$(b-a)\int_a^b g(x)\frac{1}{b-a}\,dx$

Importance Sampling

① apply to unbounded intervals

Chapter 4 (2)

② efficient to draw samples

STATS 102C: Introduction to Monte Carlo Methods

$\theta = \int_D g(x)\,dx$

$= \int_D \frac{g(x)}{f(x)} f(x)\,dx$

$= E_f\left[\frac{g(x)}{f(x)}\right]$

① $X_1,\dots,X_m \sim f(x)$

② $\frac{1}{m}\sum_{i=1}^{m} g(X_i)/f(X_i)$

$X_1,\dots,X_m \sim \phi(x)$

importantnle sampling function

$X \sim exp(\lambda)$

$E[\sqrt{X}]$

$Y=\sqrt{X} \sim$ Rayleigh (λ)

$E[Y] = \frac{\sqrt{\pi}}{2\sqrt{\lambda}}$

Introduction

- Suppose $X \sim f(x)$, for $x \in D$, where D is the support of X:
  - $f(x) > 0$, for $x \in D$
  - $f(x) = 0$, for $x \notin D$
  - $\int_D f(x)dx = 1$
- Suppose we can sample from $f(x)$. We can use the simple Monte Carlo method to estimate $\int_D g(x)dx$.
- Suppose we want to compute $\theta = E_f[g(X)] = \boxed{\int_D g(x)f(x)dx}$, but we are unable to sample from $f(x)$ directly. How do we estimate θ?
- Suppose $\phi(x) > 0$ on D, then θ can be written

$$\theta = \int_D g(x)\frac{f(x)}{\phi(x)}\phi(x)dx = E_\phi\left[\frac{g(x)f(x)}{\phi(x)}\right]$$

$$\theta = E_f[g(x)] = \int_D g(x) \cdot \underline{f(x)}\, dx$$

$$= \int_D g(x) \cdot f(x) \frac{\phi(x)}{\phi(x)}\, dx$$

$$= \int_D \frac{g(x)f(x)}{\phi(x)} \cdot \phi(x)\, dx$$

$$= E_\phi\left[\frac{g(x)f(x)}{\phi(x)}\right]$$

① $x_1, \dots, x_m \sim \phi(x)$

② $\sum_{i=1}^{m} \frac{g(x_i) \cdot f(x_i)}{\phi(x_i)} \frac{1}{m} = \underline{\hat{\theta}}$

$$MSE(\hat{\theta}) = \underline{\underline{Var(\hat{\theta})}} + \boxed{\underline{[E[\hat{\theta}] - \theta]}^2} = 0$$

$$Var_f(\hat{\theta}) = Var_\phi\left[\frac{g(x)f(x)}{\phi(x)}\right]\frac{1}{m}$$

$$= \left\{ \underline{E_\phi\left[\left[\frac{g(x)f(x)}{\phi(x)}\right]^2\right]} - E_\phi\left[\frac{g(x)f(x)}{\phi(x)}\right]^2 \right\}\frac{1}{m}$$

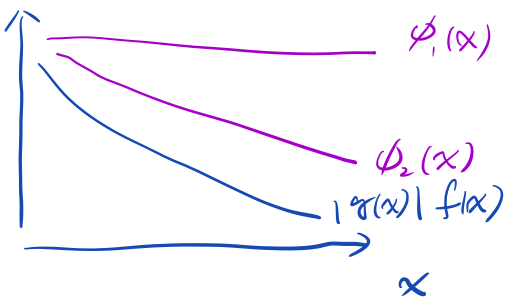
$$= \left\{ \int_D \frac{g(x)^2 f(x)^2}{\phi(x)^2} \phi(x)\, dx - \theta^2 \right\}\frac{1}{m}$$

Find the optimal ϕ that leads $\int_D \frac{g(x)^2 f(x)^2}{\phi(x)^2} \phi(x)\, dx = \theta^2$

$\Rightarrow$ $Var(\hat{\theta}) = 0$

$$\phi(x) = \frac{|g(x)|\, f(x)}{\boxed{\int_D g(x)\, f(x)\, dx}} = \frac{|g(x)|\, f(x)}{\text{normalizing constant}} \underset{\text{prop. to}}{\propto} |g(x)|\, f(x)$$

the shap of $\phi(x)$ is "close to" $|g(x)|\, f(x)$



① find $\phi(x) \cong |g(x)|\, f(x)$

② $x_1, \dots, x_m \sim \phi(x)$

③ Capute $\frac{1}{m} \sum_{i=1}^{m} \frac{g(x_i)\cdot f(x_i)}{\phi(x_i)}$
$= \hat{\theta}$

Let $w(x_i) = \frac{f(x_i)}{\phi(x_i)}$

$$\hat{\theta} = \frac{1}{m} \sum_{i=1}^{m} g(x_i) \cdot \underbrace{\boxed{w(x_i)}}_{\text{imortance weight}}$$

$$\int_0^1 \underbrace{\boxed{\frac{e^{-x}}{1+x^2}}}_{g(x)} (1)\, dx = \int_0^1 \frac{g(x)}{f(x)} \cdot f(x)\, dx = E_f\left[\frac{g(x)}{f(x)}\right]$$

$$E_U\left[\frac{e^{-x}}{1+x^2}\right] = \int_0^1 \frac{e^{-x}}{1+x^2} \cdot (1)\, dx$$

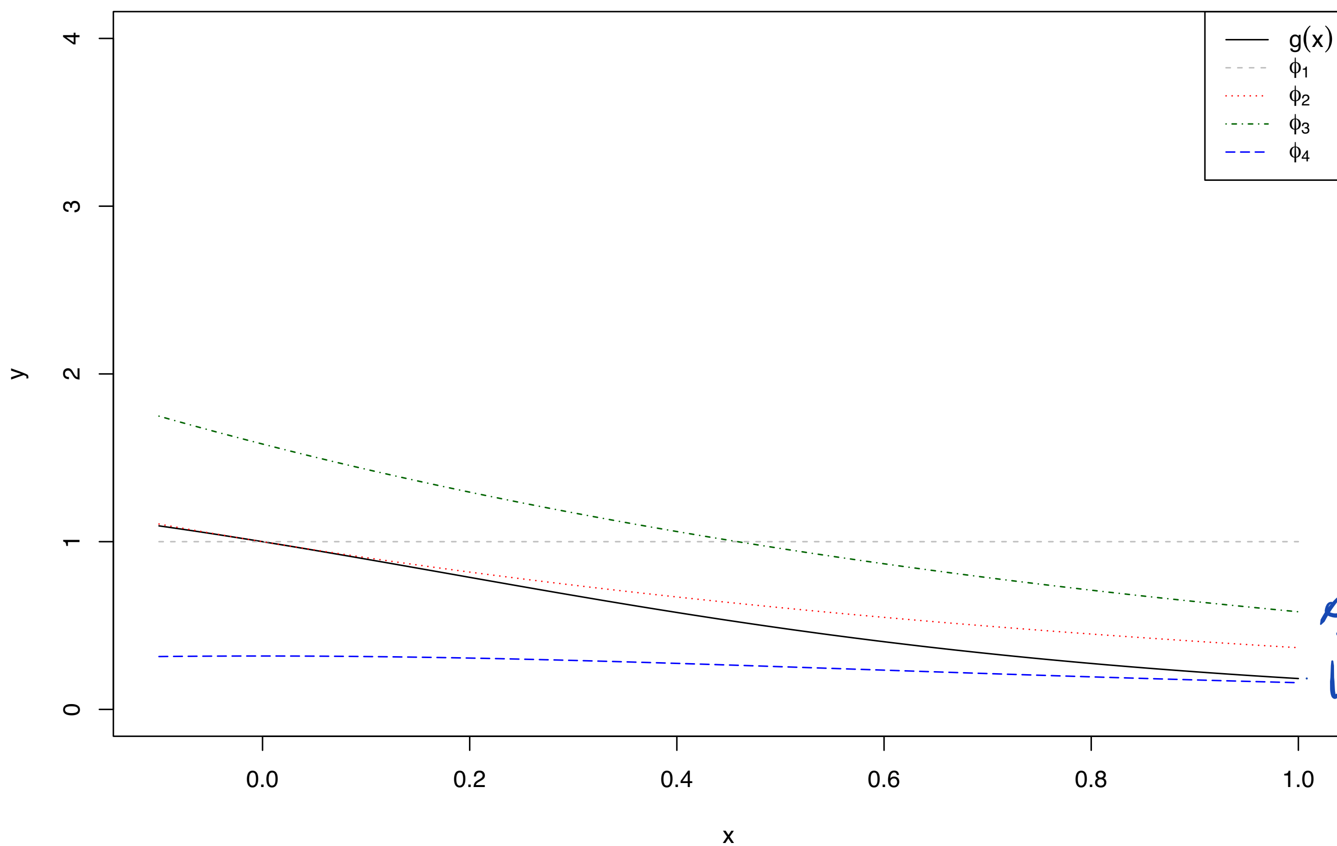
$$= \int_0^1 \frac{\frac{e^{-x}}{1+x^2}}{\phi(x)} \cdot \phi(x)\, dx = \int_0^1 \boxed{\frac{f(x)\cdot g(x)}{\phi(x)}} \phi(x)\, dx$$

$$= E_\phi\left[\frac{g(x)\, f(x)}{\phi(x)}\right]$$

Example: Estimate $\int_0^1 \frac{e^{-x}}{1+x^2}dx = \theta$

The four possible choices of importance functions:

1. $\phi_1(x) = 1,\ 0 < x < 1$
2. $\phi_2(x) = e^{-x},\ 0 < x < \infty$
3. $\phi_3(x) = e^{-x}/(1 - e^{-1}),\ 0 < x < 1$
4. $\phi_4(x) = 1/(\pi * (1 + x^2)),\ -\infty < x < \infty$



$\phi_3(x)$

$|g(x)| \cdot f(x)$

R Code

```r
g <- function(x){
 exp(-x - log(1 + x^2)) * (x > 0) * (x < 1)
}
m <- 10000
#Use Uniform(0, 1) as the candidate function
is1 <- replicate(1000, expr = {
    x <- runif(m)
    phi <- 1
    mean(g(x) / phi)
})
#Use Exponential(1) as the candidate function
is2 <- replicate(1000, expr = {
    u <- runif(m)
    x <- -log(u)
    x <- x[x <= 1]
    phi <- exp(-x)
    sum(g(x) / phi) / m
})
```

```r
#Use exp(-x) / (1 - exp(-1)) as the candidate function
is3 <- replicate(1000, expr = {
    u <- runif(m)
    x <- -log(1 - u * (1 - exp(-1)))
    phi <- exp(-x) / (1 - exp(-1))
    mean(g(x) / phi)
 })
#Use Standard Cauchy as the candidate function
is4 <- replicate(1000, expr = {
    x <- rcauchy(m)
    x <- x[x >= 0 & x <=1]
    phi <- dcauchy(x)
    outcome <- g(x) / phi
    sum(outcome) / m
 })
c(mean(is1), mean(is2), mean(is3), mean(is4))
```

```
## [1] 0.5246914 0.5245130 0.5247545 0.5248252
```

```r
c(var(is1), var(is2), var(is3), var(is4))
```

```
## [1] 6.011627e-06 1.681950e-05 9.240642e-07 9.484520e-05
```

Example: Folded Normal Distribution

- Suppose we want to estimate $E_f(X)$, where $f(x)$ is the PDF of the folded normal distribution,

$$f(x) = 2 \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = \underbrace{\boxed{\sqrt{\frac{2}{\pi}}}}_{z_g} \underbrace{\boxed{e^{-x^2/2}}}_{g(x)}, \quad \text{for } x \geq 0.$$

- The support of $f(x)$ is $D = [0, \infty)$:
  - $f(x) > 0$ for all $x \in D$
  - $\int_0^\infty f(x)dx = 1$

$$\underbrace{\frac{1}{\sqrt{2\pi}}}_{z_g} \underbrace{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}_{g(x)}$$

- We want to use importance sampling to estimate $E_f(X)$.

① choose $\phi(x) = 2e^{-2x}$, $x \geq 0$ $\Rightarrow \exp(2)$

② $w(x) = \frac{f(x)}{\phi(x)} = \sqrt{\frac{2}{\pi}} e^{-x^2/2} \Big/ 2e^{-2x} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}+2x}$

③ $x_1, \dots, x_m \sim \exp(2)$

$x = -\log(u)/2$

④ $\frac{\sum_{i=1}^{m} x_i \cdot w(x_i)}{m} = \hat{\theta}$

R Code to estimate $E_f(X)$ for the folded normal distribution

```
set.seed(9999)
n <- 10000 # Specify the number of points to generate

# Generate n points from Exp(lambda = 2)
x <- rexp(n, rate = 2)

# Compute importance weights
w <- exp(-x^2 / 2 + 2 * x) / sqrt(2 * pi)

# Compute mean(w(x) * g(x)) (g(x) = x here)
mean(w * x)
```

```
## [1] 0.801565
```

$$\hat{\theta} = \frac{1}{m} \sum g(x) \cdot w(x) \longrightarrow E_f[g(x)]$$

If $\phi(x) \approx f(x)$, $w(x) \approx 1$ $\qquad \sum g(x) \cdot \overset{\approx 1}{\boxed{w(x)}}$

If $\phi(x) >> f(x)$, $w(x) << 1$ $\Rightarrow \sum \underline{g(x) \cdot w(x)}$

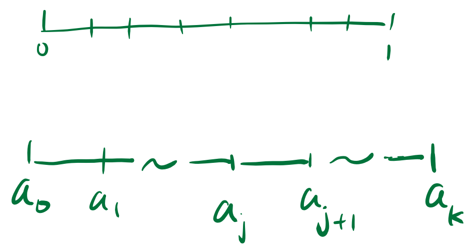
 $\phi(x) << f(x)$, $w(x) >> 1$

$$\phi(x) \approx f(x) \Rightarrow w(x) \approx 1$$

Define Efficiency $= \dfrac{1}{\underline{Var_{\phi}(w(x))}}$ $\qquad E_{\phi}(w(x)) = \int_D w(x) \cdot \phi(x) \, dx$

Stratified Importance Sampling

$$E[g(x)] = \int_D g(x)\,dx = \int_D \frac{g(x)}{f(x)}\, f(x)\,dx$$



$$Var_f\left[\frac{g(x)}{f(x)}\right] = \sigma^2$$

$$\hat{\theta}_I = \frac{1}{M}\sum_{i=1}^{M}\frac{g(x_i)}{f(x_i)}$$

$$\boxed{Var(\hat{\theta}_I) = \frac{1}{M}\sigma^2}$$

$$I_j = \{x : a_j \le x < a_{j+1}\}$$

$j = 1, 2, \dots, k-1$

$$g(x \mid J=j) = g_j(x) = \begin{cases} g(x) & \text{if } x \in I_j \\ 0 & \text{o.w.} \end{cases}$$

$$\theta = \theta_1 + \theta_2 + \dots + \theta_k$$

$$\frac{\sigma^2}{M} - \frac{k}{M}\sum_{j=1}^{k}\sigma_j^2 \ge 0$$

$$\sigma^2 - k\sum_{j=1}^{k}\sigma_j^2 \ge 0$$

$$f_j(x) = f(x \mid I_j) = \frac{f(x, I_j)}{f(I_j)} = \frac{f(x, I_j)}{1/k}$$

$$= \frac{f(x)}{1/k} = k \cdot f(x)$$

$$E(\hat{\theta}^{SI}) = \theta$$

$$Var(\hat{\theta}^{SI}) \le Var(\hat{\theta}^{I})$$

$$x_1^{(j)}, x_2^{(j)}, \dots, x_m^{(j)} \sim f_j(x)$$

$j = 1, \dots, k$

$$\hat{\theta}_j = \frac{1}{m_j}\sum_{i=1}^{m_j}\frac{g_j(x_i)}{f_j(x_i)}$$

$$\theta = \sum_{j=1}^{k} E_{f|j}\left[\frac{g_j(x)}{f_j(x)}\right]$$

$$\hat{\theta}^{SI} = \hat{\theta}_1 + \hat{\theta}_2 + \dots + \hat{\theta}_k$$

$$Var\left(\sum_{j=1}^{k}\hat{\theta}_j\right) \le \boxed{Var(\hat{\theta}^{I})}$$

$$\sum_{j=1}^{k} Var(\hat{\theta}_j) \le Var(\hat{\theta}^{I})$$

$$\sum_{j=1}^{k} Var\left(\frac{1}{m}\sum_{i=1}^{m}\frac{g_j(x_i)}{f_j(x_i)}\right)$$

$$= \sum_{j=1}^{k}\frac{1}{m} Var\left(\frac{g_j(x)}{f_j(x)}\right)$$

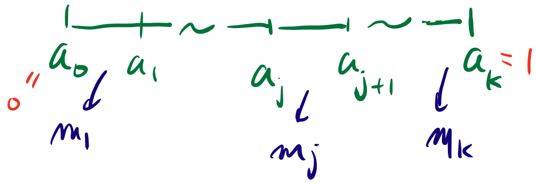
$\sum_{j=1}^{k} m_j = M$

$m = M/k$

$$= \sum_{j=1}^{k}\frac{1}{m}\sigma_j^2 = \boxed{\frac{k}{M}\sum_{j=1}^{k}\sigma_j^2}$$

Stratified sampling

$$\theta = \int_{a_0}^{a_k} g(x)\, dx$$



assume $m_1 = m_2 = \cdots = m_k = m$

$$\theta = (\theta_1 + \theta_2 + \cdots + \theta_k)\frac{1}{k}$$

$$= \theta_1 \frac{1}{k} + \theta_2 \frac{1}{k} + \cdots + \theta_k \frac{1}{k}$$

$$\sum_{i=1}^{m_1} \frac{g(x_i^{(1)})}{m_1}\frac{1}{k} + \cdots + \sum_{i=1}^{m_k} \frac{g(x_i^{(k)})}{m_k}\cdot\frac{1}{k}$$

$$\downarrow$$

$$E[g(x) \mid J=k]$$

$$E[E[g(x) \mid J]] = E[g(x)]$$

$X \sim \text{Unif}$

$$\boxed{E[E_J[g(x) \mid J]]} = \sum_{j=1}^{k} E[g(x) \mid J=j] \cdot P(J=j)$$

$$= \sum_{j=1}^{k} \sum_{x} g(x) \cdot P(x \mid J=j) \cdot P(J=j)$$

$$= \sum_{j=1}^{k} \sum_{x} g(x \mid J=j) \cdot P(x \mid J=j)$$

$$= \sum_{j=1}^{k} \sum_{x} g_j(x) \cdot P_j(x)$$

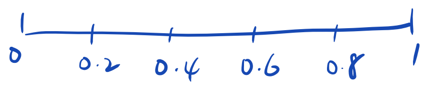
$$= \boxed{\sum_{j=1}^{k} E_{P|j}[g_j(x)]} = E[g(x)]$$

$$E_f\left[\frac{g(x)}{f(x)}\right] = E\left[E_f\left[\frac{g(x \mid J)}{f(x \mid J)}\right]\right] = \sum_{j=1}^{k} E_{f|j}\left[\frac{g(x \mid J=j)}{f(x \mid J=j)}\right]$$

$$= \sum_{j=1}^{k} E_{f|j}\left[\frac{g_j(x)}{f_j(x)}\right]$$

E.g.

$$\theta = \int_0^1 \boxed{\frac{e^{-x}}{1+x^2}}^{g(x)} dx \qquad \underline{f(x)} = \frac{e^{-x}}{1-e^{-1}}$$



$k=5$

① find $\underline{f_j(x)} = 5 \cdot \frac{e^{-x}}{1-e^{-1}} \qquad j=1,2,3,4,5$

$\qquad x_1^{(j)}, \ldots, x_{m_j}^{(j)} \sim f_j(x)$

② $\hat{\theta}_j = \underline{\frac{1}{m_j} \sum_{i=1}^{m_j} \frac{g_j(x_i)}{f_j(x_i)}}$

③ $\hat{\theta}_1 + \hat{\theta}_2 + \hat{\theta}_3 + \hat{\theta}_4 + \hat{\theta}_5 = \hat{\theta}^{S2}$

Find Z_g

e.g.

$$g(x) = e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}$$

$$\int_{-\infty}^{\infty} g(x)\, dx = Z_g \qquad \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$$

$$g(x) = e^{-3x} \quad \text{for} \quad x > 0 \qquad Z_g = \int_0^{\infty} e^{-3x} dx$$

$$\int_0^{\infty} 3 e^{-3x} dx = 1 \quad \Rightarrow \quad Z_g = \frac{1}{3}$$

Unnormalized Density

Let $q(x)$ be a function defined on a region D. Suppose that

- $q(x) > 0$, for $x \in D$
- $\int_D q(x)dx = Z_q < \infty$.

Then $q(x)$ is an unnormalized density on D. The corresponding normalized density is

$$f(x) = \frac{q(x)}{Z_q}.$$

Note that:

- For any normalized density, there are many unnormalized densities.
- For any unnormalized density, there is only one normalized density.

Self-Normalized Importance Sampling

- Let $f(x)$ be a normalized density, for $x \in D$, where D is the support of X.
- Let $q(x)$ be an unnormalized density for $f(x)$ with normalizing constant $Z_q = \int_D q(x)dx$, i.e., $f(x) = \frac{q(x)}{Z_q}$.
- Suppose we want to estimate $E_f[g(X)] = \int_D g(x)f(x)dx = \int_D g(x)\frac{q(x)}{Z_q}dx$, but Z_q is unknown and we are not able to sample from $f(x)$ directly.
- How can we estimate $E_f[g(X)]$ when we only know the unnormalized density $q(x)$?

$$f(x|y) = \frac{f(y|x)\cdot f(x)}{\int f(y|x)\cdot f(x)\,dx} = \frac{g(x)}{\int g(x)\,dx} = \frac{g(x)}{Z_g}$$

$$\boxed{E_{f(x|y)}(g(x))}$$

$$f(\theta|Data) = \frac{f(Data|\theta)\cdot f(\theta)}{\int f(Data|\theta)\cdot f(\theta)\,d\theta} \propto f(Data|\theta)\,f(\theta)$$

A trial distribution function $h(x) = \frac{r(x)}{Z_r}$, $Z_r = \int r(x)\,dx$

(1) generate $x_1, \ldots, x_m \overset{iid}{\sim} h(x)$ and compute the weight

$$w(x) = \frac{g(x)}{r(x)}$$

$$\boxed{f(x) = \frac{g(x)}{Z_g}}$$

The goal : $E_f[g(x)]$

(2) Compute $\hat{\theta} = \frac{\sum_{i=1}^{m} g(x_i)\cdot w(x_i)}{\sum_{i=1}^{m} w(x_i)}$

$$E(\hat{\theta}) = E_f[g(x)] = \theta$$

$$\frac{1}{m}E\left(\sum_{i=1}^{m} w(x_i)\right) = \frac{1}{m}E\left(\sum_{i=1}^{m}\frac{g(x_i)}{r(x_i)}\right) \xrightarrow{\text{by LLN}} E_h\left[\frac{g(x)}{r(x)}\right]$$

$$E_h\left[\frac{g(x)}{r(x)}\right] = \int \frac{g(x)}{r(x)}\cdot h(x)\,dx = \int \frac{g(x)}{\cancel{r(x)}}\,\frac{\cancel{r(x)}}{Z_r}\,dx$$

$$= \frac{1}{Z_r}Z_g = \frac{Z_g}{Z_r} \quad \cdots (1)$$

$$\frac{1}{m} E\left[\underbrace{\sum_{i=1}^{m} g(x_i)\frac{g(x_i)}{r(x_i)}}\right] \longrightarrow \underbrace{E_h\left[g(x)\frac{g(x)}{r(x)}\right]}$$

$$E_h\left[g(x)\cdot\frac{g(x)}{r(x)}\right] = \int g(x)\cdot\frac{g(x)}{r(x)}\cdot h(x)\,dx$$

$$= \frac{Z_g}{Z_r} E_f[g(x)] \quad \cdots \text{②}$$

$$E(\hat{\theta}) = E_f[g(x)]$$

Example: Folded Normal Distribution

- Let $q(x)$ be an unnormalized density for $f(x)$, given by
  $q(x) = e^{-x^2/2}, \quad \text{for } x \geq 0.$
- We want to use self-normalized importance sampling to estimate

$g(x) = x$

$$\theta = \boxed{E_f(X)} = \int_0^\infty x f(x) dx = \int_0^\infty x \frac{q(x)}{Z_q} dx.$$

$\frac{1}{Z_r} = \boxed{\lambda}\, \boxed{e^{-\lambda x}}$, $r(x)$

- We previously found the theoretical $E_f(X) = \sqrt{\frac{2}{\pi}}$.

trial density : exp(2) $\quad h(x) = \frac{r(x)}{Z_r} \quad r(x) = e^{-2x}$

① $x_1, x_2, \ldots, x_m \overset{iid}{\sim} \exp(2)$

$w(x) = \frac{q(x)}{r(x)} = e^{-\frac{x^2}{2}} / e^{-2x} = e^{-\frac{x^2}{2} + 2x}$

② $\hat{\theta} = \frac{\sum x_i w(x_i)}{\sum w(x_i)}$

R Code to estimate $E_f(X)$ (self-normalized importance sampling):

```
set.seed(9999) # for reproduceability

n <- 10000 # Specify the number of points to generate

# Generate n points from Exp(lambda = 2)
X <- rexp(n, rate = 2)

# Compute importance weights
W <- exp(-X^2 / 2 + 2 * X)

# Compute sum(w(X) * X) / sum(w(X))
sum(W * X) / sum(W)
```

```
## [1] 0.800945
```