STAT400 - Homework 7

Henrique Magalhaes Rio

Due 10/22/2020 by 4pm

Be sure to `set.seed(400)` at the beginning of your homework.

```r
#reproducibility
set.seed(400)
```

```r
# useful libraries
library(tidyverse)
```

1. Find two importance functions ϕ_1 and ϕ_2 that are supported on $(1, \infty)$ and are "close" to

$$h(x) = \frac{x^2}{\sqrt{2\pi}} e^{-x^2/2}, \qquad x > 1.$$

Which of your two importance functions should produce the smallest variance in estimating

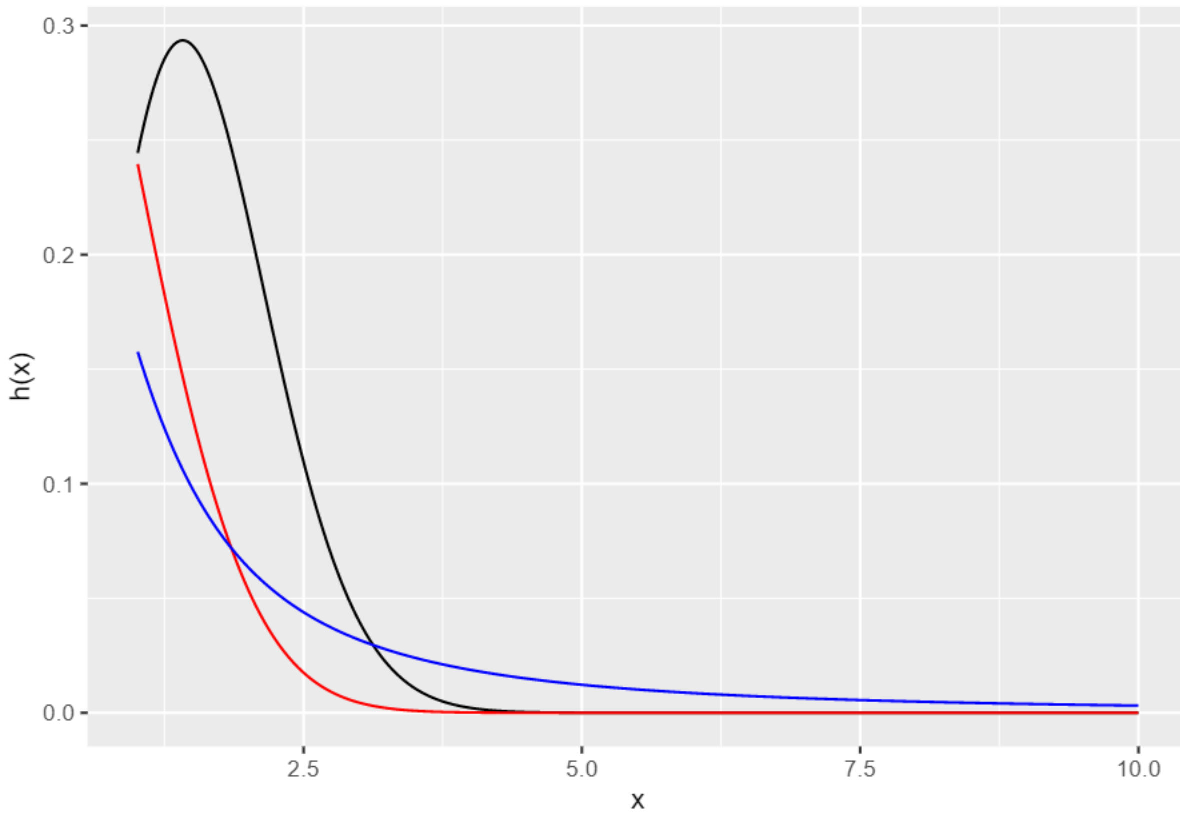
$$\int_1^\infty \frac{x^2}{\sqrt{2\pi}} e^{-x^2/2} dx$$

by importance sampling? Explain.

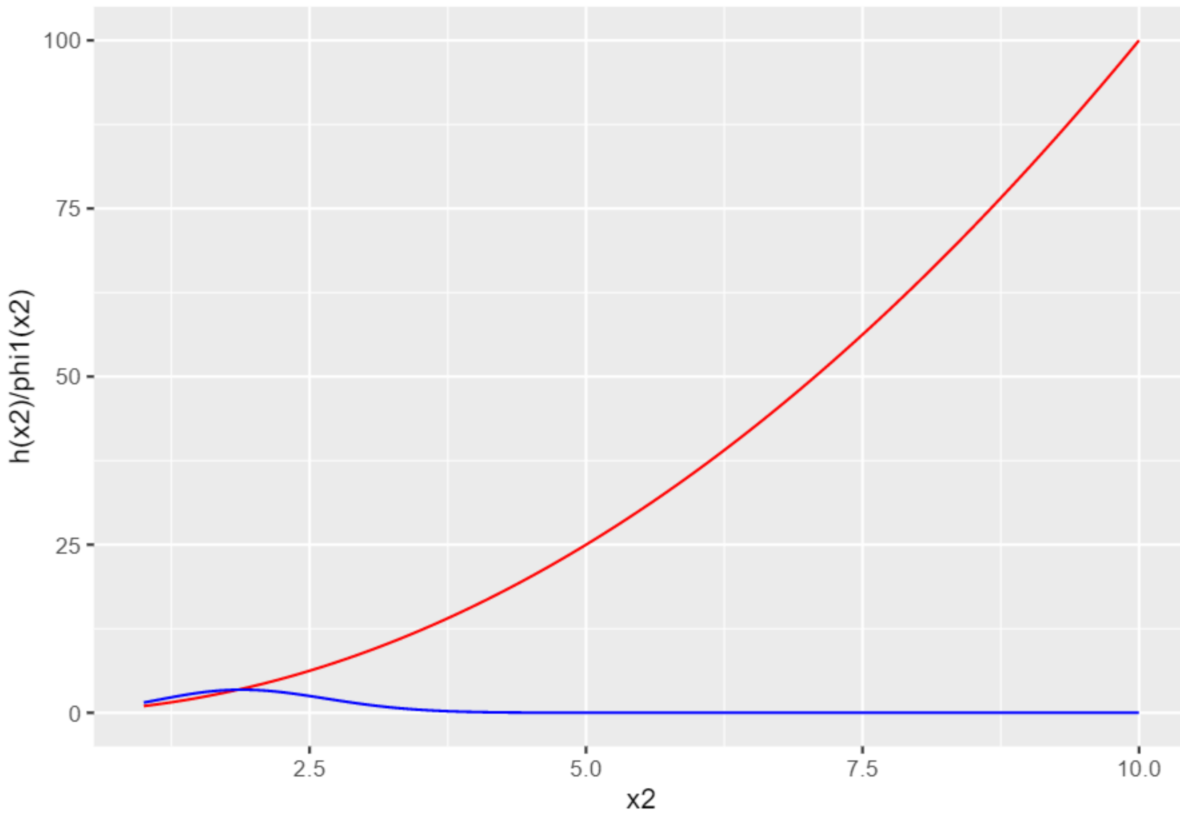
Hint: You will need to create plots of ϕ_1, ϕ_2, and $g(x)f(x)$ as well as $g(x)f(x)/\phi(x)$ to answer this question.

```r
## create functions for h(x), phi_1(x), and phi_2(x)
h <- function(x){
    x^2/sqrt(2*pi)*exp(-x^2/2)
}

f <- function(x){
     dnorm(x,0,1)
}

phi1 <- function(x){
    dnorm(x,0,1)
}

phi2 <- function(x){
    dcauchy(x,0,1)
}

## plot h(x), phi_1(x), and phi_2(x) together
x <- seq(1.01, 10, length.out = 500)
 ggplot() +
  geom_line(aes(x, h(x)), color = "black") +
```

```r
  geom_line(aes(x, phi1(x)), color = "red") +
  geom_line(aes(x, phi2(x)), color = "blue")
```




```r
## plot h(x)/phi_1(x) and h(x)/phi_1(x)
 x2 <- seq(1, 10, length.out = 500)
 ggplot() +

  geom_line(aes(x2, h(x2)/phi1(x2)), color = "red") +
  geom_line(aes(x2, h(x2)/phi2(x2)), color = "blue")
```




```
## which is better? Why?
```

Looking at the second graph, I believe that ϕ_2 should have a smaller variance when compared to ϕ_1, which is seem by the graph being constant.

2. Obtain a Monte Carlo estimate of

$$\int_1^{\infty} \frac{x^2}{\sqrt{2\pi}} e^{-x^2/2} dx$$

using importance sampling with the two importance sampling functions you chose in Problem 4 (ϕ_1 and ϕ_2). Obtain an estimate of the variance for each and compare.

```
## number of samples to use
m <- 100000

## estimate the integral using importance sampling
## phi 1

theta_1 <- function(m) {

x <- rnorm(m,0,1)
g <- x^2*f(x)/phi1(x)*(x>1)
}
theta1<- theta_1(m)

mean(theta_1(m))
```

```
## [1] 0.4010379
```

```r
## phi 2
theta_2 <- function(m) {

x<- rcauchy(m,0,1)
g <- x^2*f(x)/phi2(x)*(x>1)
}
theta2<- theta_2(m)

mean(theta_2(m))
```

```
## [1] 0.3993751
```

```r
## estimates


## true value for theta is approx 0.4006


## compare variances
var_theta_hat1 <- 1/m * mean((theta1 -  mean(theta_1(m)))^2)
var_theta_hat2 <- 1/m * mean((theta2 -  mean(theta_2(m)))^2)


var_theta_hat1
```

```
## [1] 1.274208e-05
```

```r
var_theta_hat2
```

```
## [1] 9.093357e-06
```

For ϕ_1:

$\hat{\theta} = 0.3989501$ $Var(\hat{\theta}) = 1.28249e^{-05}$

For ϕ_2:

$\hat{\theta} = 0.401192$ $Var(\hat{\theta}) = 9.099747e^{-06}$

These results confirm what we saw in the graph, that is, ϕ_2 has a smaller variance then ϕ_1.