Outline

Importance Sampling
- Unbiased importance sampling
- Normalised Importance Sampling
Outline

Importance Sampling
  Unbiased importance sampling
  Normalised Importance Sampling
Importance Sampling

- We want to estimate
  \[ \theta = \mathbb{E}(\phi(X)) \]

  where \( X \) is a rv with pdf or pmf \( p \) and \( \phi : \Omega \to \mathbb{R} \).

- The Monte Carlo estimator uses samples from \( p \) to estimate \( \theta \), but this choice is in general suboptimal.

- Importance sampling uses samples from another distribution \( q \), called importance or proposal distribution, and reweight them.

- Importance sampling (IS) can be thought, among other things, as a strategy for recycling samples.

- It is also useful when we need to make an accurate estimate of the probability that a random variable exceeds some very high threshold.

- In this context it is referred to as a variance reduction technique.
Importance sampling identity

Let $Y \sim q$ and $X \sim p$ be continuous or discrete rv on $\Omega$. Assume $p(x) > 0 \Rightarrow q(x) > 0$, then for any function $\phi : \Omega \rightarrow \mathbb{R}$ we have

$$\mathbb{E}_p(\phi(X)) = \mathbb{E}_q(\phi(Y)w(Y))$$

where $w : \Omega \rightarrow \mathbb{R}^+$ is the importance weight function

$$w(x) = \frac{p(x)}{q(x)}.$$
Importance Sampling Identity

Proof: We have

\[ \mathbb{E}_p(\phi(X)) = \int_\Omega \phi(x)p(x)dx \]

\[ = \int_\Omega \phi(x) \frac{p(x)}{q(x)} q(x)dx \]

\[ = \int_\Omega \phi(x) w(x) q(x)dx \]

\[ = \mathbb{E}_q(\phi(Y)w(Y)). \]

Similar proof holds in the discrete case.
Importance Sampling Estimator

**Definition**

Let \( q \) and \( p \) be pdfs or pmfs on \( \Omega \). Assume \( p(x)\phi(x) \neq 0 \Rightarrow q(x) > 0 \). Let \( \phi : \Omega \to \mathbb{R} \) and \( X \sim p \) such that \( \theta = \mathbb{E}_p(\phi(X)) \) exists. Let \( Y_1, ..., Y_n \) be a sample of independent random variables distributed according to \( q \). The importance sampling estimator is defined as

\[
\hat{\theta}_n^{\text{IS}} = \frac{1}{n} \sum_{i=1}^{n} \phi(Y_i)w(Y_i).
\]

**Properties**

The IS estimator is

- **Unbiased**: \( \mathbb{E}[\hat{\theta}_n^{\text{IS}}] = \theta \)

- (Weakly and strongly) **consistent**: \( \hat{\theta}_n^{\text{IS}} \to \theta \) a.s. as \( n \to \infty \).
Proof.

\[ E[\hat{\theta}_n^{IS}] = \frac{1}{n} \sum_{i=1}^{n} E(\phi(Y_i)w(Y_i)) \]

\[ = E(\phi(Y_1)w(Y_1)) \]

\[ = E(\phi(X)) = \theta \]

Let \( Z_i = \phi(Y_i)w(Y_i) \). \( Z_1, \ldots, Z_n \) are iid with mean \( E(Z_i) = E(\phi(Y_i)w(Y_i)) = \theta \). From the strong law of large numbers

\[ \frac{1}{n} \sum_{i=1}^{n} Z_i \rightarrow \theta \quad \text{a.s. as } n \rightarrow \infty \]
Target and Proposal Distributions

- **Target:** \( p(x) = \frac{1}{2} e^{-|x|} \)
- **Proposal:** \( q(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \)
- **Weight function:** \( w(x) = \sqrt{\frac{\pi}{2}} e^{-|x| + \frac{x^2}{2}} \)
Target and Proposal Distributions

- **Target:** \( p(x) = \frac{1}{2} e^{-|x|} \)
- **Proposal:** \( q(x) = \frac{1}{\pi (1 + x^2)} \)
- **Weight function:** \( w(x) = \frac{\pi}{2} (1 + x^2) e^{-|x|} \)
Example: Gamma Distribution

Say we have simulated $Y_i \sim \text{Gamma}(a, b)$ and we want to estimate $\mathbb{E}_p(\phi(X))$ where $X \sim \text{Gamma}(\alpha, \beta)$.

Recall that the $\text{Gamma}(\alpha, \beta)$ density is

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x)$$

so

$$w(x) = \frac{p(x)}{q(x)} = \frac{\Gamma(a) \beta^\alpha}{\Gamma(\alpha) b^a} x^{\alpha-a} e^{-(\beta-b)x}$$

Hence

$$\hat{\theta}_{IS} = \frac{\Gamma(a) \beta^\alpha}{\Gamma(\alpha) b^a} \frac{1}{n} \sum_{i=1}^{n} \phi(Y_i) Y_i^{\alpha-a} e^{-(\beta-b)Y_i}$$

is an unbiased and consistent estimate of $\mathbb{E}_p(\phi(X))$. 

**Proposition.** Assume \( \theta = \mathbb{E}_p(\phi(X)) \) and \( \mathbb{E}_p(w(X)\phi^2(X)) \) are finite. Then \( \hat{\theta}_n^{IS} \) satisfies

\[
\mathbb{E} \left( \left( \hat{\theta}_n^{IS} - \theta \right)^2 \right) = \mathbb{V} (\hat{\theta}_n^{IS}) = \frac{1}{n} \mathbb{V}_q (w(Y_1)\phi(Y_1)) \\
= \frac{1}{n} \left( \mathbb{E}_q \left( \frac{p^2(Y_1)}{q^2(Y_1)} \phi^2(Y_1) \right) - \mathbb{E}_q \left( \frac{p(Y_1)}{q(Y_1)} \phi(Y_1) \right)^2 \right) \\
= \frac{1}{n} \left( \mathbb{E}_p \left( w(X)\phi^2(X) \right) - \theta^2 \right).
\]

**Each time we do IS we should check that this variance is finite, otherwise our estimates are somewhat untrustworthy!** We check \( \mathbb{E}_p(w(X)\phi^2(X)) \) is finite.
Variance of the Importance Sampling Estimator

- **Target:** \( p(x) = \frac{1}{2} e^{-|x|} \)
- **Proposal:** \( q(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \)
- \( w(x) = \frac{\sqrt{\pi}}{2} e^{-|x|+x^2/2}, \phi(x) = x \)
- \( \mathbb{E}_p(w(X)\phi^2(X)) = \infty \)
Variance of the Importance Sampling Estimator

- **Target:** $p(x) = \frac{1}{2} e^{-|x|}$
- **Proposal:** $q(x) = \frac{1}{\pi(1 + x^2)}$
- $w(x) = \pi/2 (1 + x^2) e^{-|x|}$, $\phi(x) = x$
- $\mathbb{E}_p(w(X)\phi^2(X)) < \infty$
Variance of the Importance Sampling Estimator

- If $\nabla_p(\phi(X))$ is finite, a sufficient condition is that $w$ is a bounded function: there is $M$ such that $w(x) = \frac{p(x)}{q(x)} \leq M$ for all $x \in \Omega$.
- Note that this is the same condition as for rejection sampling,
- For IS it is enough just for $M$ to exist—we do not have to work out its value.
- Proof:

\[
\mathbb{E}_p(w(X)\phi^2(X)) \leq M\mathbb{E}_p(\phi^2(X)) < \infty
\]

as $\nabla_p(\phi(X)) < \infty$. 
Example: Gamma Distribution

Let us check that the variance of $\hat{\theta}_n^{\text{IS}}$ in previous Example is finite if $\theta = \mathbb{E}_p(\phi(X))$ and $\nabla_p(\phi(X))$ are finite.

It is enough to check that $\mathbb{E}_p\left( w(Y_1)\phi^2(Y_1) \right)$ is finite.

The normalisation constants are finite so we can ignore those, and begin with

$$w(x)\phi^2(x) \propto x^{\alpha-a}e^{-(\beta-b)X} \phi^2(x).$$

The expectation of interest is

$$\mathbb{E}_p\left( w(X)\phi^2(X) \right) \propto \mathbb{E}_p\left( X^{\alpha-a}e^{-(\beta-b)X} \phi^2(X) \right)$$

$$= \int_0^\infty p(x) x^{\alpha-a} \exp(-(\beta-b)x))\phi^2(x) \, dx$$

$$\leq M \int_0^\infty p(x)\phi(x)^2 \, dx = M\mathbb{E}_p(\phi^2(X)).$$

where $M = \max_{x>0} x^{\alpha-a} \exp(-(\beta-b)x)$ is finite if $a < \alpha$ and $b < \beta$ (see rejection sampling section).
Since $\theta = \mathbb{E}_p(\phi(X))$ and $\nabla_p(\phi(X))$ are finite, we have
$\mathbb{E}_p(\phi^2(X)) < \infty$ if these conditions on $a, b$ are satisfied. If not, we cannot conclude as it depends on $\phi$.

These same (sufficient) conditions apply to our rejection sampler for $\text{Gamma}(\alpha, \beta)$.
Choice of the Importance Sampling Distribution

- While $p$ is given, $q$ needs to cover $p\phi$ (i.e. $p(x)\phi(x) \neq 0 \Rightarrow q(x) > 0$) and be simple to sample.
- The requirement $\mathbb{V}\left(\hat{\theta}_n^{\text{IS}}\right) < \infty$ further constrains our choice: we need $\mathbb{E}_p \left( w(X)\phi^2(X) \right) < \infty$.
- If $\mathbb{V}_p(\phi(X))$ is known finite then, it may be easy to get a sufficient condition for $\mathbb{E}_p \left( w(X)\phi^2(X) \right) < \infty$; e.g. $w(x) \leq M$. Further analysis will depend on $\phi$. 
Choice of the Importance Sampling Distribution

- What is the choice $q_{\text{opt}}$ of $q$ that actually minimizes the variance of the IS estimator? Consider for now $\phi : \Omega \rightarrow [0, \infty)$ then

$$q_{\text{opt}}(x) = \frac{p(x)\phi(x)}{\mathbb{E}_p(\phi(X))} \Rightarrow \mathbb{V}(\hat{\theta}_n^{\text{IS}}) = 0.$$

- This optimal zero-variance estimator cannot be implemented as

$$w(x) = \frac{p(x)}{q_{\text{opt}}(x)} = \frac{\mathbb{E}_p(\phi(X))}{\phi(x)}$$

where $\mathbb{E}_p(\phi(X))$ is the quantity we are trying to estimate! This can however be used as a guideline to select $q$. 
Choice of the Importance Sampling Distribution

- For general function $\phi : \Omega \rightarrow \mathbb{R}$, the optimal importance distribution is

$$q_{\text{opt}}(x) = \frac{p(x)|\phi(x)|}{\mathbb{E}_p(|\phi(X)|)}$$

with variance

$$\mathbb{V}(\hat{\theta}_n^{\text{IS}}) = \frac{1}{n} \left( \mathbb{E}_p(|\phi(X)|)^2 - \theta^2 \right).$$
Choice of the Importance Sampling Distribution

- Proof:

\[
\mathbb{E}_p \left( w(X) \phi^2(X) \right) = \mathbb{E}_q \left( \frac{p^2(Y_1)}{q^2(Y_1)} \phi^2(Y_1) \right) \\
= \mathbb{V}_q \left( \frac{p(Y_1)}{q(Y_1)} |\phi(Y_1)| \right) + \left( \mathbb{E}_q \left( \frac{p(Y_1)}{q(Y_1)} |\phi(Y_1)| \right) \right)^2 \\
\geq \left( \mathbb{E}_q \left( \frac{p(Y_1)}{q(Y_1)} |\phi(Y_1)| \right) \right)^2 \\
= (\mathbb{E}_p (|\phi(X)|))^2
\]

where the lower bound does not depend on \( q \). This lower bound is achieved for \( q = q_{opt} \)

\[
\mathbb{E}_p \left( \frac{p(X)}{q_{opt}(X)} \phi^2(X) \right) = (\mathbb{E}_p (|\phi(X)|))^2
\]
Importance Sampling for Rare Event Estimation

- One important class of applications of IS is to problems in which we estimate the probability for a rare event.
- In such scenarios, we may be able to sample from \( p \) directly but this does not help us. If, for example, \( X \sim p \) with \( \mathbb{P}(X > x_0) = \mathbb{E}_p(\mathbb{I}[X > x_0]) = \theta \) say, with \( \theta \ll 1 \), we may not get any samples \( X_i > x_0 \) and our estimate \( \hat{\theta}_n = \sum_i \mathbb{I}(X_i > x_0)/n \) is simply zero.
- Generally, we have
  \[
  \mathbb{E}\left(\hat{\theta}_n\right) = \theta, \quad \mathbb{V}\left(\hat{\theta}_n\right) = \frac{\theta(1 - \theta)}{n}
  \]
  but the relative variance
  \[
  \frac{\mathbb{V}\left(\hat{\theta}_n\right)}{\theta^2} = \frac{(1 - \theta)}{\theta n} \xrightarrow{\theta \to 0} \infty.
  \]
- By using IS, we can actually reduce the variance of our estimator.
Importance Sampling for Rare Event Estimation

- Let $X \sim \mathcal{N}(\mu, \sigma^2)$ be a scalar normal random variable and we want to estimate $\theta = \mathbb{P}(X > x_0)$ for some $x_0 \gg \mu + 3\sigma$.
- If $p$ is the pdf of $X$ then

  $$q(x) = \frac{p(x)e^{tx}}{\mathbb{E}_p(e^{tX})}$$

is called an exponentially tilted version of $p$ where $\mathbb{E}_p(e^{tX})$ is the moment generating function of $X$.
- For many standard pdfs, the exponentially tilted pdf is in the same family as $p$, with different parameters.
- For $p$ the pdf of a Gaussian variable with mean $\mu$ and variance $\sigma^2$,

  $$q(x) \propto e^{-(x-\mu)^2/2\sigma^2} e^{tx} = e^{-(x-\mu-t\sigma^2)^2/2\sigma^2} e^{\mu t + t^2\sigma^2/2}$$

so we have

  $$q(x) = \mathcal{N}(x; \mu + t\sigma^2, \sigma^2), \quad \mathbb{E}_p(e^{tX}) = e^{\mu t + t^2\sigma^2/2}.$$
Importance Sampling for Rare Event Estimation

- The IS weight function is \( p(x)/q(x) = e^{-tx} M_p(t) \) so
  \[
  w(x) = e^{-t(x-\mu-t\sigma^2/2)}.
  \]

- We take samples \( Y_i \sim \mathcal{N}(\mu + t\sigma^2, \sigma^2) \), and form our IS estimator for
  \( \theta = \mathbb{P}(X > x_0) \)
  \[
  \hat{\theta}_{n}^{\text{IS}} = \frac{1}{n} \sum_{i=1}^{n} w(Y_i) \mathbb{I}(Y_i > x_0)
  \]
  since \( \phi(Y_i) = \mathbb{I}(Y_i > x_0) \).

- We have not said how to choose \( t \). The point here is that we want samples in the region of interest. We choose the mean of the tilted distribution so that it equals \( x_0 \), this ensure we have samples in the region of interest; that is \( \mu + t\sigma^2 = x_0 \), or \( t = (x_0 - \mu)/\sigma^2 \).
Original and Exponentially Tilted Densities

$p(x) = N(x; 0, 1)$ and $q(x) = N(x; t, 1)$, $x_0 = t = 4$
Optimal Tilted Densities

- We selected $t$ such that $\mu + t\sigma^2 = x_0$ somewhat heuristically.

- In practice, we might be interested in selecting the $t$ value which minimizes the variance of $\hat{\theta}_{IS}$ where

\[
\nabla(\hat{\theta}_{IS}) = \frac{1}{n} \left( \mathbb{E}_p (w(X)I(X > x_0)) - \mathbb{E}_p (I(X > x_0))^2 \right) = \frac{1}{n} \left( \mathbb{E}_p (w(X)I(X > x_0)) - \theta^2 \right).
\]

- Hence we need to minimize $\mathbb{E}_p (w(X)I(X > x_0))$ w.r.t $t$ where

\[
\mathbb{E}_p (w(X)I(X > x_0)) = \int_{x_0}^{\infty} p(x)e^{-t(x-\mu-t\sigma^2/2)}dx = M_p(t)\int_{x_0}^{\infty} p(x)e^{-tx}dx
\]
Optimal Tilted Densities

- Relative variance $\frac{\mathbb{V} (\hat{\theta}_1^{IS})}{\theta^2}$ of the IS estimators for different values of $t$
Importance Sampling in High Dimension

- Purely for illustration, consider that we want to estimate

$$\theta = \mathbb{E}_p(1) = 1$$

where the target pdf is a \(d\)-dimensional Gaussian

$$p(x_1, \ldots, x_d) = (2\pi)^{-d/2} \exp \left( -\frac{1}{2} \sum_{k=1}^{d} x_k^2 \right).$$

- Consider the proposal density

$$q(x_1, \ldots, x_d) = (2\pi\sigma^2)^{-d/2} \exp \left( -\frac{1}{2\sigma^2} \sum_{k=1}^{d} x_k^2 \right).$$

- We have

$$w(x) = \frac{p(x_1, \ldots, x_d)}{q(x_1, \ldots, x_d)} = \sigma^d \exp \left( -\frac{1}{2} \left( 1 - \sigma^{-2} \right) \sum_{k=1}^{d} x_k^2 \right).$$
For $Y_i \sim q$, $\hat{\theta}_n^{\text{IS}} = \frac{1}{n} \sum_{i=1}^{n} w(Y_i)$ is a consistent estimate of $\theta = 1$.

The estimator has finite variance for $\sigma^2 > \frac{1}{2}$, with

$$\mathbb{V} \left( \hat{\theta}_n^{\text{IS}} \right) = \frac{\mathbb{V}_q (w(Y_1))}{n} = \frac{1}{n} \left( \left( \frac{\sigma^4}{2\sigma^2 - 1} \right)^{d/2} - 1 \right)$$

with $\frac{\sigma^4}{2\sigma^2 - 1} > 1$ for $\sigma^2 > \frac{1}{2}$, $\sigma^2 \neq 1$.

Variance of the IS estimator grows exponentially with the dimension $d$. 
Importance Sampling
  Unbiased importance sampling
  Normalised Importance Sampling
Normalised Importance Sampling

- In most practical scenarios,
  
  \[ p(x) = \tilde{p}(x)/Z_p \text{ and } q(x) = \tilde{q}(x)/Z_q \]

  where \( \tilde{p}(x), \tilde{q}(x) \) are known but \( Z_p = \int_{\Omega} \tilde{p}(x)dx, \) \( Z_q = \int_{\Omega} \tilde{q}(x)dx \) are unknown or difficult to compute.

- The previous IS estimator is not applicable as it requires evaluating
  \[ w(x) = p(x)/q(x). \]

- An alternative IS estimator can be proposed based on the following alternative IS identity.

- **Proposition.** Let \( Y \sim q \) and \( X \sim p \) be continuous or discrete rv on \( \Omega. \) Assume \( p(x) > 0 \Rightarrow q(x) > 0, \) then for any function \( \phi : \Omega \to \mathbb{R} \) we have
  
  \[ \mathbb{E}_p(\phi(X)) = \frac{\mathbb{E}_q(\phi(Y)\tilde{w}(Y))}{\mathbb{E}_q(\tilde{w}(Y))} \]

  where \( \tilde{w} : \Omega \to \mathbb{R}^+ \) is the importance weight function
  \[ \tilde{w}(x) = \tilde{p}(x)/\tilde{q}(x). \]
Normalised Importance Sampling

Proof: Observe that

\[
\mathbb{E}_q(\tilde{w}(Y)) = \int \frac{\tilde{p}(x)}{\tilde{q}(x)} q(x) dx
\]

\[
= \int \frac{p(x)}{q(x)} \frac{Z_q}{Z_p} q(x) dx
\]

\[
= \frac{Z_q}{Z_p}
\]

and noting that \(\tilde{w}/\frac{Z_q}{Z_p} = w\) we have that

\[
\frac{\mathbb{E}_q(\phi(Y)\tilde{w}(Y))}{\mathbb{E}_q(\tilde{w}(Y))} = \mathbb{E}_q(\phi(Y)w(Y))
\]

Remark: Even if we are interested in a simple function \(\phi\), we do need \(p(x) > 0 \Rightarrow q(x) > 0\) to hold instead of \(p(x)\phi(x) \neq 0 \Rightarrow q(x) > 0\) for the previous IS identity.
Normalised Importance Sampling

Proof: We have

\[ \mathbb{E}_p(\phi(X)) = \int_{\Omega} \phi(x)p(x)dx \]

\[ = \frac{\int_{\Omega} \phi(x)p(x)\frac{p(x)}{q(x)}q(x)dx}{\int_{\Omega} \frac{p(x)}{q(x)}q(x)dx} \]

\[ = \frac{\int_{\Omega} \phi(x)\tilde{w}(x)q(x)dx}{\int_{\Omega} \tilde{w}(x)q(x)dx} \]

\[ = \frac{\mathbb{E}_q(\phi(Y)\tilde{w}(Y))}{\mathbb{E}_q(\tilde{w}(Y))}. \]

Remark: Even if we are interested in a simple function \( \phi \), we do need \( p(x) > 0 \Rightarrow q(x) > 0 \) to hold instead of \( p(x)\phi(x) \neq 0 \Rightarrow q(x) > 0 \) for the previous IS identity.
Normalised Importance Sampling Pseudocode

1. Inputs:
   - Function to draw samples from \( q \)
   - Function \( \tilde{w}(x) = \tilde{p}(x)/\tilde{q}(x) \)
   - Function \( \phi \)
   - Number of samples \( n \)

2. For \( i = 1, \ldots, n \):
   2.1 Draw \( y_i \sim q \).
   2.2 Compute \( \tilde{w}_i = \tilde{w}(y_i) \).

3. Return
   \[
   \frac{\sum_{i=1}^{n} \tilde{w}_i \phi(y_i)}{\sum_{i=1}^{n} \tilde{w}_i}.
   \]
Normalised Importance Sampling Estimator

Proposition

Let $q$ and $p$ be pdf or pmf on $\Omega$, with $q(x) \propto \tilde{q}(x)$ and $p(x) \propto \tilde{p}(x)$. Assume $p(x) > 0 \Rightarrow q(x) > 0$. Let $X \sim p$, and $\phi : \Omega \rightarrow \mathbb{R}$ such that $\theta = \mathbb{E}_p(\phi(X))$ exists. Let $Y_1, \ldots, Y_n$ be a sample of independent random variables distributed according to $q$ then the normalized importance sampling estimator, defined by

$$
\hat{\theta}_{\text{NIS}}_n = \frac{1}{n} \sum_{i=1}^n \frac{\phi(Y_i)\tilde{w}(Y_i)}{\frac{1}{n} \sum_{i=1}^n \tilde{w}(Y_i)} = \frac{\sum_{i=1}^n \phi(Y_i)\tilde{w}(Y_i)}{\sum_{i=1}^n \tilde{w}(Y_i)},
$$

with $\tilde{w}(x) = \frac{\tilde{p}(x)}{q(x)}$.

This estimator is consistent.

Remark: It is easy to show that $\hat{A}_n = \frac{1}{n} \sum_{i=1}^n \phi(Y_i)\tilde{w}(Y_i)$ (resp. $\hat{B}_n = \frac{1}{n} \sum_{i=1}^n \tilde{w}(Y_i)$) is an unbiased and consistent estimator of $A = \mathbb{E}_q(\phi(Y)\tilde{w}(Y))$ (resp. $B = \mathbb{E}_q(\tilde{w}(Y))$). However $\hat{\theta}_{\text{NIS}}_n$, which is a ratio of estimates, is biased for finite $n$. 

Part A Simulation. HT 2018. J. Berestycki. 35 / 37
Proof strong consistency (not examinable). The strong law of large numbers yields

\[ \mathbb{P} \left( \lim_{n \to \infty} \hat{A}_n \to A \right) = \mathbb{P} \left( \lim_{n \to \infty} \hat{B}_n \to B \right) = 1 \]

This implies

\[ \mathbb{P} \left( \lim_{n \to \infty} \hat{A}_n \to A, \lim_{n \to \infty} \hat{B}_n \to B \right) = 1 \]

and

\[ \mathbb{P} \left( \lim_{n \to \infty} \frac{\hat{A}_n}{\hat{B}_n} \to \frac{A}{B} \right) = 1. \]
Example Revisited: Gamma Distribution

- We are interested in estimating $\mathbb{E}_p(\phi(X))$ where $X \sim \text{Gamma}(\alpha, \beta)$ using samples from a $\text{Gamma}(a, b)$ distribution; i.e.

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad q(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$$

- Suppose we do not remember the expression of the normalising constant for the Gamma, so that we use

$$\tilde{p}(x) = x^{\alpha-1} e^{-\beta x}, \quad \tilde{q}(x) = x^{a-1} e^{-bx}$$

$$\Rightarrow \tilde{w}(x) = x^{\alpha-a} e^{-(\beta-b)x}$$

- Practically, we simulate $Y_i \sim \text{Gamma}(a, b)$, for $i = 1, 2, \ldots, n$ then compute

$$\tilde{w}(Y_i) = Y_i^{\alpha-a} e^{-(\beta-b)Y_i},$$

$$\hat{\theta}_n^{\text{NIS}} = \frac{\sum_{i=1}^{n} \phi(Y_i) \tilde{w}(Y_i)}{\sum_{i=1}^{n} \tilde{w}(Y_i)}.$$