Splitting for Heavy-tailed Stochastic Processes

Jose Blanchet (based on work with Y. Shi)

Columbia University

December, 2013

Outline

Goal of the talk

- **Our Goal**: Study splitting strategies for heavy-tailed systems.

Outline

Light-tailed Systems: Random Walk Dynamics

- Assume $\psi(\theta) = \log E \exp(\theta X_i)$ finite around zero $<-$ X_i's are increments of random walk.

Light-tailed Systems: Random Walk Dynamics

- Assume $\psi(\theta) = \log E \exp(\theta X_i)$ finite around zero <– X_i's are increments of random walk.
- *Efficient importance sampling: Subsolution approach given rare event of interest MUST find classical (smooth)* $I(x, 1-t)$ *satisfying:*

$$
\begin{aligned}
2\psi(-I_x(x,1-t)/2) &\leq -I_t(x,1-t) \\
I(0,0) &\geq 2J(0,0) \quad \text{where } J(0,0) \text{ is large deviations}
\end{aligned}
$$

Light-tailed Systems: Random Walk Dynamics

- Assume $\psi(\theta) = \log E \exp(\theta X_i)$ finite around zero <– X_i's are increments of random walk.
- *Efficient importance sampling: Subsolution approach given rare event of interest MUST find classical (smooth)* $I(x, 1-t)$ *satisfying:*

$$\begin{aligned} 2\psi\left(-I_x(x,1-t)/2\right) &\leq -I_t(x,1-t) \\ I(0,0) &\geq 2J(0,0) \quad \text{where } J(0,0) \text{ is large deviations} \end{aligned}$$

- *Efficient splitting: Suffices to solve equation in weak sense.*

Light-tailed Systems: Random Walk Dynamics

- Assume $\psi(\theta) = \log E \exp(\theta X_i)$ finite around zero <– X_i's are increments of random walk.
- *Efficient importance sampling: Subsolution approach given rare event of interest MUST find classical (smooth)* $I(x, 1-t)$ *satisfying:*

$$
\begin{aligned}
2\psi\left(-I_x(x, 1-t)/2\right) &\leq -I_t(x, 1-t) \\
I(0,0) &\geq 2J(0,0) \quad \text{where } J(0,0) \text{ is large deviations}
\end{aligned}
$$

- *Efficient splitting: Suffices to solve equation in weak sense.*
- See Dupuis & Wang '09, Dean and Dupuis '10.

Light-tailed Systems: Random Walk Dynamics

- Assume $\psi(\theta) = \log E \exp(\theta X_i)$ finite around zero <– X_i's are increments of random walk.
- *Efficient importance sampling: Subsolution approach given rare event of interest MUST find classical (smooth)* $I(x, 1-t)$ *satisfying:*

$$\begin{aligned} 2\psi\left(-I_x(x,1-t)/2\right) &\leq -I_t(x,1-t) \\ I(0,0) &\geq 2J(0,0) \quad \text{where } J(0,0) \text{ is large deviations} \end{aligned}$$

- *Efficient splitting: Suffices to solve equation in weak sense.*
- See Dupuis & Wang '09, Dean and Dupuis '10.
- **Moral:** *Efficient* splitting easier than *efficient* importance sampling

Efficiency in Splitting vs Importance Sampling in Light-tailed Systems

- Efficiency means (traditionally) number of replications grows *subexponentially* to get, say 5% precision with 95% confidence.

Efficiency in Splitting vs Importance Sampling in Light-tailed Systems

- Efficiency means (traditionally) number of replications grows *subexponentially* to get, say 5% precision with 95% confidence.
- In Crude Monte Carlo's number of replications grows *exponentially*.

Efficiency in Splitting vs Importance Sampling in Light-tailed Systems

- Efficiency means (traditionally) number of replications grows *subexponentially* to get, say 5% precision with 95% confidence.
- In Crude Monte Carlo's number of replications grows *exponentially*.
- But what if you do a sharp analysis (i.e. beyond exponential rates, down to polynomials)?

Efficiency in Splitting vs Importance Sampling in Light-tailed Systems

- Efficiency means (traditionally) number of replications grows *subexponentially* to get, say 5% precision with 95% confidence.
- In Crude Monte Carlo's number of replications grows *exponentially*.
- But what if you do a sharp analysis (i.e. beyond exponential rates, down to polynomials)?
- Efficient splitting LESS efficient than well crafted importance sampling: Jackson networks sharp analysis in Blanchet, Leder, Shi (2009) & Blanchet (2013).

Efficiency in Splitting vs Importance Sampling in Light-tailed Systems

- Efficiency means (traditionally) number of replications grows *subexponentially* to get, say 5% precision with 95% confidence.
- In Crude Monte Carlo's number of replications grows *exponentially*.
- But what if you do a sharp analysis (i.e. beyond exponential rates, down to polynomials)?
- Efficient splitting LESS efficient than well crafted importance sampling: Jackson networks sharp analysis in Blanchet, Leder, Shi (2009) & Blanchet (2013).
- **Moral:** *Simplicity comes with an price in (somewhat modest) efficiency loss...*

Motivating question

What about heavy tails? Is "efficient splitting" somehow easier than "efficient importance sampling" even if there is some (modest) loss in efficiency?

Outline

Importace Sampling Benchmark

- Correct family of importance sampling distributions

$$
\begin{aligned}
f(x|s) &= p_0 \frac{f_X(x)\, I(x \leq (b-s) - A(s))}{P(X \leq (b-s) - A(s))} \\
&+ p_1 \frac{f_X(x)\, I(x \in ((b-s) - A(s), a_1(b-s)])}{P(X \in ((b-s) - A(s), a_1(b-s)])} \\
&+ \sum_{k=1}^{m-1} p_{k+1}; \frac{f_X(x)\, I(x \in (a_k(b-s), a_{k+1}(b-s)])}{P(X \in (a_k(b-s), a_{k+1}(b-s)])} \\
&+ p_{m+1} \frac{f_X(b-s-x)\, I(x \in (a_m(b-s), b-s-A(b-s)]}{P(X \in (a_m(b-s), b-s-A(b-s)])} \\
&+ p_{m+2} \frac{f_X(b-s-x)\, I(x > b-s-A(b-s))}{P(X > b-s-A(b-s))}.
\end{aligned}
$$

Importace Sampling Benchmark

- Correct family of importance sampling distributions

$$
\begin{aligned}
f(x|s) &= p_0 \frac{f_X(x)\, I(x \le (b-s) - A(s))}{P(X \le (b-s) - A(s))} \\
&+ p_1 \frac{f_X(x)\, I(x \in ((b-s) - A(s), a_1(b-s)])}{P(X \in ((b-s) - A(s), a_1(b-s)])} \\
&+ \sum_{k=1}^{m-1} p_{k+1}; \frac{f_X(x)\, I(x \in (a_k(b-s), a_{k+1}(b-s)])}{P(X \in (a_k(b-s), a_{k+1}(b-s)])} \\
&+ p_{m+1} \frac{f_X(b-s-x)\, I(x \in (a_m(b-s), b-s-A(b-s)]}{P(X \in (a_m(b-s), b-s-A(b-s)])} \\
&+ p_{m+2} \frac{f_X(b-s-x)\, I(x > b-s-A(b-s))}{P(X > b-s-A(b-s))}.
\end{aligned}
$$

- $A(\cdot)$ is a sublinear function called auxiliary function, a_k's recursively defined, m smaller for heavier tails.

Importace Sampling Benchmark

- Specification depends on asymptotic behavior of hazard rate function.

Importace Sampling Benchmark

- Specification depends on asymptotic behavior of hazard rate function.
- General: Blanchet and Liu (2012); Pareto case: Dupuis, Leder, Wang (2009).

Importace Sampling Benchmark

- Specification depends on asymptotic behavior of hazard rate function.
- General: Blanchet and Liu (2012); Pareto case: Dupuis, Leder, Wang (2009).
- **Moral:** *Importance sampling depends on fine structure of hazard rates, and needs somewhat complex specification...*

Importace Sampling Benchmark

- Specification depends on asymptotic behavior of hazard rate function.
- General: Blanchet and Liu (2012); Pareto case: Dupuis, Leder, Wang (2009).
- **Moral:** *Importance sampling depends on fine structure of hazard rates, and needs somewhat complex specification...*
- *BUT:* strongly efficient – often asymptotically zero relative error.

Outline

Splitting Initial Thoughts and Considerations

- Traditional splitting: At time k, if particle random walk reaches *milestone* C_k, replace particle walk is r **children** & reweight each children using
$$w_{new} = w_{old} / r.$$

Splitting Initial Thoughts and Considerations

- Traditional splitting: At time k, if particle random walk reaches *milestone* C_k, replace particle walk is r **children** & reweight each children using

$$w_{new} = w_{old}/r.$$

- After n steps estimator is

$$\sum_k I\left(\text{particle}\left(k\right) \in \text{Set of interest}\right) \frac{1}{r^n}.$$

Splitting Initial Thoughts and Considerations

- Traditional splitting: At time k, if particle random walk reaches *milestone* C_k, replace particle walk is r **children** & reweight each children using

$$w_{new} = w_{old}/r.$$

- After n steps estimator is

$$\sum_k I\left(\text{particle}\left(k\right) \in \text{Set of interest}\right)\frac{1}{r^n}.$$

- Key property: Rewards walks that behave according to conditional distribution given rare event of interest...

Splitting Initial Thoughts and Considerations

- Traditional splitting: At time k, if particle random walk reaches *milestone* C_k, replace particle walk is r **children** & reweight each children using
$$w_{new} = w_{old}/r.$$
- After n steps estimator is
$$\sum_k I\left(\text{particle}\left(k\right) \in \text{Set of interest}\right)\frac{1}{r^n}.$$
- Key property: Rewards walks that behave according to conditional distribution given rare event of interest...
- **Problem:** In heavy tailed systems "normal" behavior is typical conditional behavior prior to extreme event... (i.e. no warning).

Outline

Splitting for Heavy-tailed Walks: Two Approaches

- Branching processes in "hazard rate" space:

$$\begin{array}{rcl} \mathbf{P}\left(X_1+...+X_n>b\right) & \longleftarrow & \textbf{PROBLEM OF INTEREST} \\ \text{WE'LL USE:} & & P\left(X>t\right)=P\left(\Lambda^{-1}\left(V\right)>t\right) \\ V & \sim & Exp\left(1\right). \\ \text{Pareto example:} & & \Lambda^{-1}\left(v\right)=\exp\left(v/\alpha\right)-1. \end{array}$$

Splitting for Heavy-tailed Walks: Two Approaches

- Branching processes in "hazard rate" space:

$$\mathbf{P}\left(X_1 + ... + X_n > b\right) \quad \longleftarrow \quad \textbf{PROBLEM OF INTEREST}$$

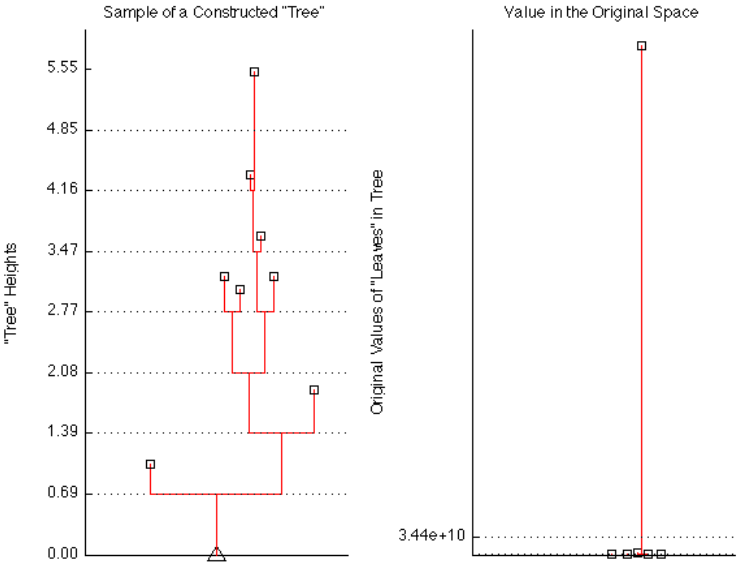
$$\text{WE'LL USE:} \qquad P\left(X > t\right) = P\left(\Lambda^{-1}\left(V\right) > t\right)$$

$$V \sim Exp\left(1\right).$$

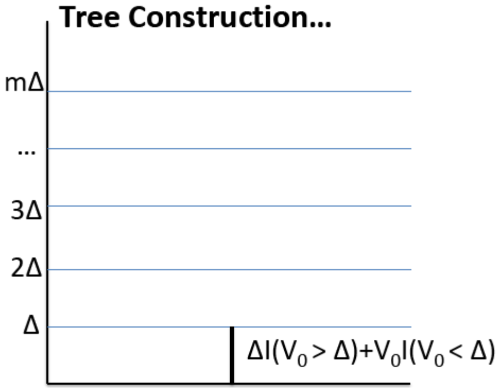
$$\text{Pareto example:} \qquad \Lambda^{-1}\left(v\right) = \exp\left(v/\alpha\right) - 1.$$

- New (not today): Branching and conditional Monte Carlo: Based on "normal" behavior prior to jump...

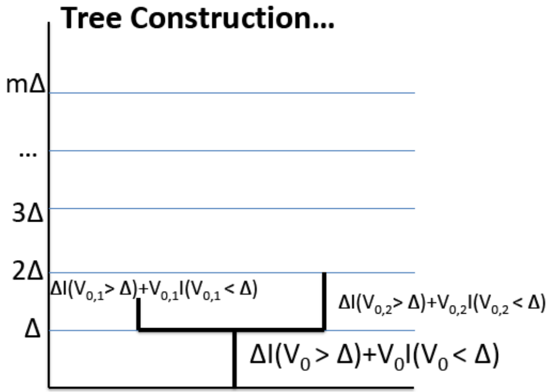
Splitting I: Branching Processes in Hazard Rate Space



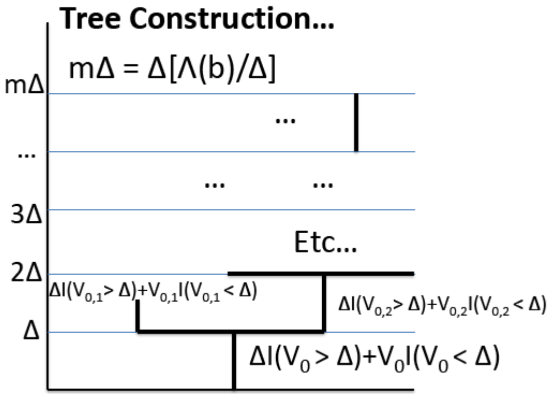
Splitting Estimator: Construction



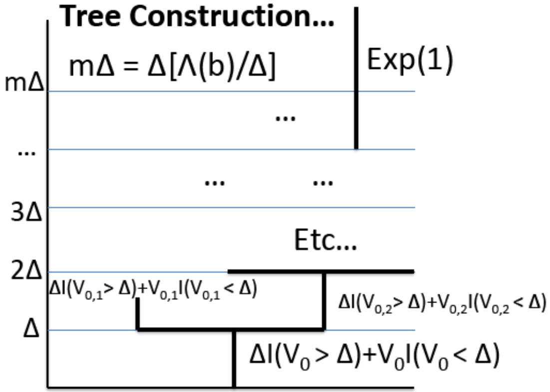
Splitting Estimator: Construction



Splitting Estimator: Construction



Splitting Estimator: Construction



Splitting Estimator: Final Form

- Step 1: Let $\Pi_m(0) =$ Tree rooted at origin as discussed. Let

$$\begin{aligned}
\overline{\Pi}_1 &= \Pi_m(0) \\
\mathcal{L}\left(\overline{\Pi}_1\right) &= \text{ leaves on top of } \overline{\Pi}_1 \\
\{\Pi_m(s) &: s \in \mathcal{L}\left(\overline{\Pi}_1\right)\} \text{ i.i.d. copies of } \textit{tree} \textbf{ BUT rooted at } s
\end{aligned}$$

Splitting Estimator: Final Form

- Step 1: Let $\Pi_m(0) =$ Tree rooted at origin as discussed. Let

$$\begin{aligned} \overline{\Pi}_1 &= \Pi_m(0) \\ \mathcal{L}\left(\overline{\Pi}_1\right) &= \text{ leaves on top of } \overline{\Pi}_1 \\ \{\Pi_m(s) &: s \in \mathcal{L}\left(\overline{\Pi}_1\right)\} \text{ i.i.d. copies of } \textit{tree} \textbf{ BUT rooted at } s \end{aligned}$$

- Step 2:

$$\overline{\Pi}_k = \cup_{s\in\mathcal{L}\left(\overline{\Pi}_{k-1}\right)} \Pi_m(s).$$

Splitting Estimator: Final Form

- Step 1: Let $\Pi_m(0)$ = Tree rooted at origin as discussed. Let

$$\begin{aligned} \overline{\Pi}_1 &= \Pi_m(0) \\ \mathcal{L}\left(\overline{\Pi}_1\right) &= \text{ leaves on top of } \overline{\Pi}_1 \\ \{\Pi_m(s) &: s \in \mathcal{L}\left(\overline{\Pi}_1\right)\} \text{ i.i.d. copies of } \textit{tree} \textbf{ BUT rooted at } s \end{aligned}$$

- Step 2:

$$\overline{\Pi}_k = \cup_{s \in \mathcal{L}\left(\overline{\Pi}_{k-1}\right)} \Pi_m(s).$$

- Step 3: Get $\overline{\Pi}_n$ and for each $s \in \mathcal{L}\left(\overline{\Pi}_n\right)$ define

$$((L_1(s), V_1(s)), ..., (L_n(s), V_n(s)))$$

where $L_i(s)$ is the length produced in the i-th tree in the genealogy of s.

Splitting Estimator: Final Form

- Step 1: Let $\Pi_m(0) =$ Tree rooted at origin as discussed. Let

$$\begin{aligned} \overline{\Pi}_1 &= \Pi_m(0) \\ \mathcal{L}\left(\overline{\Pi}_1\right) &= \text{ leaves on top of } \overline{\Pi}_1 \\ \{\Pi_m(s) &: s \in \mathcal{L}\left(\overline{\Pi}_1\right)\} \text{ i.i.d. copies of } \textit{tree} \textbf{ BUT rooted at } s \end{aligned}$$

- Step 2:

$$\overline{\Pi}_k = \cup_{s \in \mathcal{L}\left(\overline{\Pi}_{k-1}\right)} \Pi_m(s).$$

- Step 3: Get $\overline{\Pi}_n$ and for each $s \in \mathcal{L}\left(\overline{\Pi}_n\right)$ define

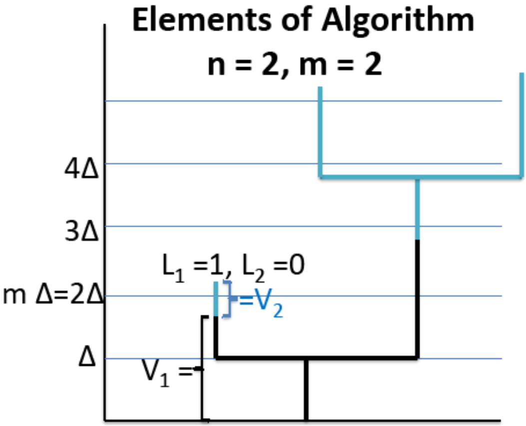
$$((L_1(s), V_1(s)), ..., (L_n(s), V_n(s)))$$

where $L_i(s)$ is the length produced in the i-th tree in the genealogy of s.

- **Estimator:**

$$\widehat{p}_b = \sum_{s \in \mathcal{L}(\overline{\Pi}_n)} r^{-L_1(s) - ... - L_n(s)} I\left(\Lambda^{-1}(V_1(s)) + ...\Lambda^{-1}(V_n(s)) > b\right).$$

Splitting Estimator: Picture of Elements



Outline

Splitting Estimator: How to Select Parameters?

- Which parameters are needed to create the estimator?

Splitting Estimator: How to Select Parameters?

- Which parameters are needed to create the estimator?
- Correct, Δ and r... What do we want? Efficiency (work normalized variance):

$$E\left(\widehat{p}_b^2\right) \times E\left|\mathcal{L}(\overline{\Pi}_n)\right| \approx O(P\left(X_1 + ... + X_n > b\right)^2).$$

Splitting Estimator: How to Select Parameters?

- Which parameters are needed to create the estimator?
- Correct, Δ and r... What do we want? Efficiency (work normalized variance):

$$E\left(\widehat{p}_b^2\right) \times E\left|\mathcal{L}(\overline{\Pi}_n)\right| \approx O(P\left(X_1 + ... + X_n > b\right)^2).$$

- Let's see... (elementary Galton-Watson processes)

$$\begin{aligned} E\left|\mathcal{L}(\Pi_{k+1}(0))\right| &\approx rP\left(V > \Delta\right) E\left|\mathcal{L}(\Pi_k(0))\right| \\ E\left|\mathcal{L}(\Pi_m(0))\right| &\approx \left(r\exp\left(-\Delta\right)\right)^m \approx r^m \exp\left(-\Lambda\left(b\right)\right) \\ \text{Recall } m &= \left[\Lambda\left(b\right)/\Delta\right]. \end{aligned}$$

Splitting Estimator: How to Select Parameters?

- Which parameters are needed to create the estimator?
- Correct, Δ and r... What do we want? Efficiency (work normalized variance):

$$E\left(\widehat{p}_b^2\right) \times E\left|\mathcal{L}(\overline{\Pi}_n)\right| \approx O(P\left(X_1 + ... + X_n > b\right)^2).$$

- Let's see... (elementary Galton-Watson processes)

$$\begin{aligned} E\left|\mathcal{L}(\Pi_{k+1}(0))\right| &\approx rP(V > \Delta) E\left|\mathcal{L}(\Pi_k(0))\right| \\ E\left|\mathcal{L}(\Pi_m(0))\right| &\approx (r\exp(-\Delta))^m \approx r^m \exp(-\Lambda(b)) \\ \text{Recall } m &= [\Lambda(b)/\Delta]. \end{aligned}$$

- How would you pick r and Δ?

Splitting Estimator: Efficiency Result

Theorem

Assuming $d\Lambda(t)/dt$ eventually non-increasing we obtain that for each $\varepsilon > 0$. Suppose that $r \exp(-\Delta) = 1$, then

$$E\left(\widehat{p}_b^2\right) \times E\left|\mathcal{L}(\overline{\Pi}_n)\right| = O(P\left(X_1 + ... + X_n > b\right)^{2-\varepsilon})$$

as $b \to \infty$.

Why Decreasing Hazard Rates?

- Where did we use heavy-tailed property?

$$\widehat{p}_b = \sum_{s\in\mathcal{L}(\overline{\Pi}_n)} r^{-L_1(s)-...-L_n(s)} I\left(\Lambda^{-1}(V_1(s)) + ...\Lambda^{-1}(V_n(s)) > b\right),$$

$$\widehat{p}_b^2 = \sum_{s\in\mathcal{L}(\overline{\Pi}_n)} \sum_{u\in\mathcal{L}(\overline{\Pi}_n)} r^{-L_1(s)-...-L_n(s)} r^{-L_1(u)-...-L_n(u)} \mathbf{I}(s)\,\mathbf{I}(u).$$

Why Decreasing Hazard Rates?

- Where did we use heavy-tailed property?

$$\widehat{p}_b = \sum_{s \in \mathcal{L}(\overline{\Pi}_n)} r^{-L_1(s)-\ldots-L_n(s)} I\left(\Lambda^{-1}(V_1(s)) + \ldots \Lambda^{-1}(V_n(s)) > b\right),$$

$$\widehat{p}_b^2 = \sum_{s \in \mathcal{L}(\overline{\Pi}_n)} \sum_{u \in \mathcal{L}(\overline{\Pi}_n)} r^{-L_1(s)-\ldots-L_n(s)} r^{-L_1(u)-\ldots-L_n(u)} \mathbf{I}(s)\, \mathbf{I}(u).$$

- Condition on last common ancestor of s and u, say t

$$\begin{aligned} E(\mathbf{I}(s)\,\mathbf{I}(u)\,|\text{tree up to } t) &= P^2\Big(X_1 + \ldots + X_{n-k(t)} > b - l_1 - \ldots - \\ &= \exp\left(-2\Lambda\left(b - l_1 - \ldots - l_t\right)\right) \\ &\leq \exp\left(-2\Lambda(b) + 2\Delta L_1 + \ldots + 2\Delta L_t\right) \\ \text{Due to } \Lambda(b-x) &\leq \Lambda(b) - \Lambda(x) \ \text{ because of decreasing ha} \end{aligned}$$

Conclusions

- Light tails: Splitting easier than importance sampling –> weak vs classical sense solutions).

Conclusions

- Light tails: Splitting easier than importance sampling –> weak vs classical sense solutions).
- Heavy tails: Splitting easier than importance sampling? (Uniform setup vs fine information on tails).

Conclusions

- Light tails: Splitting easier than importance sampling –> weak vs classical sense solutions).
- Heavy tails: Splitting easier than importance sampling? (Uniform setup vs fine information on tails).
- Analysis of natural splitting estimator on hazard space.