# Importance Sampling

Dr. Uma Ravat

PSTAT 194CS

# Classic Monte Carlo Integration estimator

$$\theta = E_f[h(X)] = \int h(x)f(x)dx \approx \frac{1}{n}\sum_{i=1}^{n} h(x_i) = \bar{h}_n = \hat{\theta},$$

- is unbiased
- we have computed its variance
- $\hat{\theta}$ converges in probability to $\theta$

**Drawback**

- If $f(x)$ is uniform, calculation of $\bar{h}_n$ easy but may not be efficient if the function $h$ has considerable variation.
- cannot be directly applied to unbounded intervals
- If $f(x)$ is not uniform then calculation of $\bar{h}_n$ depends on our ability to produce random samples from the p.d.f $f(x)$

In such cases, we can perform importance sampling

# Importance sampling

Consider evaluating $\theta = E_f[h(X)] = \int_A h(x)f(x)dx$

The set $A$ is the support of the density $f(x)$, ie $f(x) = 0, x \notin A$

If we can't generate from $f(x)$ or generating from $f(x)$ is inefficient:

- We instead choose an entirely different distribution $g(x)$, the importance sampling distribution and perform importance sampling
- The only requirement for $g(x)$ is that the support of $f(x)$ has to be within the support of $g(x)$. ($g(x) > 0$ for all $x \in A$)

# Importance sampling

We express $E_f[h(X)]$ as an expectation in terms of $g(x)$

$$
\begin{aligned}
E_f[h(X)] &= \int_A h(x) f(x) dx \\
&= \int_A h(x) \frac{f(x)}{g(x)} g(x) dx \\
&= \int [h(x) \frac{f(x)}{g(x)}] g(x) dx \quad f(x) = 0 \text{ for } x \notin A \\
&= E_g[h(X) \frac{f(X)}{g(X)}]
\end{aligned}
$$

# Importance sampling algorithm

Estimate $E_f[h(X)]$ by choosing an importance sampling function $g(x)$ such that support of $f(x)$ is contained in support of $g(x)$ then

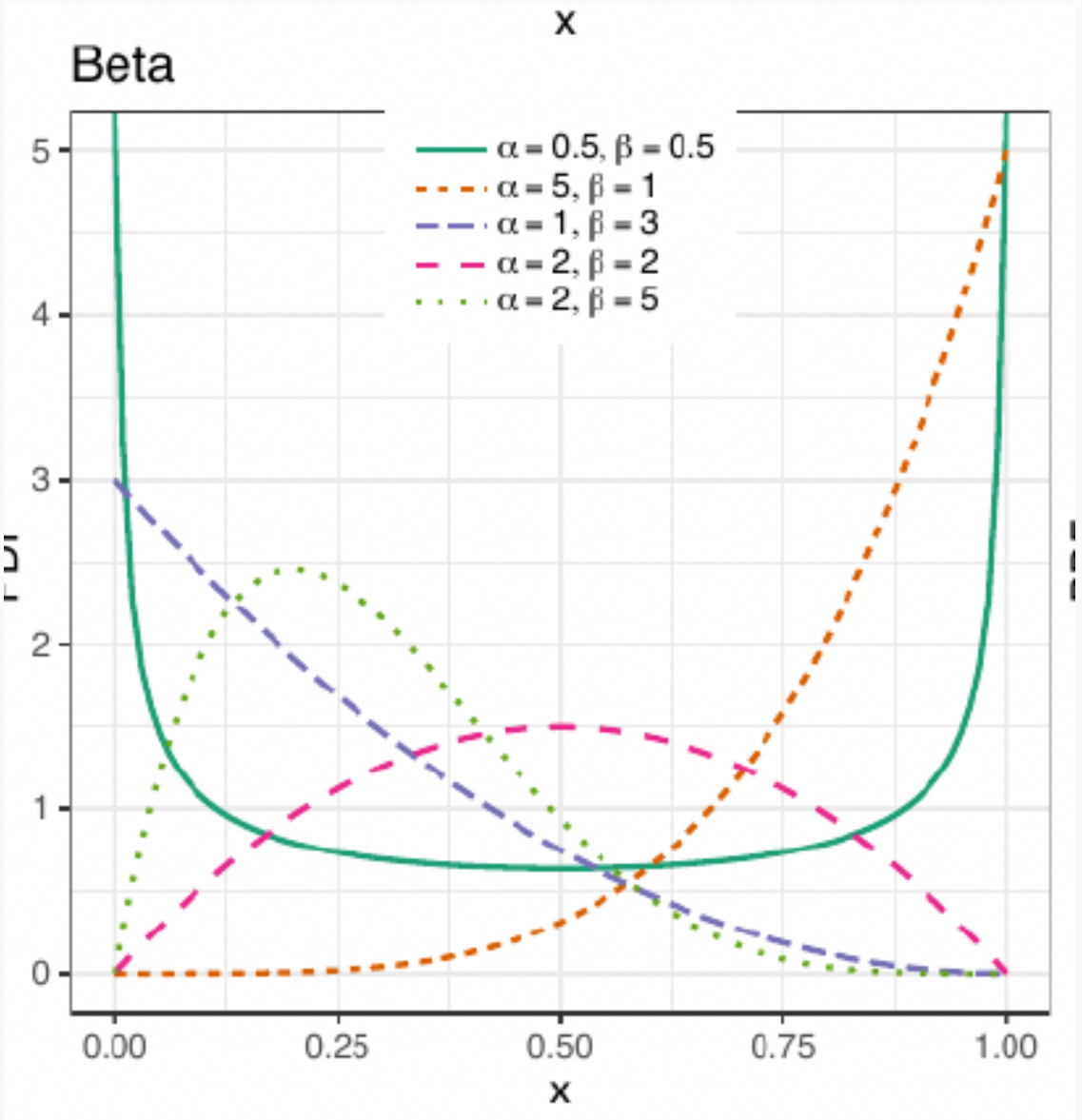
1. Generate $x_1, x_2, \cdots x_n$ iid $\sim g(x)$
2. Compute $h(x_1)\frac{f(x_1)}{g(x_1)}, h(x_2)\frac{f(x_2)}{g(x_2)}, \cdots h(x_n)\frac{f(x_n)}{g(x_n)}$
3. Estimate $E_f[h(X)]$ by

$$E_f[h(X)] = E_g[h(X)\frac{f(X)}{g(X)}] \approx \frac{1}{n}\sum_{i=1}^{n} h(x_i)\frac{f(x_i)}{g(x_i)}$$

It works by intelligently reweighting the distribution $g(x)$ based on "importance" to $f(x)$ distribution

Choosing $g(x)$ wisely, will ensure reduction in variance.

# Example with Beta's



# Example with Beta's

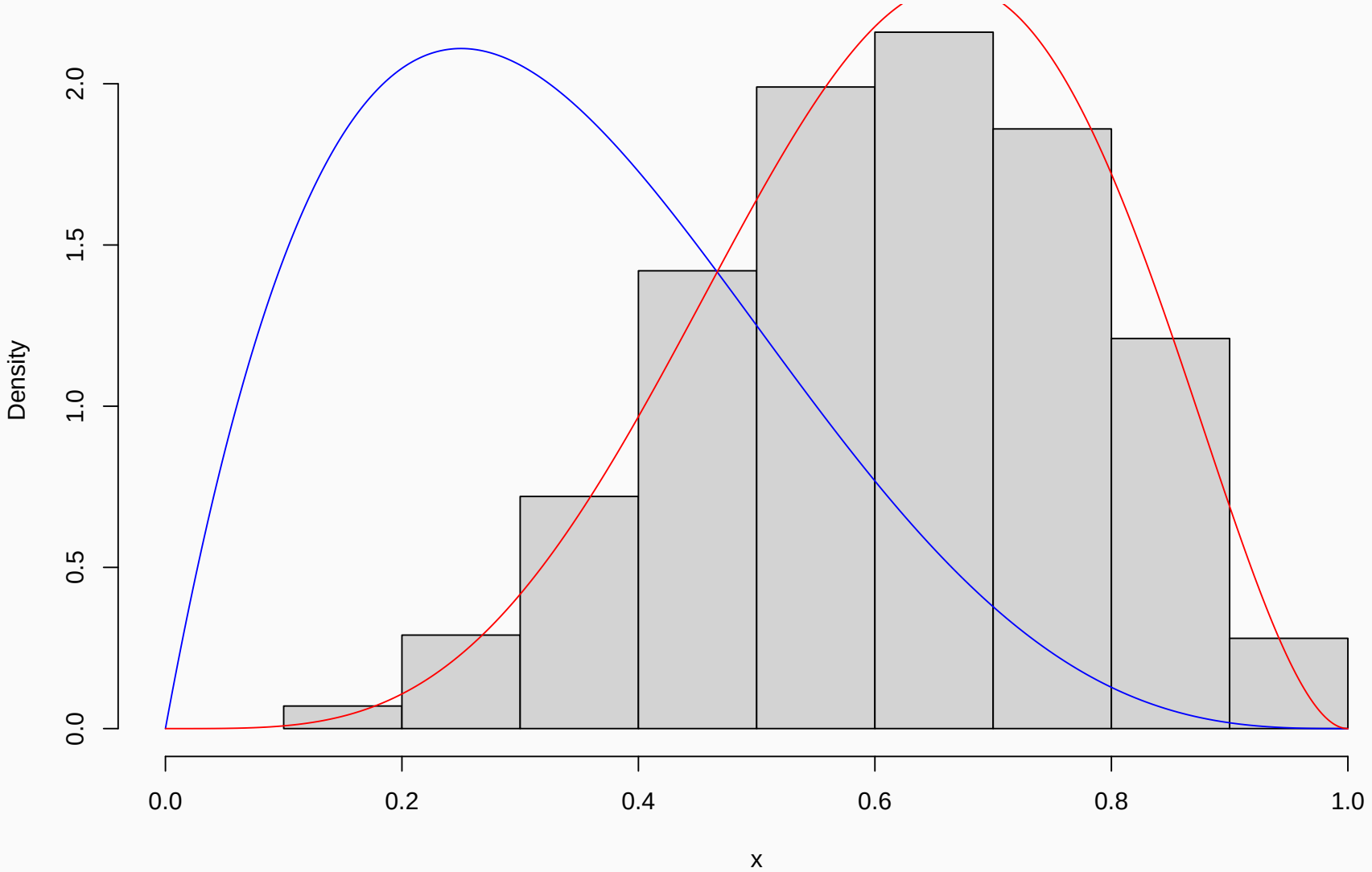
- We want to estimate the mean of $f$, where $f \sim Beta(2, 4)$ distribution
- $g \sim Beta(5, 3)$ distribution

Assume we can draw from $g$ but not from $f$ (silly assumption, but yes)

- The mean of $f$ is $\frac{2}{2+4} = 0.333$
- The mean of $g$ is $\frac{5}{5+3} = 0.625$

```
set.seed(100)
n <-1000
x <- rbeta(n,5,3)
mean(x) #Monte carlo estimate for mean of g
```

```
## [1] 0.6161036
```


histogram of x with density curves: f in blue, g in red
Density
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0
x

# Importance weight

To estimate the mean of $f$ from samples from $g$, I need to reweight each observation from $g$ by it's "importance" in $f$.

$$w(x_i) = \frac{f(x_i)}{g(x_i)}$$

When I draw from $g$, I will get many values close to 0.7 and not enough close 0.3. To correct for this, I reweight the values appropriately.

(higher weight for areas of high density of $f$ and lower weight for areas of low density of $f$)

# Importance weights

To get the mean of $f(x)$ when we sample from $g(x)$, we just multiply the sample value $x_i$ from $g$ by it's weight $w(x_i) = \frac{f(x_i)}{g(x_i)}$

```
set.seed(100)
n <-10000
x <- rbeta(n,5,3)
mean(x) #Monte carlo estimate for mean of g
```

```
## [1] 0.623744
```

```
w <- dbeta(x,2,4)/dbeta(x,5,3) # vectors of weights
mean(x*w) #Monte carlo estimate
```

```
## [1] 0.3390631
```

```
#mean of f using $g$ as importance function
```

# Importance sampling algorithm using importance weights

Estimate $E_f[h(X)]$ by choosing an importance sampling function $g(x)$ such that support of $f(x)$ is contained in support of $g(x)$ then

1. Generate $x_1, x_2, \cdots x_n$ iid $\sim g(x)$
2. Compute weights
   $w(x_1) = \frac{f(x_1)}{g(x_1)}, w(x_2) = \frac{f(x_2)}{g(x_2)}, \cdots w(x_n) = \frac{f(x_n)}{g(x_n)}$
3. Estimate $E_f[h(X)]$ by

$$E_f[h(X)] = E_g[h(X)\frac{f(X)}{g(X)}] \approx \frac{1}{n}\sum_{i=1}^{n} h(x_i)w(x_i)$$

# How to choose the importance sampling function $g(x)$

The choice of the **importance sampling function** $g(x)$ influences the variance of our importance sampling estimator $\hat{\theta}$ for $E_f[h(X)]$

- Importance sampling activity will show how to choose the best importance sampling function

Importance sampling achieves the greatest variance reduction compared with the standard Monte Carlo estimator if shape of $g(x)$ is roughly proportional to shape of $|h(x)|f(x)$.

# Variance of the importance sampling estimator

Let's take a look at the variance of the importance sampling estimator, and see how it relates to $g$. Consider the parameter

$$\theta = \int_A h(x)f(x)dx = \int_A h(x)\frac{f(x)}{g(x)}g(x)dx = E_g\left[h(X)\frac{f(X)}{g(X)}\right]$$

Draw a sample $X_1, X_2, ..., X_n$ with pdf $g$ and construct the estimator

$$\hat{\theta} = \frac{1}{m}\sum_{i=1}^{m} h(X_i)\frac{f(X_i)}{g(X_i)}$$

$$Var(\hat{\theta}) = E[\hat{\theta}^2] - (E[\hat{\theta}])^2$$

This variance is minimized when

$$g(x) = \frac{\mid h(x)f(x) \mid}{\int_A \mid h(x)f(x) \mid dx}$$

However, we cannot directly use such a function, because it involves knowing the very thing we are trying to compute.

The idea is to try and find a density function $g(x)$ that is close to a positive and normalized version of $h(x)f(x)$, but can be drawn from.

# Stratified Sampling

# Stratified Sampling

Another approach to variance reduction is to use simple random sampling from uniform distributions, but stratify these to ensure balance over a partition into $k$ subintervals of an interval $(a, b)$.

Let $a = x_0 < x_1 < ... < x_k = b$.

Then

$$\theta = \int_a^b g(t)dt = \sum_{j=1}^{k} \int_{x_{j-1}}^{x_j} g(t)dt.$$

If we define

$$\theta_j = \int_{x_{j-1}}^{x_j} g(t)dt$$

for $j = 1, 2, 3, ..., k$, then $\theta = \sum_{j=1}^{k} \theta_j$.

Use the Monte Carlo estimator

$$\hat{\theta}_j = \frac{x_j - x_{j-1}}{m_j} \sum_{i=1}^{m_j} g(X_{ij})$$

Then we compute $\hat{\theta} = \sum_{j=1}^{k} \hat{\theta}_j$. This method guarantees some level of balance in sampling over the interval $(a, b)$ if all the $m_j$s are nearly the same.

In fact, the variance of the stratified estimator will be less than the variance of the standard Monte Carlo estimator with $m$ replications if $m = \sum_{j=1}^{k} m_j$ where $m_j = m/k$.

- Activity verifies this

# Stratified Importance Sampling

# Stratified Importance Sampling

$$\theta = \int_a^b g(t)dt = \sum_{j=1}^{k} \int_{x_{j-1}}^{x_j} g(t)dt$$

and

$$\theta_j = \int_{x_{j-1}}^{x_j} g(t)dt$$

for $j = 1, 2, 3, ..., k$.

In the activity, we verified stratification can lead to variance reduction compared with the standard Monte Carlo estimator.

Further gains can be made using importance sampling, with an importance sampling function $\phi_j$ chosen wisely for each stratum.

# Stratified Importance Sampling

Let $\hat{\theta} = \sum_{j=1}^{k} \hat{\theta}_j$ where

$$\hat{\theta}_j = \frac{1}{m_j} \sum_{i=1}^{m_j} \frac{g(X_{ij})}{\phi_j(X_{ij})}$$

where $X_{1j}, X_{2j}, ..., X_{m_j j}$ are drawn from a density $\phi_j(x)$, the importance sampling function, that is supported on the interval $(x_{j-1}, x_j)$.

This idea combines the benefits of stratification and importance sampling. The importance sampling function should be something that may be conveniently drawn from and would ideally be roughly proportional to $|g(x)|$ for $x \in (x_{j-1}, x_j)$.

# Summary

# Summary

To summarize our discussion of Monte Carlo integration, we have seen

1. the standard Monte Carlo estimator
2. importance sampling and
3. stratification.

Importance sampling provides a direct way to integrate over unbounded intervals and also can reduce the variance of the integral estimator if the importance function is chosen carefully.

Stratification can also be used to reduce variance and can be combined with importance sampling.