(x\right) I\left(x \le a\left(b-s\right)\right)}{P\left(X \le a\left(b-s\right)\right)} \\
&\approx p\left(s\right)\frac{f_X\left(x\right) I\left(x > a\left(b-s\right)\right)}{P\left(X > a\left(b-s\right)\right)} + \left(1-p\left(s\right)\right) f_X\left(x\right)
\end{aligned}
$$

- Dupuis, Leder & Wang '07.

Step 1: Parametric family (discussion)

- **Role of** $a \in (0, 1)$: Capture rogue sample paths, $(X|X > x) \approx x + xZ$, where $Z \geq 0$ is Pareto

Step 1: Parametric family (discussion)

- **Role of** $a \in (0,1)$: Capture rogue sample paths, $(X|X > x) \approx x + xZ$, where $Z \geq 0$ is Pareto
- How to choose $p(s)$? Use large deviations results

Step 1: Parametric family (discussion)

- **Role of** $a \in (0,1)$: Capture rogue sample paths, $(X|X > x) \approx x + xZ$, where $Z \geq 0$ is Pareto
- How to choose $p(s)$? Use large deviations results
- Given no jump by time t, $S_t \approx -\mu t$ & jumping to b at $t+1$ *given* $\tau_b < \infty$

$$p(s) \approx \frac{P(X - \mu t > b)}{\int_0^\infty P(X - \mu t > b)\, dt} = \frac{\mu P(X > b + \mu t)}{\int_b^\infty P(X > u)\, du} = O\left(\frac{1}{b}\right).$$

Step 2: Lyapunov inequalities (variance control)

Lemma (B. & Glynn '07)

Suppose that there is a positive function $g(\cdot)$ such that

$$E\left(\frac{g(s+X)}{g(s)} \times \frac{f(X)}{f(X|s)}\right) \leq 1$$

and $g(s) \geq 1$ for $s > b$. Then,

$$E_s^Q(\text{2nd moment IS}) \leq g(s).$$

Step 2: Guessing Lyapunov function

- Want *bounded relative error*, so pick (for $\kappa > 0$)

$$g(s) = \min\left(\kappa\left(\int_{b-s}^{\infty} P(X > u)\, du\right)^2, 1\right).$$

Step 2: Guessing Lyapunov function

- Want *bounded relative error*, so pick (for $\kappa > 0$)

$$g(s) = \min\left(\kappa\left(\int_{b-s}^{\infty} P(X > u)\, du\right)^2, 1\right).$$

- Earlier discussion suggests (θ to be selected)

$$p(s) = \theta \frac{P(X > b - s)}{\int_{b-s}^{\infty} P(X > s)\, du}$$

- **On $g(s) < 1$ suffices to check**

$$E\left(\frac{g(s+X)}{g(s)} \times \frac{f(X)}{f(X|s)}\right)$$

$$\leq \quad \frac{P(X > a(b-s))^2}{p(s)\,g(s)} + \frac{E(g(s+X)\,; X \leq a(b-s))}{(1-p(s))\,g(s)} \leq 1$$

- Recall

$$p(s) = \theta \frac{P(X > b - s)}{\int_{b-s}^{\infty} P(X > s)\,du} \approx \frac{\theta(\alpha - 1)}{b - s},$$

$$g(s) = \kappa \left(\int_{b-s}^{\infty} P(X > u)\,du \right)^2,$$

- Recall

$$p(s) = \theta \frac{P(X > b-s)}{\int_{b-s}^{\infty} P(X > s)\, du} \approx \frac{\theta(\alpha - 1)}{b-s},$$

$$g(s) = \kappa \left(\int_{b-s}^{\infty} P(X > u)\, du \right)^2,$$

- Thus, as $b - s \nearrow \infty$

$$\frac{P(X > a(b-s))^2}{p(s)\, g(s)} \approx \frac{a^{-\alpha} P(X > a(b-s))}{\theta\kappa \int_{b-s}^{\infty} P(X > u)\, du} \approx \frac{a^{-\alpha}(\alpha - 1)}{\theta\kappa(b-s)}$$

Step 3: Tunning the parameters...

- Taylor approximation...ξ between 0 and X

$$\frac{E\left(g\left(s+X\right);X\leq a\left(b-s\right)\right)}{g\left(s\right)\left(1-p\left(s\right)\right)}$$

$$\approx\quad \left(1+p\left(s\right)\right)\left(1+\frac{\partial g\left(s\right)}{g\left(s\right)}E\left(\frac{\partial g\left(s+\xi\right)}{\partial g\left(s\right)}X;X\leq a\left(b-s\right)\right)\right)$$

Step 3: Tunning the parameters...

- Taylor approximation...ξ between 0 and X

$$\frac{E\left(g\left(s+X\right);X\leq a\left(b-s\right)\right)}{g\left(s\right)\left(1-p\left(s\right)\right)}$$

$$\approx \quad \left(1+p\left(s\right)\right)\left(1+\frac{\partial g\left(s\right)}{g\left(s\right)}E\left(\frac{\partial g\left(s+\xi\right)}{\partial g\left(s\right)}X;X\leq a\left(b-s\right)\right)\right)$$

- Note on $X\leq a\left(b-s\right)$

$$\left|\frac{\partial g\left(s+\xi\right)}{\partial g\left(s\right)}\right|\leq const\left|\frac{\overline{F}\left(b-s-a\left(b-s\right)\right)}{\overline{F}\left(b-s\right)}\right|\leq O\left(1\right)$$

uniformly over $b-s>0$... apply dominated convergence & obtain...

Step 3: Tunning the parameters...

- Given $\varepsilon > 0$, if $b - s$ is large (depending on ε)

$$
\begin{aligned}
& E\left(\frac{g(s+X)}{g(s)} \times \frac{f(X)}{f(X|s)}\right) \\
\leq \quad & \frac{a^{-\alpha}(\alpha-1)}{\theta\kappa(b-s)} + \underbrace{\left(1+\theta\frac{\alpha-1}{b-s}\right)}_{(1+p(s))} \underbrace{\left(1-2\mu\frac{(\alpha-1)(1-\varepsilon)}{b-s}\right)}_{(1+\partial g(s)/g(s))} \\
\approx \quad & \frac{a^{-\alpha}(\alpha-1)}{\theta\kappa(b-s)} + 1 + \theta\frac{(\alpha-1)}{b-s} - 2\mu\frac{(\alpha-1)(1-\varepsilon)}{b-s} \leq 1
\end{aligned}
$$

Step 3: Tunning the parameters...

- Given $\varepsilon > 0$, if $b - s$ is large (depending on ε)

$$E\left(\frac{g(s+X)}{g(s)} \times \frac{f(X)}{f(X|s)}\right)$$

$$\leq \frac{a^{-\alpha}(\alpha-1)}{\theta\kappa(b-s)} + \underbrace{\left(1+\theta\frac{\alpha-1}{b-s}\right)}_{(1+p(s))} \underbrace{\left(1-2\mu\frac{(\alpha-1)(1-\varepsilon)}{b-s}\right)}_{(1+\partial g(s)/g(s))}$$

$$\approx \frac{a^{-\alpha}(\alpha-1)}{\theta\kappa(b-s)} + 1 + \theta\frac{(\alpha-1)}{b-s} - 2\mu\frac{(\alpha-1)(1-\varepsilon)}{b-s} \leq 1$$

- If $a \lesssim 1$, $\varepsilon \gtrsim 0$ then $\theta \approx \mu$, $\kappa \approx 1/\mu^2$. And

$$g(s) \approx \frac{1}{\mu^2}\left(\int_{b-s} P(X > t)\, dt\right)^2 \sim u(b-s)^2$$

Summary (out of these calculations)

- 2 MIXTURES: Reg. varying – Need two mixtures & introduce $a \in (0,1)$ for rogue paths

Summary (out of these calculations)

- 2 MIXTURES: Reg. varying – Need two mixtures & introduce $a \in (0,1)$ for rogue paths
- ROLE OF $a \in (0,1)$ DCT: Analysis shows $a \in (0,1)$ for the **D**om. **C**onv. **T**hm.

$$\left|\frac{\partial g\left(s+\xi\right)}{\partial g\left(s\right)}\right| \leq const \left|\frac{\overline{F}\left(b-s-a\left(b-s\right)\right)}{\overline{F}\left(b-s\right)}\right| \leq O\left(1\right)$$

Summary (out of these calculations)

- 2 MIXTURES: Reg. varying – Need two mixtures & introduce $a \in (0,1)$ for rogue paths
- ROLE OF $a \in (0,1)$ DCT: Analysis shows $a \in (0,1)$ for the **D**om. **C**onv. **T**hm.

$$\left|\frac{\partial g\left(s+\xi\right)}{\partial g\left(s\right)}\right| \leq const \left|\frac{\overline{F}\left(b-s-a\left(b-s\right)\right)}{\overline{F}\left(b-s\right)}\right| \leq O\left(1\right)$$

- ZERO VARIANCE SELECTION: Can pick $\theta \approx \mu$, $\kappa \approx 1/\mu^2$ which yields for b large enough

(Jensen) $u\left(b\right)^2 \leq E^Q\left[\text{2nd moment}\right] \leq \left(1+\varepsilon\right) u\left(b\right)^2$ (Lyapunov)

Summary (out of these calculations)

- 2 MIXTURES: Reg. varying – Need two mixtures & introduce $a \in (0,1)$ for rogue paths
- ROLE OF $a \in (0,1)$ DCT: Analysis shows $a \in (0,1)$ for the **D**om. **C**onv. **T**hm.

$$\left|\frac{\partial g\,(s+\xi)}{\partial g\,(s)}\right| \leq const\,\left|\frac{\overline{F}\,(b-s-a\,(b-s))}{\overline{F}\,(b-s)}\right| \leq O\,(1)$$

- ZERO VARIANCE SELECTION: Can pick $\theta \approx \mu$, $\kappa \approx 1/\mu^2$ which yields for b large enough

$$\text{(Jensen) } u\,(b)^2 \leq E^Q\,[\text{2nd moment}] \leq (1+\varepsilon)\,u\,(b)^2 \text{ (Lyapunov)}$$

- PARAMETER CONSTRAINT: Selection came from

$$\frac{1}{\theta\kappa} + \theta - 2\mu \leq 0$$

Total variation approximation to conditional distribution

- ZERO VARIANCE SELECTION: Can pick $\theta \approx \mu$, $\kappa \approx 1/\mu^2$ which yields for b large enough

(Jensen) $u(b)^2 \leq E^Q[\text{2nd moment}] \leq (1+\varepsilon)u(b)^2$ (Lyapunov)

Total variation approximation to conditional distribution

- ZERO VARIANCE SELECTION: Can pick $\theta \approx \mu$, $\kappa \approx 1/\mu^2$ which yields for b large enough

  (Jensen) $u(b)^2 \leq E^Q[\text{2nd moment}] \leq (1+\varepsilon)u(b)^2$ (Lyapunov)

Lemma

If

$$\widetilde{E}\left(\left(\frac{dP}{d\widetilde{P}} \times I(\tau_b < \infty)\right)^2 \frac{1}{P(\tau_b < \infty)}\right) \leq 1 + \varepsilon$$

Then,

$$\sup_A \left| P\left((S_1, ..., S_{\tau_b}) \in A | \tau_b < \infty\right) - \widetilde{P}\left((S_1, ..., S_{\tau_b}) \in A\right)\right| < \varepsilon^{1/2}.$$

Total variation approximation to conditional distribution

- ZERO VARIANCE SELECTION: Can pick $\theta \approx \mu$, $\kappa \approx 1/\mu^2$ which yields for b large enough

  (Jensen) $u(b)^2 \leq E^Q[\text{2nd moment}] \leq (1+\varepsilon)u(b)^2$ (Lyapunov)

Lemma

If

$$\widetilde{E}\left(\left(\frac{dP}{d\widetilde{P}} \times I(\tau_b < \infty)\right)^2 \frac{1}{P(\tau_b < \infty)}\right) \leq 1 + \varepsilon$$

Then,

$$\sup_A \left| P((S_1, ..., S_{\tau_b}) \in A | \tau_b < \infty) - \widetilde{P}((S_1, ..., S_{\tau_b}) \in A) \right| < \varepsilon^{1/2}.$$

- **Conclusion:** $\widetilde{P}(\cdot)$ approximates conditional distribution given $\tau_b < \infty$

Conditional functional central limit theorems

- Easy to see that $\tau_b \approx$ time to get one large jump under $\widetilde{P}(\cdot)$ &...
- Using

$$f_{X|s}(x|\,s) = p(s)\frac{f_X(x)\,I(x > a(b-s))}{P(X > a(b-s))} + (1-p(s))\,f_X(x)$$

recover & refine Asmussen-Kluppelberg...

Theorem (B. & Liu)

[Simplified] If $P(X_i > t) \sim ct^{-\alpha}$ for $\alpha > 2$ and $Var(X_i) = \sigma^2 < \infty$,

$$\left(\frac{S_{u\tau_b} + \mu\tau_b}{\sqrt{\tau_b}}, \frac{S_{\tau_b} - b}{b}, \frac{\tau_b}{b}\right) \Longrightarrow (\sigma B(u), Z_1, Z_2)$$

on $D(0,1) \times R \times R$ as $b \nearrow \infty$, where Z_1 and Z_2 are Pareto with index $\alpha - 1$.

Parameter selection for termination time

- PARAMETER CONSTRAINT: Selection came from

$$\frac{1}{\theta\kappa} + \theta - 2\mu \leq 0$$

Parameter selection for termination time

- PARAMETER CONSTRAINT: Selection came from

$$\frac{1}{\theta\kappa} + \theta - 2\mu \leq 0$$

- Select $\theta \lesssim 2\mu$ and κ large enough... still get bounded relative error BUT faster termination time (inducing more jumps!).

Parameter selection for termination time

- PARAMETER CONSTRAINT: Selection came from

$$\frac{1}{\theta\kappa} + \theta - 2\mu \leq 0$$

- Select $\theta \lesssim 2\mu$ and κ large enough... still get bounded relative error BUT faster termination time (inducing more jumps!).
- How fast can it be?

Termination time

- **Intuition:** $\tau_b \approx$ time to get one large jump

$$\widetilde{P}_0(\tau_b > tb) \approx \prod_{j=0}^{tb}(1 - p(-j\mu)) \approx \exp\left(-\sum_{j=0}^{tb}\frac{\theta(\alpha-1)}{b+\mu j}\right)$$

$$\approx \exp\left(-\int_0^{tb}\frac{\theta(\alpha-1)}{b+\mu s}ds\right) = \left(\frac{1}{1+\mu t}\right)^{\theta(\alpha-1)/\mu}.$$

Termination time

- **Intuition:** $\tau_b \approx$ time to get one large jump

$$\widetilde{P}_0(\tau_b > tb) \approx \prod_{j=0}^{tb}(1 - p(-j\mu)) \approx \exp\left(-\sum_{j=0}^{tb}\frac{\theta(\alpha-1)}{b+\mu j}\right)$$

$$\approx \exp\left(-\int_0^{tb}\frac{\theta(\alpha-1)}{b+\mu s}ds\right) = \left(\frac{1}{1+\mu t}\right)^{\theta(\alpha-1)/\mu}.$$

- So, under $\widetilde{P}(\cdot)$, τ_b/b is pareto with index $\theta(\alpha-1)/\mu$

Termination time

- **Intuition:** $\tau_b \approx$ time to get one large jump

$$\widetilde{P}_0(\tau_b > tb) \approx \prod_{j=0}^{tb}(1 - p(-j\mu)) \approx \exp\left(-\sum_{j=0}^{tb}\frac{\theta(\alpha-1)}{b+\mu j}\right)$$
$$\approx \exp\left(-\int_0^{tb}\frac{\theta(\alpha-1)}{b+\mu s}ds\right) = \left(\frac{1}{1+\mu t}\right)^{\theta(\alpha-1)/\mu}.$$

- So, under $\widetilde{P}(\cdot)$, τ_b/b is pareto with index $\theta(\alpha-1)/\mu$
- $\theta \lesssim 2\mu$, **then** $\widetilde{E}(\tau_b) = O(b)$ **if** $2(\alpha-1) > 1$ **OR** $\alpha > 3/2$

Summary of results on bounded relative error and termination time

Theorem (B. & Liu)

if X_i's are regularly varying (power law tails) can select mixture parameters using Lyapunov inequalities to guarantee: A) Total variation approximation (asymptotically zero relative error), B) Bounded relative error and $O(b)$ termination time if $\alpha > 3/2$.

More on efficiency and termination time

- What about $\alpha \in (1, 3/2)$?

More on efficiency and termination time

- What about $\alpha \in (1, 3/2)$?
- Basically impossible to do bounded relative error IS such that $\widetilde{E}\left(\tau_b\right) = O\left(b\right)$

More on efficiency and termination time

- What about $\alpha \in (1, 3/2)$?
- Basically impossible to do bounded relative error IS such that $\widetilde{E}(\tau_b) = O(b)$
- ANY reasonable IS should make

$$\widetilde{P}(\tau_b / b > t) \approx c t^{-\gamma+1}$$

More on efficiency and termination time

- What about $\alpha \in (1, 3/2)$?
- Basically impossible to do bounded relative error IS such that $\widetilde{E}(\tau_b) = O(b)$
- ANY reasonable IS should make

$$\widetilde{P}(\tau_b/b > t) \approx ct^{-\gamma+1}$$

- Bounded relative error says

$$\begin{aligned}\infty &> \int_0^\infty \frac{1}{P(\tau_b < \infty)^2} \left(\frac{P(\tau_b/b \in dt)}{\widetilde{P}(\tau_b/b \in dt)} \right)^2 \widetilde{P}(\tau_b/b \in dt) \\ &= \int_0^\infty \left(\frac{P(\tau_b/b \in dt | \tau_b < \infty)}{\widetilde{P}(\tau_b/b \in dt)} \right)^2 \widetilde{P}(\tau_b/b \in dt).\end{aligned}$$

More on efficiency and termination time

- Asmussen & Kluppelberg's result suggests

$$P\left(\tau_b/b \in dt | \tau_b < \infty\right) \approx (\alpha - 1)\mu(1 + \mu t)^{-\alpha} dt$$

More on efficiency and termination time

- Asmussen & Kluppelberg's result suggests

$$P\left(\tau_b/b \in dt \middle| \tau_b < \infty\right) \approx (\alpha - 1)\mu(1 + \mu t)^{-\alpha} dt$$

- So, must have

$$\int_1^\infty \left(\frac{t^{-\alpha}}{t^{-\gamma}}\right)^2 t^{-\gamma} dt = \int_1^\infty t^{-2\alpha+\gamma} dt < \infty$$

or $2\alpha > \gamma + 1$.

More on efficiency and termination time

- Asmussen & Kluppelberg's result suggests

$$P\left(\tau_b/b \in dt \middle| \tau_b < \infty\right) \approx (\alpha - 1)\mu(1 + \mu t)^{-\alpha} dt$$

- So, must have

$$\int_1^\infty \left(\frac{t^{-\alpha}}{t^{-\gamma}}\right)^2 t^{-\gamma} dt = \int_1^\infty t^{-2\alpha+\gamma} dt < \infty$$

  or $2\alpha > \gamma + 1$.
- At the same time, $\widetilde{E}\tau_b = O\left(b\right)$ demands

$$\int_1^\infty ct^{-\gamma+1} dt < \infty$$

  or $\gamma > 2$.

More on efficiency and termination time

- Asmussen & Kluppelberg's result suggests

$$P\left(\tau_b/b \in dt \middle| \tau_b < \infty\right) \approx (\alpha - 1)\mu(1 + \mu t)^{-\alpha} dt$$

- So, must have

$$\int_1^\infty \left(\frac{t^{-\alpha}}{t^{-\gamma}}\right)^2 t^{-\gamma} dt = \int_1^\infty t^{-2\alpha+\gamma} dt < \infty$$

  or $2\alpha > \gamma + 1$.

- At the same time, $\widetilde{E}\tau_b = O\left(b\right)$ demands

$$\int_1^\infty ct^{-\gamma+1} dt < \infty$$

  or $\gamma > 2$.

- So $\alpha > 3/2$...

Summary of results of termination time beyond critical case

Theorem (B. & Liu)

if X_i's are regularly varying (power law tails) $\alpha \in (1, 3/2)$ can select mixture parameters using Lyapunov inequalities to guarantee for $0 < \rho < (\alpha - 1)/(2 - \alpha)$

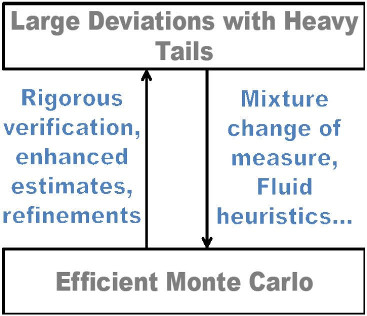
$$\widetilde{E}[(1+\rho)\text{-moment of IS Est}] \leq cu\,(b)^{1+\rho}$$

and $O\,(b)$ termination time.

- Bound $\rho < (\alpha - 1)/(2 - \alpha)$ optimal, similar argument as before...

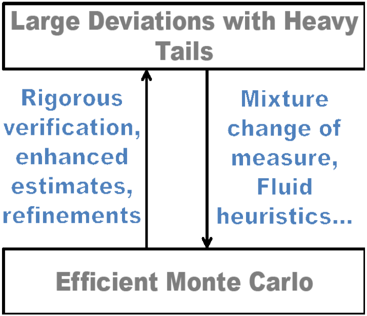
Keep in mind the Big Picture

- Efficient changes of measure for *heavy tails*



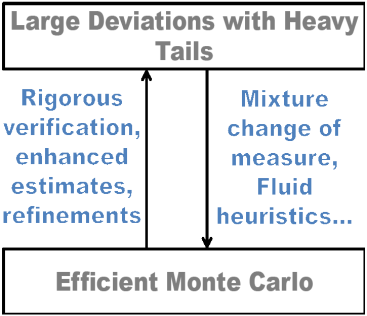
Keep in mind the Big Picture

- Efficient changes of measure for *heavy tails*
- Optimal complexity properties **(bounded rel. error & optimal termination time)**



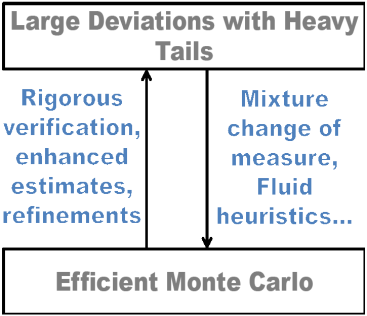
Keep in mind the Big Picture

- Efficient changes of measure for *heavy tails*
- Optimal complexity properties **(bounded rel. error & optimal termination time)**
- Methodology based on Lyapunov inequalities

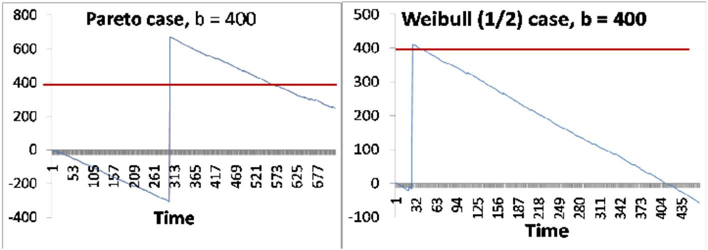


Keep in mind the Big Picture

- Efficient changes of measure for *heavy tails*
- Optimal complexity properties **(bounded rel. error & optimal termination time)**
- Methodology based on Lyapunov inequalities
- Supporting conditional limit theorems & sampling



Pareto vs Weibullian tails



How to control "rogue" paths

- 2 MIXTURES: Reg. varying – Need two mixtures & introduce $a \in (0, 1)$ for rogue paths

How to control "rogue" paths

- 2 MIXTURES: Reg. varying – Need two mixtures & introduce $a \in (0,1)$ for rogue paths
- ROLE OF $a \in (0,1)$ DCT: Analysis shows $a \in (0,1)$ for the **D**om. **C**onv. **T**hm.

$$\left|\frac{\partial g\left(s+\xi\right)}{\partial g\left(s\right)}\right| \leq const \left|\frac{\overline{F}\left(b-s-a\left(b-s\right)\right)}{\overline{F}\left(b-s\right)}\right| \leq O\left(1\right)$$

How to control "rogue" paths

- 2 MIXTURES: Reg. varying – Need two mixtures & introduce $a \in (0,1)$ for rogue paths
- ROLE OF $a \in (0,1)$ DCT: Analysis shows $a \in (0,1)$ for the **D**om. **C**onv. **T**hm.

$$\left| \frac{\partial g\left(s+\xi\right)}{\partial g\left(s\right)} \right| \leq const \left| \frac{\overline{F}\left(b-s-a\left(b-s\right)\right)}{\overline{F}\left(b-s\right)} \right| \leq O\left(1\right)$$

- Weibull case clearly 2 mixtures are not enough...

How to control "rogue" paths

- 2 MIXTURES: Reg. varying – Need two mixtures & introduce $a \in (0,1)$ for rogue paths
- ROLE OF $a \in (0,1)$ DCT: Analysis shows $a \in (0,1)$ for the **D**om. **C**onv. **T**hm.

$$\left| \frac{\partial g(s+\xi)}{\partial g(s)} \right| \leq const \left| \frac{\overline{F}(b-s-a(b-s))}{\overline{F}(b-s)} \right| \leq O(1)$$

- Weibull case clearly 2 mixtures are not enough...
- How to deal with different scales issue?

How to control "rogue" paths

- 2 MIXTURES: Reg. varying – Need two mixtures & introduce $a \in (0,1)$ for rogue paths
- ROLE OF $a \in (0,1)$ DCT: Analysis shows $a \in (0,1)$ for the **D**om. **C**onv. **T**hm.

$$\left|\frac{\partial g\left(s+\xi\right)}{\partial g\left(s\right)}\right| \leq const \left|\frac{\overline{F}\left(b-s-a\left(b-s\right)\right)}{\overline{F}\left(b-s\right)}\right| \leq O\left(1\right)$$

- Weibull case clearly 2 mixtures are not enough...
- How to deal with different scales issue?
- Can one even do finitely many mixtures?

The mixture family: Assumptions

- A1 $=$ NOT lighter than Weibull(β), $\beta \in (0,1)$

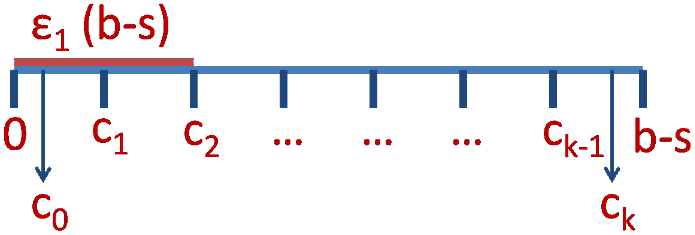
The mixture family: Assumptions

- A1 = NOT lighter than Weibull(β), $\beta \in (0, 1)$
- A2 = lighter than any Pareto

The mixture family: Assumptions

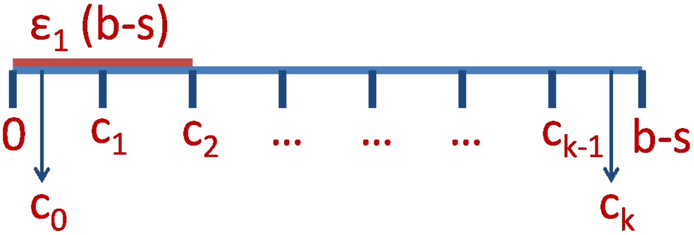
- A1 $=$ NOT lighter than Weibull(β), $\beta \in (0,1)$
- A2 $=$ lighter than any Pareto
- A3 eventually concave hazard function $\Lambda(\cdot)$

The mixture family: Construction



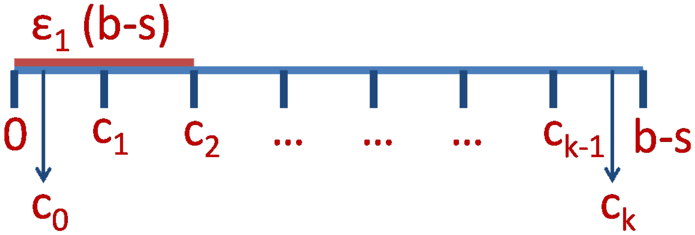
- $k+2$ - mixtures ($k = 0$ if Reg. Var. & $k \nearrow \infty$ if $\beta \nearrow 1$)

The mixture family: Construction



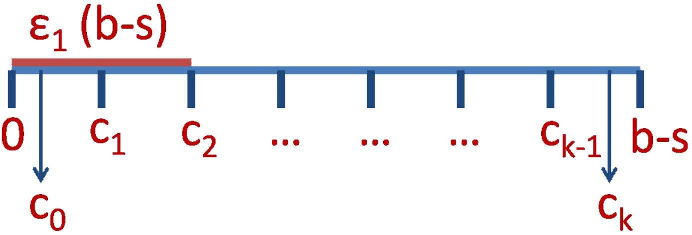
- $k+2$ - mixtures ($k = 0$ if Reg. Var. & $k \nearrow \infty$ if $\beta \nearrow 1$)
- $(-\infty, c_0]$ —> **regular increment**

The mixture family: Construction



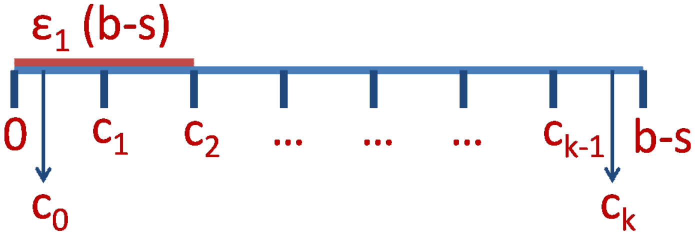
- $k+2$ - mixtures ($k=0$ if Reg. Var. & $k \nearrow \infty$ if $\beta \nearrow 1$)
- $(-\infty, c_0]$ —> **regular increment**
- $(c_0, c_1]$ —> transition component

The mixture family: Construction



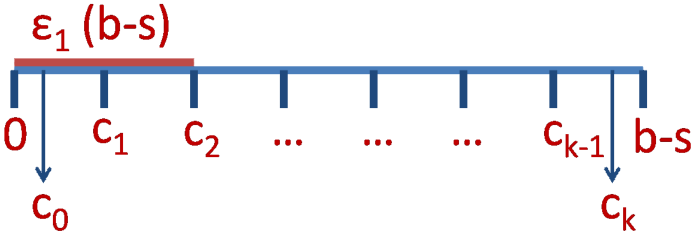
- $k+2$ - mixtures ($k=0$ if Reg. Var. & $k \nearrow \infty$ if $\beta \nearrow 1$)
- $(-\infty, c_0]$ —> **regular increment**
- $(c_0, c_1]$ —> transition component
- $(c_{j-1}, c_j]$, $j = 2, ..., k-1$ —> interpolating components

The mixture family: Construction



- $k+2$ - mixtures ($k=0$ if Reg. Var. & $k \nearrow \infty$ if $\beta \nearrow 1$)
- $(-\infty, c_0]$ —> **regular increment**
- $(c_0, c_1]$ —> transition component
- $(c_{j-1}, c_j]$, $j = 2, ..., k-1$ —> interpolating components
- $(c_{k-1}, c_k]$ —> transition component

The mixture family: Construction



- $k+2$ - mixtures ($k=0$ if Reg. Var. & $k \nearrow \infty$ if $\beta \nearrow 1$)
- $(-\infty, c_0]$ —> **regular increment**
- $(c_0, c_1]$ —> transition component
- $(c_{j-1}, c_j]$, $j = 2, ..., k-1$ —> interpolating components
- $(c_{k-1}, c_k]$ —> transition component
- (c_k, ∞) —> **large jump component**

Mixture family: Regular component

- Regular component: 3 STEP procedure...

Mixture family: Regular component

- Regular component: 3 STEP procedure...
- $c_0 = b - s - \Lambda^{-1}\left(\Lambda\left(b - s\right) - a_*\right)$ for $a_* > 0$

Mixture family: Regular component

- Regular component: 3 STEP procedure...
- $c_0 = b - s - \Lambda^{-1}\left(\Lambda\left(b-s\right) - a_*\right)$ for $a_* > 0$
- $X|X \leq c_0$: **Taylor expansion, DCT argument**

$$\left|\frac{\overline{F}\left(b-s-c_0\right)}{\overline{F}\left(b-s\right)}\right| = \left|\frac{\overline{F}\left(\Lambda^{-1}\left(\Lambda\left(b-s\right)-a_*\right)\right)}{\overline{F}\left(b-s\right)}\right| = \exp\left(a_*\right)$$

Mixture family: Large jump component

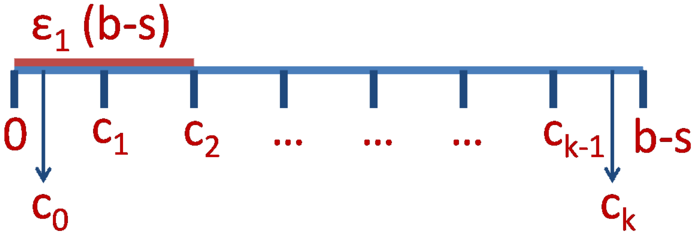
- $c_k = \Lambda^{-1}\left(\Lambda\left(b - s\right) - a_{**}\right)$ for $a_{**} > 0$

Mixture family: Large jump component

- $c_k = \Lambda^{-1}\left(\Lambda\left(b - s\right) - a_{**}\right)$ for $a_{**} > 0$
- $X|X > c_k$: **Capturing rogue paths**

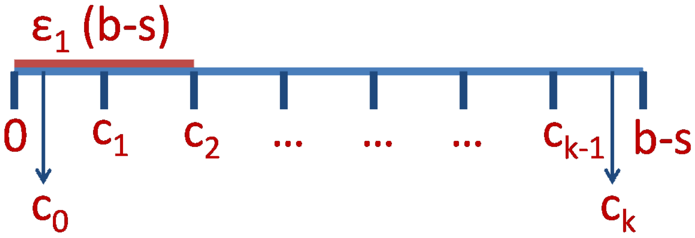
$$P\left(X > b - s | X > c_k\right) = \frac{\overline{F}\left(b - s\right)}{\overline{F}\left(\Lambda^{-1}(\Lambda(b - s) - a_{**})\right)} = \exp\left(-a_{**}\right)$$

The mixture family: Interpolating components



- Designed to be negligible when doing 3 STEP verification procedure

The mixture family: interpolating components

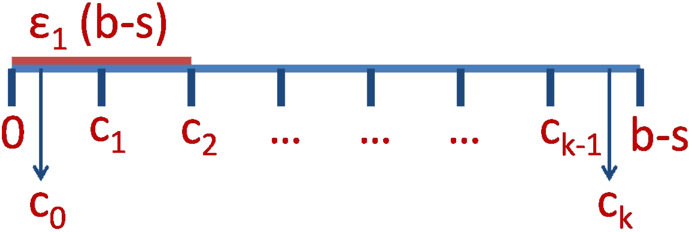


- Pick $\varepsilon_1, \varepsilon_2 > 0$ & $a_{j+1} = a_j + \varepsilon_1/2$ with

$$a_j^\beta + (1 - a_{j+1})^\beta \geq 1 + \varepsilon_2.$$

and $a_{k-1} \geq 1 - \varepsilon_1$, $a_1 \leq \varepsilon_1$

The mixture family: interpolating components



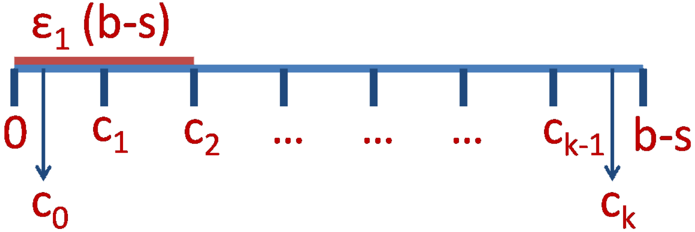
- Pick $\varepsilon_1, \varepsilon_2 > 0$ & $a_{j+1} = a_j + \varepsilon_1/2$ with

$$a_j^\beta + (1 - a_{j+1})^\beta \geq 1 + \varepsilon_2.$$

  and $a_{k-1} \geq 1 - \varepsilon_1$, $a_1 \leq \varepsilon_1$
- Existence since $x \in (0, 1)$ implies

$$x^\beta + (1 - x)^\beta > 1$$

The mixture family: interpolating components



- Pick $\varepsilon_1, \varepsilon_2 > 0$ & $a_{j+1} = a_j + \varepsilon_1/2$ with

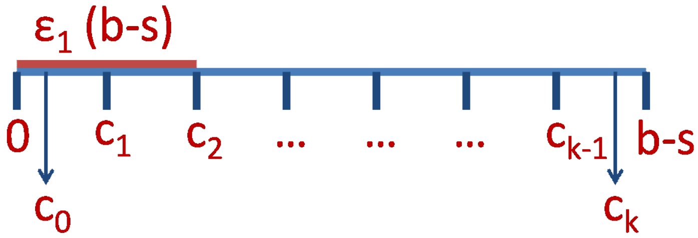
$$a_j^\beta + (1 - a_{j+1})^\beta \geq 1 + \varepsilon_2.$$

  and $a_{k-1} \geq 1 - \varepsilon_1,\ a_1 \leq \varepsilon_1$
- Existence since $x \in (0,1)$ implies

$$x^\beta + (1-x)^\beta > 1$$

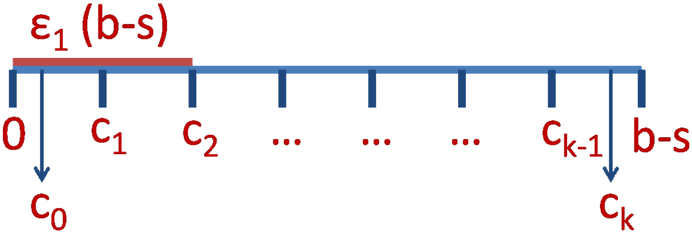
- $c_1 = a_1\,(b-s)\,, \ldots, c_{k-1} = a_{k-1}\,(b-s)$

The mixture family: transition components



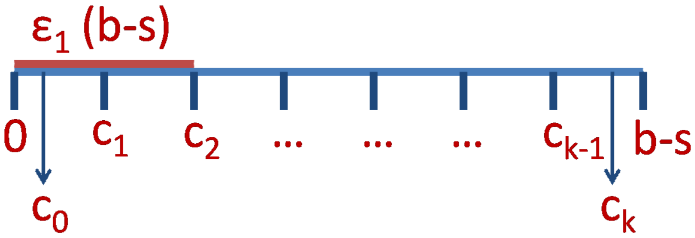
- Required since $c_0 = o\left(b - s\right)$

The mixture family: transition components



- Required since $c_0 = o\left(b - s\right)$
- $X|X \in (c_0, c_1]$ —> transition component

The mixture family: transition components



- Required since $c_0 = o\,(b - s)$
- $X | X \in (c_0, c_1]$ —> transition component
- $b - s - X | b - s - X \in (c_{k-1}, c_k]$ —> transition component

Summary of results on bounded relative error and termination time

Theorem (B. & Liu)

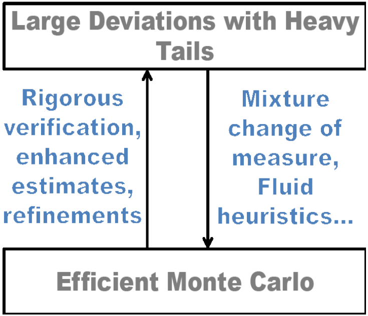
if X_i's A1-A3 apply then can select mixture parameters using Lyapunov inequalities to guarantee: A) Total variation approximation (asymptotically zero relative error), B) Bounded relative error and $O\left(b^{1-\beta}\right)$ termination time. If in addition X_i is in MDA of Gumble law then

$$\left(\frac{S_{u\tau_b}+\mu\tau_b}{\sqrt{\tau_b}},\Lambda'\left(b\right)\times\left(S_{\tau_b}-b\right),\Lambda'\left(b\right)\times\tau_b\right)$$
$$\Longrightarrow\quad\left(\sigma B\left(u\right),Z_1,Z_2\right)$$

on $D(0,1)\times R\times R$ as $b\nearrow\infty$, where Z_1 and Z_2 are independent exponentials.

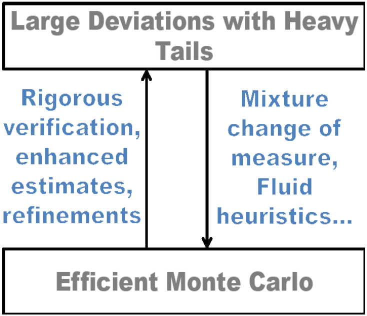
Keep in mind the Big Picture!

- Efficient changes of measure for *heavy tails*



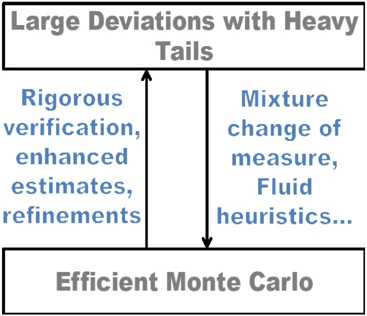
Keep in mind the Big Picture!

- Efficient changes of measure for *heavy tails*
- Optimal complexity properties **(bounded rel. error & optimal termination time)**



Keep in mind the Big Picture!

- Efficient changes of measure for *heavy tails*
- Optimal complexity properties **(bounded rel. error & optimal termination time)**
- Methodology based on Lyapunov inequalities and fluid heuristics



Keep in mind the Big Picture!

- Efficient changes of measure for *heavy tails*
- Optimal complexity properties **(bounded rel. error & optimal termination time)**
- Methodology based on Lyapunov inequalities and fluid heuristics
- Supporting conditional limit theorems & sampling