Importance Sampling

Estimation Theory Project 2

Braghadeesh Lakshminarayanan

November 1, 2021

Division of DCS, KTH Royal Institute of Technology

Table of contents

Motivation

Motivation

Motivation

- Consider finding the expectation of a function of a random variable

Motivation

- Consider finding the expectation of a function of a random variable

$$\mu = \mathbb{E}\left[f(X)\right] = \int_{x \in \mathcal{X}} f(x)p(x)dx$$

Motivation

- Consider finding the expectation of a function of a random variable

$$\mu = \mathbb{E}\left[f(X)\right] = \int_{x \in \mathcal{X}} f(x)p(x)dx$$

- Why are we interested in calculating expectation?

Motivation

- Consider finding the expectation of a function of a random variable

$$\mu = \mathbb{E}\left[f(X)\right] = \int_{x \in \mathcal{X}} f(x)p(x)dx$$

- Why are we interested in calculating expectation?
  - Many quantities of interest may be cast as expectation

Motivation

Motivation

- *Probabilities:*

$$\mathbb{P}\left(Y \in A\right) = \mathbb{E}\left[I_{\{A\}}(Y)\right]$$

$$\text{where } I_{\{A\}}(Y) = \begin{cases} 1, \text{if } Y \in A \\ 0, \text{if } Y \notin A \end{cases}$$

Motivation

- *Probabilities:*

$$\mathbb{P}(Y \in A) = \mathbb{E}\left[I_{\{A\}}(Y)\right]$$

$$\text{where } I_{\{A\}}(Y) = \begin{cases} 1, \text{if } Y \in A \\ 0, \text{if } Y \notin A \end{cases}$$

- *Integrals:*

$$\int_a^b q(x)dx$$

The above integral can be computed as

$$\begin{aligned} \int_a^b q(x)dx &= (b-a)\int_a^b q(x)\frac{1}{b-a}dx \\ &= (b-a)\int_a^b q(x)P_U(x)dx \\ &= (b-a)\mathbb{E}\left[q(U)\right] \end{aligned}$$

Motivation

Motivation

- Issue in computing expectation:

Motivation

- Issue in computing expectation:
    - The explicit computation is often cumbersome

Motivation

- Issue in computing expectation:
  - The explicit computation is often cumbersome
    - The integrand may have complicated expressions!

Motivation

- Issue in computing expectation:
    - The explicit computation is often cumbersome
        - The integrand may have complicated expressions!
- Remedy?

Motivation

- Issue in computing expectation:
    - The explicit computation is often cumbersome
        - The integrand may have complicated expressions!
- Remedy?
    - We can compute the expectation by simple Monte-Carlo sampling

MC Sampling

Monte-Carlo Sampling

Monte-Carlo Sampling

- Draw N samples (*i.i.d*) from the distribution P

Monte-Carlo Sampling

- Draw N samples (*i.i.d*) from the distribution P

$$X_1, \ldots, X_N \overset{i.i.d}{\sim} P$$

Monte-Carlo Sampling

- Draw N samples (*i.i.d*) from the distribution P

$$X_1, \ldots, X_N \overset{i.i.d}{\sim} P$$

- Compute the sample mean

Monte-Carlo Sampling

- Draw N samples (*i.i.d*) from the distribution P

$$X_1, \ldots, X_N \overset{i.i.d}{\sim} P$$

- Compute the sample mean

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} f(X_i)$$

Monte-Carlo Sampling

- Draw N samples (*i.i.d*) from the distribution P

$$X_1, \ldots, X_N \stackrel{i.i.d}{\sim} P$$

- Compute the sample mean

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^{N} f(X_i)$$

- Law of large numbers guarantees that as long as N is “large” enough,

Monte-Carlo Sampling

- Draw N samples (*i.i.d*) from the distribution P

$$X_1, \ldots, X_N \stackrel{i.i.d}{\sim} P$$

- Compute the sample mean

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} f(X_i)$$

- Law of large numbers guarantees that as long as N is “large” enough,

$$\hat{\mu} \stackrel{a.s}{\to} \mathbb{E}\left[f(X)\right]$$

Monte-Carlo Sampling

Monte-Carlo Sampling

- $\hat{\mu}$ is unbiased

Monte-Carlo Sampling

- $\hat{\mu}$ is unbiased
- $\text{Var}(\hat{\mu}) = \frac{\sigma^2}{N}$, where

$$\sigma^2 = \int_{x \in \mathcal{X}} f^2(x)p(x)dx - \mu^2$$

Monte-Carlo Sampling

- $\hat{\mu}$ is unbiased
- $\text{Var}(\hat{\mu}) = \frac{\sigma^2}{N}$, where

$$\sigma^2 = \int_{x \in \mathcal{X}} f^2(x) p(x) dx - \mu^2$$

- Does Monte-Carlo sampling always yield "good" approximation?

Monte-Carlo Sampling

- $\hat{\mu}$ is unbiased
- $\text{Var}(\hat{\mu}) = \frac{\sigma^2}{N}$, where

$$\sigma^2 = \int_{x \in \mathcal{X}} f^2(x)p(x)dx - \mu^2$$

- Does Monte-Carlo sampling always yield “good” approximation?

Not really!

Pitfall of Monte-Carlo Sampling

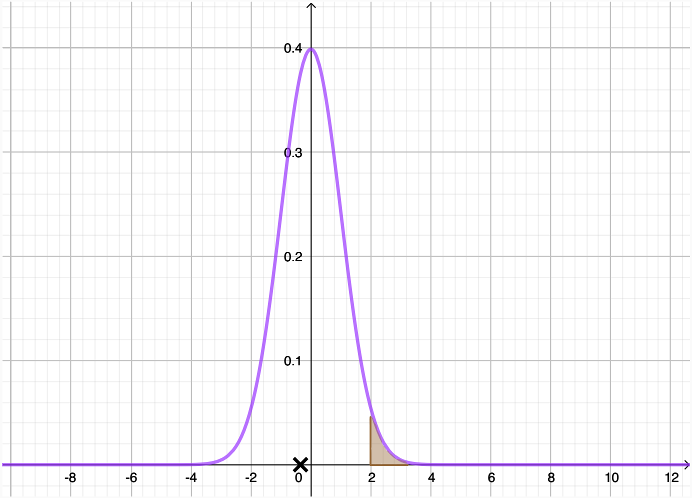
Pitfall of Monte-Carlo Sampling

Pitfall of Monte-Carlo Sampling

- Consider estimating the following tail probability

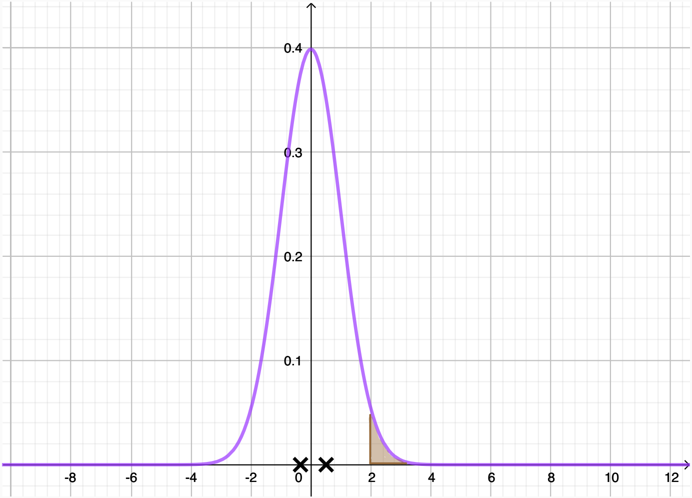
Pitfall of Monte-Carlo Sampling

- Consider estimating the following tail probability



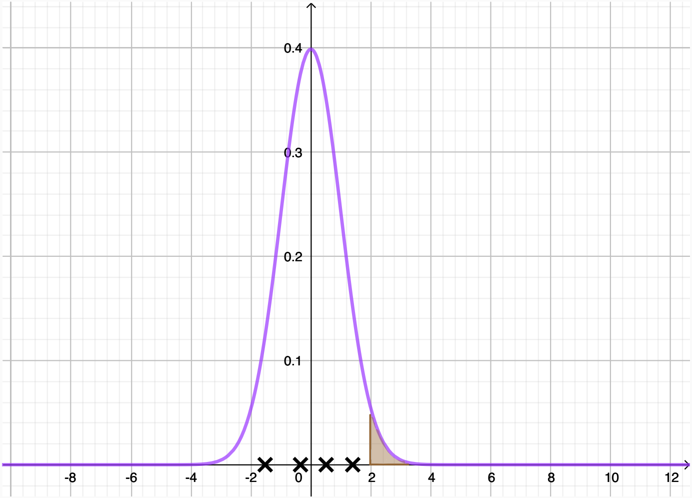
Pitfall of Monte-Carlo Sampling

- Consider estimating the following tail probability



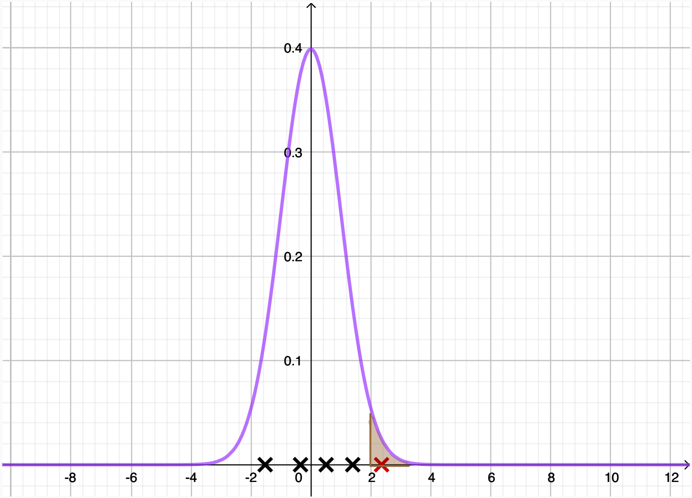
Pitfall of Monte-Carlo Sampling

- Consider estimating the following tail probability



Pitfall of Monte-Carlo Sampling

- Consider estimating the following tail probability



Pitfall of Monte-Carlo Sampling

Pitfall of Monte-Carlo Sampling

- Monte-Carlo sampling may fail to sample points from the tail

Pitfall of Monte-Carlo Sampling

- Monte-Carlo sampling may fail to sample points from the tail
- Sample mean estimate for the tail probability may have huge variance

Pitfall of Monte-Carlo Sampling

- Monte-Carlo sampling may fail to sample points from the tail
- Sample mean estimate for the tail probability may have huge variance
    - Leads to inaccurate approximation

Pitfall of Monte-Carlo Sampling

- Monte-Carlo sampling may fail to sample points from the tail
- Sample mean estimate for the tail probability may have huge variance
    - Leads to inaccurate approximation
- Remedy?

Pitfall of Monte-Carlo Sampling

- Monte-Carlo sampling may fail to sample points from the tail
- Sample mean estimate for the tail probability may have huge variance
    - Leads to inaccurate approximation
- Remedy?

Importance Sampling!

Importance Sampling

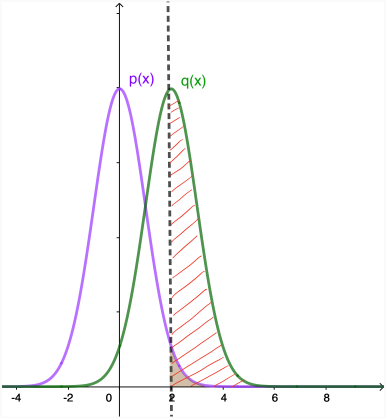
Importance Sampling

Importance Sampling

- Key idea: Sample from a different distribution q that has “heavy” tail over our region of interest

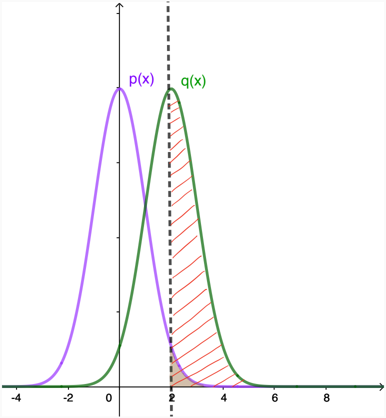
Importance Sampling

- Key idea: Sample from a different distribution q that has "heavy" tail over our region of interest



Importance Sampling

- Key idea: Sample from a different distribution q that has “heavy” tail over our region of interest



- p - true distribution, q - proposal distribution

Importance Sampling

Importance Sampling

- Recall: Original problem of finding expectation

Importance Sampling

- Recall: Original problem of finding expectation

$$\mu = \mathbb{E}[f(X)] = \int_{x \in \mathcal{X}} f(x)p(x)dx$$

Importance Sampling

- Recall: Original problem of finding expectation

$$\mu = \mathbb{E}[f(X)] = \int_{x \in \mathcal{X}} f(x)p(x)dx$$

- We can rewrite the above expression in terms of our proposal distribution q

Importance Sampling

- Recall: Original problem of finding expectation

$$\mu = \mathbb{E}[f(X)] = \int_{x \in \mathcal{X}} f(x)p(x)dx$$

- We can rewrite the above expression in terms of our proposal distribution q

$$\mu = \int_{x \in \mathcal{X}} \frac{f(x)p(x)}{q(x)} q(x)dx = \mathbb{E}_q\left[\frac{f(X)p(X)}{q(X)}\right]$$

$$X \overset{i.i.d}{\sim} q$$

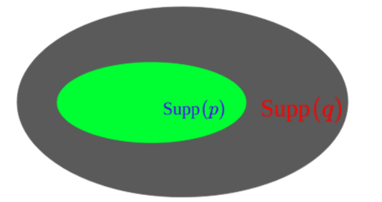
Requirements for Proposal Distribution

Requirements for Proposal Distribution

- $q(x) > 0$ whenever $f(x)p(x) \neq 0$

Requirements for Proposal Distribution

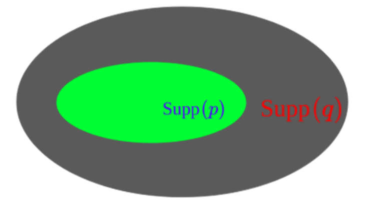
- $q(x) > 0$ whenever $f(x)p(x) \neq 0$
  - $q(x) > 0$ whenever $p(x) > 0$



- $\text{Supp(p)} = \{x \in \mathcal{X} : p(x) > 0\}$

Requirements for Proposal Distribution

- $q(x) > 0$ whenever $f(x)p(x) \neq 0$
    - $q(x) > 0$ whenever $p(x) > 0$



- $\text{Supp(p)} = \{x \in \mathcal{X} : p(x) > 0\}$
- Let $\mathcal{D}$ denote the support of p and $\mathcal{Q}$ denote the support of q

Importance Sampling Estimate

Importance Sampling Estimate

- IS estimate: $\hat{\mu}_q = \frac{1}{N}\sum_{i=1}^{N}\frac{f(X_i)p(X_i)}{q(X_i)} \quad X_i \overset{i.i.d}{\sim} q$

Importance Sampling Estimate

- IS estimate: $\hat{\mu}_q = \frac{1}{N}\sum_{i=1}^{N}\frac{f(X_i)p(X_i)}{q(X_i)} \qquad X_i \overset{i.i.d}{\sim} q$

- IS estimate is unbiased

$$\begin{aligned}
\mathbb{E}\left[\hat{\mu}_q\right] &= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_q\left[\frac{f(X_i)p(X_i)}{q(X_i)}\right] \\
&= \mathbb{E}_q\left[\frac{f(X_1)p(X_1)}{q(X_1)}\right] \\
&= \int_{\mathcal{Q}}\frac{f(x)p(x)}{q(x)}q(x)\,dx \\
&= \int_{\mathcal{Q}} f(x)p(x)\,dx
\end{aligned}$$

-

Importance Sampling Estimate

Importance Sampling Estimate

$$\begin{aligned}
\mathbb{E}\left[\hat{\mu}_q\right] &\overset{(*)}{=} \int_{\mathcal{D}} f(x)p(x)\,dx + \int_{\mathcal{Q}\cap\mathcal{D}^c} f(x)p(x)\,dx - \int_{\mathcal{D}\cap\mathcal{Q}^c} f(x)p(x)\,dx \\
&= \int_{\mathcal{D}} f(x)p(x)\,dx \\
\implies \mathbb{E}\left[\hat{\mu}_q\right] &= \mu
\end{aligned}$$

Importance Sampling Estimate

$$\begin{aligned}
\mathbb{E}\left[\hat{\mu}_q\right] &\overset{(*)}{=} \int_{\mathcal{D}} f(x)p(x)\,dx + \int_{\mathcal{Q}\cap\mathcal{D}^c} f(x)p(x)\,dx - \int_{\mathcal{D}\cap\mathcal{Q}^c} f(x)p(x)\,dx \\
&= \int_{\mathcal{D}} f(x)p(x)\,dx \\
\implies \mathbb{E}\left[\hat{\mu}_q\right] &= \mu
\end{aligned}$$

$(*)$ follows from the fact that

1. $\mathcal{Q} \cup (\mathcal{D} \cap \mathcal{Q}^c) = \mathcal{D} \cup (\mathcal{Q} \cap \mathcal{D}^c)$
2. $\mathcal{Q}$ and $\mathcal{D} \cap \mathcal{Q}^c$ are disjoint sets, so do $\mathcal{D}$ and $\mathcal{Q} \cap \mathcal{D}^c$

Variance of IS Estimate

Variance of IS Estimate

- $\text{Var}_q(\hat{\mu}_q) = \frac{{\sigma_q}^2}{N}$ where

$$
\begin{aligned}
{\sigma_q}^2 &= \int_{\mathcal{Q}} \frac{(f(x)p(x))^2}{q(x)}\, dx - \mu^2 \\
&= \int_{\mathcal{Q}} \frac{(f(x)p(x) - \mu q(x))^2}{q(x)}\, dx
\end{aligned}
$$

Variance of IS Estimate

- $\text{Var}_q(\hat{\mu}_q) = \frac{{\sigma_q}^2}{N}$ where

$$
\begin{aligned}
{\sigma_q}^2 &= \int_{\mathcal{Q}} \frac{\left(f(x)p(x)\right)^2}{q(x)}\, dx - \mu^2 \\
&= \int_{\mathcal{Q}} \frac{\left(f(x)p(x) - \mu q(x)\right)^2}{q(x)}\, dx
\end{aligned}
$$

- **Note:** Since, for IS, $\text{supp}(q) \supset \text{supp}(p)$, $\mathcal{Q}$ can be replaced by $\mathcal{D}$ in the above integral

Variance of IS Estimate

- $\text{Var}_q(\hat{\mu}_q) = \frac{{\sigma_q}^2}{N}$ where

$$
\begin{aligned}
{\sigma_q}^2 &= \int_{\mathcal{Q}} \frac{(f(x)p(x))^2}{q(x)}\, dx - \mu^2 \\
&= \int_{\mathcal{Q}} \frac{(f(x)p(x) - \mu q(x))^2}{q(x)}\, dx
\end{aligned}
$$

- **Note:** Since, for IS, $\text{supp}(q) \supset \text{supp}(p)$, $\mathcal{Q}$ can be replaced by $\mathcal{D}$ in the above integral
- since ${\sigma_q}^2$ depends on the choice of q, we can get “optimal” q that reduces the variance

Optimal Proposal Distribution

Optimal Proposal Distribution

- Optimal q, denoted by $q^* \propto |f(x)|\, p(x)$

Optimal Proposal Distribution

- Optimal q, denoted by $q^* \propto |f(x)|\, p(x)$
- In fact, $q^* = \dfrac{|f(x)|\, p(x)}{\mathbb{E}_p\left(|f(x)|\right)}$

Optimal Proposal Distribution

- Optimal q, denoted by $q^* \propto |f(x)|\, p(x)$
- In fact, $q^* = \dfrac{|f(x)|\, p(x)}{\mathbb{E}_p\left(|f(x)|\right)}$

  Proof:

$$
\begin{aligned}
\mu^2 + {\sigma_{q^*}}^2 &= \int_{\mathcal{Q}} \frac{\left(f(x)p(x)\right)^2}{q^*(x)}\, dx \\
&= \int_{\mathcal{Q}} \frac{\left(f(x)p(x)\right)^2}{\frac{|f(x)|\, p(x)}{\mathbb{E}_p\left(|f(x)|\right)}}\, dx = \left(\mathbb{E}_p\left(|f(X)|\right)\right)^2 = \left(\mathbb{E}_q\left[\frac{|f(X)|\, p(X)}{q(X)}\right]\right)^2 \\
&= \left(\int_{\mathcal{Q}} \frac{|f(x)|\, p(x)}{q(x)} q(x)\, dx\right)^2 \\
&\overset{(*)}{\leq} \int_{\mathcal{Q}} \frac{f^2(x)p^2(x)}{q^2(x)} q(x)\, dx \underbrace{\int_{\mathcal{Q}} q(x)\, dx}_{=1} = \mu^2 + {\sigma_q}^2
\end{aligned}
$$

$$\implies {\sigma_{q^*}}^2 \leq {\sigma_q}^2$$

$(*)$ follows from Cauchy-Schwartz inequality

Quick Summary of IS

Quick Summary of IS

- IS is an unbiased estimate

Quick Summary of IS

- IS is an unbiased estimate
- By choosing “optimal” proposal distribution, we can reduce the variance of sample mean estimate

Quick Summary of IS

- IS is an unbiased estimate
- By choosing “optimal” proposal distribution, we can reduce the variance of sample mean estimate
- IS allows us to sample values from “light” tail of the distribution

Quick Summary of IS

Quick Summary of IS

However, there are some pointers to keep in mind

Quick Summary of IS

However, there are some pointers to keep in mind

- $q(x) > 0$ whenever $f(x)p(x) \neq 0$. In other words, $\text{supp}(q) \supset \text{supp}(p)$

Quick Summary of IS

However, there are some pointers to keep in mind

- $q(x) > 0$ whenever $f(x)p(x) \neq 0$. In other words, $\text{supp}(q) \supset \text{supp}(p)$
- The likelihood ratio $w(x) = \frac{p(x)}{q(x)}$ gives an idea of how "good" a proposal distribution is

Quick Summary of IS

However, there are some pointers to keep in mind

- $q(x) > 0$ whenever $f(x)p(x) \neq 0$. In other words, $\text{supp}(q) \supset \text{supp}(p)$
- The likelihood ratio $w(x) = \frac{p(x)}{q(x)}$ gives an idea of how "good" a proposal distribution is

$$\mathbb{E}_p\left[f^2(X)\,w(X)\right] = \mathbb{E}_q\left[f^2(X)\,w^2(X)\right]$$

Quick Summary of IS

However, there are some pointers to keep in mind

- $q(x) > 0$ whenever $f(x)p(x) \neq 0$. In other words, $\text{supp}(q) \supset \text{supp}(p)$
- The likelihood ratio $w(x) = \dfrac{p(x)}{q(x)}$ gives an idea of how "good" a proposal distribution is

$$\mathbb{E}_p \left[f^2(X)\, w(X) \right] = \mathbb{E}_q \left[f^2(X)\, w^2(X) \right]$$

  - The appearance of q in the denominator of w means that light-tailed q are dangerous

Quick Summary of IS

However, there are some pointers to keep in mind

- $q(x) > 0$ whenever $f(x)p(x) \neq 0$. In other words, $\text{supp}(q) \supset \text{supp}(p)$
- The likelihood ratio $w(x) = \dfrac{p(x)}{q(x)}$ gives an idea of how "good" a proposal distribution is

$$\mathbb{E}_p \left[f^2(X)\, w(X)\right] = \mathbb{E}_q \left[f^2(X)\, w^2(X)\right]$$

  - The appearance of q in the denominator of w means that light-tailed q are dangerous
  - so, q should have tails at least as heavy as p does

Simulation Study

Simulation

Simulation

- For comparing MC and IS, we consider the toy problem of calculating the tail of the normal distribution $\mathcal{N}(0, 1)$

Simulation

- For comparing MC and IS, we consider the toy problem of calculating the tail of the normal distribution $\mathcal{N}(0,1)$

$$\mathbb{P}(X > 3) = \frac{1}{\sqrt{2\pi}} \int_3^\infty \exp\left(-\frac{x^2}{2}\right)$$

- We consider two proposal distributions q_1 and q_2 with heavy tail and light tail over $(3,\infty)$ respectively

Simulation

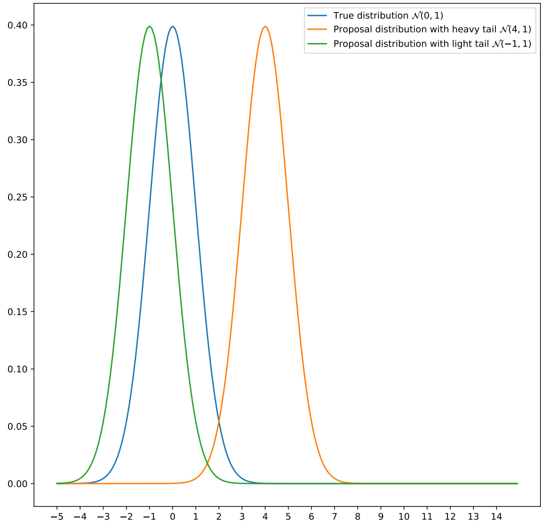
- For comparing MC and IS, we consider the toy problem of calculating the tail of the normal distribution $\mathcal{N}(0, 1)$

$$\mathbb{P}(X > 3) = \frac{1}{\sqrt{2\pi}} \int_3^\infty \exp\left(-\frac{x^2}{2}\right)$$

- We consider two proposal distributions q_1 and q_2 with heavy tail and light tail over $(3, \infty)$ respectively
    - This is to show how the choice of proposal distribution q helps or hurts our IS estimate of the tail probability

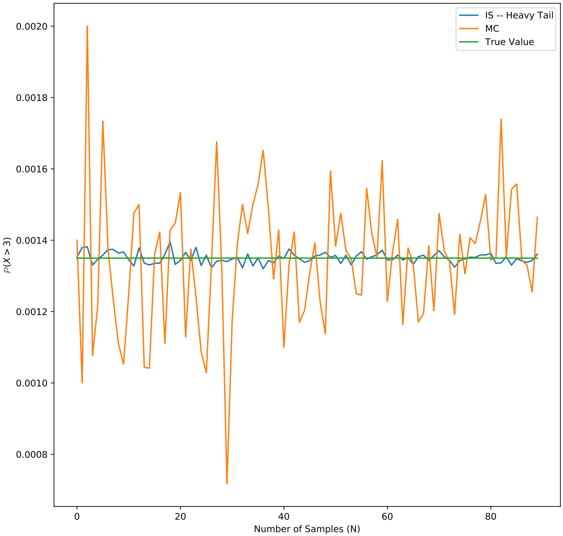
Proposal Distributions and True Distribution

Proposal Distributions and True Distribution



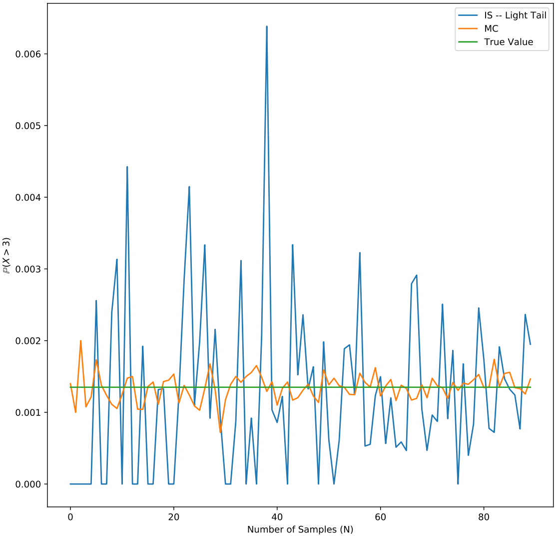
Results - Using Heavy Tail Proposal Distribution

Results - Using Heavy Tail Proposal Distribution



Results - Using Light Tail Proposal Distribution

Results - Using Light Tail Proposal Distribution



Conclusion

Conclusion

Conclusion

- We considered the problem of finding the expectation of a function of random variable

Conclusion

- We considered the problem of finding the expectation of a function of random variable
- We have seen that many quantities of interest may be cast as expectation problem

Conclusion

- We considered the problem of finding the expectation of a function of random variable
- We have seen that many quantities of interest may be cast as expectation problem
- Explicit computation of the expectation is often cumbersome, so we consider simple Monte-Carlo simulation

Conclusion

Conclusion

- Monte-Carlo simulation may lead to “poor” approximation i.e, huge variance, in some problems of interest like calculating the tail of a distribution

Conclusion

- Monte-Carlo simulation may lead to “poor” approximation i.e, huge variance, in some problems of interest like calculating the tail of a distribution
- This motivated us to look for an alternative sampling technique called *Importance Sampling* which allows us to sample values from the tail using the so called proposal distribution

Conclusion

- Monte-Carlo simulation may lead to “poor” approximation i.e, huge variance, in some problems of interest like calculating the tail of a distribution
- This motivated us to look for an alternative sampling technique called *Importance Sampling* which allows us to sample values from the tail using the so called proposal distribution
- We finally conducted a simple toy experiment to see the advantage of using IS. We have also demonstrated how could IS possibly fail

Thank You