Worksheet 3

Foundations of Bayesian Methodology

Wenje Tu Lea Bührer Jerome Sepin Zhixuan Li
Elia-Leonid Mastropietro Jonas Raphael Füglistaler

Spring Semester 2022

Exercise 3 (Conjugate Bayes: analytical derivation)

Assumptions:

$$\begin{aligned} y_1, \cdots, y_n \mid m &\overset{i.i.d}{\sim} \mathcal{N}(m, \kappa^{-1}) \\ m &\sim \mathcal{N}(\mu, \lambda^{-1}) \end{aligned}$$

3(a) The prior predictive distribution of one future observation y assuming that no observations have been collected yet.

$$\begin{aligned}
f(y) &= \int_{-\infty}^{\infty} f(y \mid m) f(m) \mathrm{d}m \\
&= \int_{-\infty}^{\infty} \sqrt{\frac{\kappa}{2\pi}} \exp\left(-\frac{\kappa}{2}(y-m)^2\right) \sqrt{\frac{\lambda}{2\pi}} \exp\left(-\frac{\lambda}{2}(m-\mu)^2\right) \mathrm{d}m \\
&= \frac{\sqrt{\kappa\lambda}}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\kappa(y^2 - 2ym + m^2) + \lambda(m^2 - 2m\mu + \mu^2)\right)\right) \mathrm{d}m \\
&= \frac{\sqrt{\kappa\lambda}}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left((\kappa+\lambda)\left(m - \frac{\kappa y + \lambda\mu}{\kappa+\lambda}\right)^2 - \frac{(\kappa y + \lambda\mu)^2}{\kappa+\lambda} + \kappa y^2 + \lambda\mu^2\right)\right) \mathrm{d}m \\
&= \frac{\sqrt{\kappa\lambda}}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left((\kappa+\lambda)\left(m - \frac{\kappa y + \lambda\mu}{\kappa+\lambda}\right)^2 + \frac{\kappa\lambda(y-\mu)^2}{\kappa+\lambda}\right)\right) \mathrm{d}m \\
&= \frac{\sqrt{\kappa\lambda}}{2\pi} \exp\left(-\frac{\kappa\lambda(y-\mu)^2}{2(\kappa+\lambda)}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{\kappa+\lambda}{2}\left(m - \frac{\kappa y + \lambda\mu}{\kappa+\lambda}\right)^2\right) \mathrm{d}m \\
&= \frac{\sqrt{\kappa\lambda}}{2\pi} \sqrt{\frac{2\pi}{\kappa+\lambda}} \exp\left(-\frac{\kappa\lambda(y-\mu)^2}{2(\kappa+\lambda)}\right) \underbrace{\int_{-\infty}^{\infty} \sqrt{\frac{\kappa+\lambda}{2\pi}} \exp\left(-\frac{\kappa+\lambda}{2}\left(m - \frac{\kappa y + \lambda\mu}{\kappa+\lambda}\right)^2\right) \mathrm{d}m}_{=1} \\
&= \frac{\sqrt{\kappa\lambda}}{2\pi} \sqrt{\frac{2\pi}{\kappa+\lambda}} \exp\left(-\frac{\kappa\lambda(y-\mu)^2}{2(\kappa+\lambda)}\right) \\
&= \sqrt{\frac{1}{2\pi\left(\frac{1}{\lambda} + \frac{1}{\kappa}\right)}} \exp\left(-\frac{(y-\mu)^2}{2\left(\frac{1}{\lambda} + \frac{1}{\kappa}\right)}\right)
\end{aligned}$$

The prior predictive distribution of one future observation y is

$$\mathcal{N}(\mu, \lambda^{-1} + \kappa^{-1})$$

3(b) The posterior predictive distribution of one future observation y_{n+1} given that $y_1, \cdots, y_n$ have been observed.

$$
\begin{aligned}
f(y_{n+1} \mid y_1, \cdots, y_n) &= \int_{-\infty}^{\infty} f(y_{n+1}, m \mid y_1, \cdots, y_n) \mathrm{d}m \\
&= \int_{-\infty}^{\infty} f(y_{n+1} \mid m, y_1, \cdots, y_n) f(m \mid y_1, \cdots, y_n) \mathrm{d}m \\
&= \int_{-\infty}^{\infty} f(y_{n+1} \mid m) f(m \mid y_1, \cdots, y_n) \mathrm{d}m
\end{aligned}
$$

$$
f(m \mid y_1, \cdots, y_n) = \sqrt{\frac{n\kappa + \lambda}{2\pi}} \exp\left(-\frac{n\kappa+\lambda}{2}\left(m - \frac{\kappa n \bar{y} + \lambda\mu}{n\kappa+\lambda}\right)^2\right)
$$

Denote:

$$
\mu_{\text{post}} = \frac{\kappa n \bar{y} + \lambda\mu}{n\kappa + \lambda}
$$

$$
\lambda_{\text{post}} = n\kappa + \lambda
$$

$$
f(y_{n+1} \mid y_1, \cdots, y_n) = \int_{-\infty}^{\infty} \sqrt{\frac{\kappa}{2\pi}} \exp\left(-\frac{\kappa}{2}(y_{n+1} - m)^2\right) \sqrt{\frac{\lambda_{\text{post}}}{2\pi}} \exp\left(-\frac{\lambda_{\text{post}}}{2}(m - \mu_{\text{post}})^2\right) \mathrm{d}m
$$

Repeating the same derivation steps as for the prior predictive distribution, we obtain the posterior predictive distribution:

$$
y_{n+1} \mid y_1, \cdots, y_n \sim \mathcal{N}(\mu_{\text{post}}, \lambda_{\text{post}}^{-1} + \kappa^{-1})
$$

Exercise 4 (Conjugate Bayesian analysis in practice)

4(a) Plot the prior predictive distribution for one observation y and compute its expectation and standard deviation. Estimate P[y > 200] for one future observation of Height.

Prior predictive distribution:

$$
y \sim \mathcal{N}(\mu, \lambda^{-1} + \kappa^{-1}) = \mathcal{N}(161, 970)
$$

So the expectation for y equals $\mu = 161$ and the standard deviation is $\lambda^{-1} + \kappa^{-1} = 900 + 70 = 970$.

Prior predictive distribution:

```r
## P[y>200] for one future observation of Height
pnorm(200, mean=161, sd=sqrt(970), lower.tail=F)
```

```
## [1] 0.1052459
```

The probability that an observation of Height larger than 200 cm is made equals 0.105. So it can be concluded that we expect around 10% of future observation to be larger than 200 cm.

4(b) Posterior distribution:

$$
m \mid y_1, \cdots, y_n \sim \mathcal{N}\left(\frac{\kappa n \bar{y} + \lambda\mu}{n\kappa + \lambda}, (n\kappa + \lambda)^{-1}\right) = \mathcal{N}(164.558, 34.80663)
$$

Posterior predictive distribution:

$$
y_{n+1} \mid y_1, \cdots, y_n \sim \mathcal{N}\left(\frac{\kappa n \bar{y} + \lambda\mu}{n\kappa + \lambda}, (n\kappa + \lambda)^{-1} + \kappa^{-1}\right) = \mathcal{N}(164.558, 934.80663)
$$

Compute the expectation and standard error:

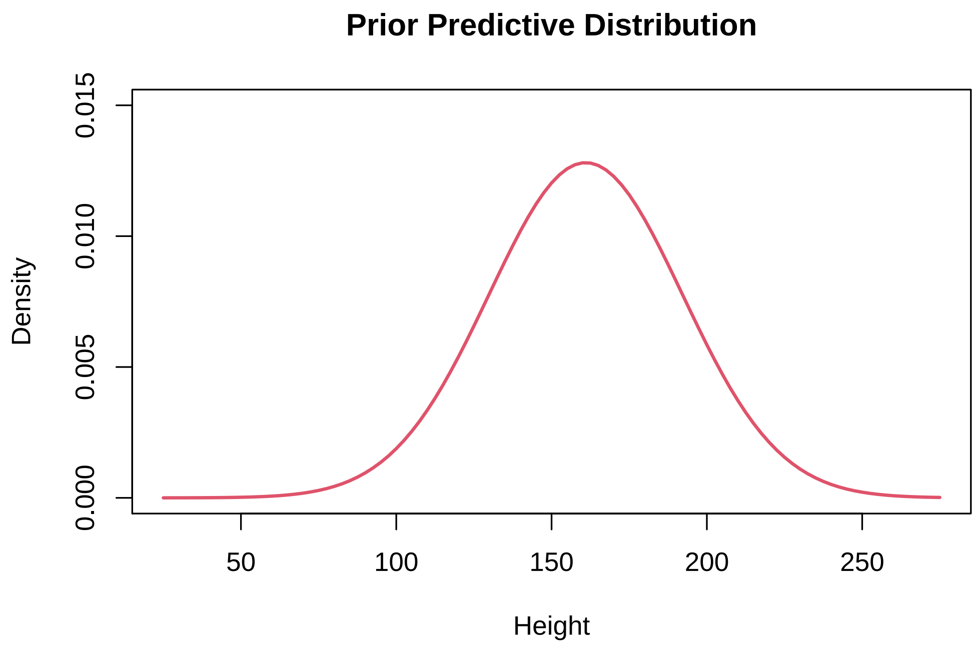


Figure 1: Prior Predictive Distribution

```
height <- c(166,168,168,177,160,170,172,159,175,164,175,167,164)
n = length(height)
y_bar = mean(height)
kappa = 1/900
mu = 161
lambda = 1/70

mu_post = (kappa*n*y_bar+lambda*mu)/(n*kappa+lambda)
mu_post
```

```
## [1] 164.558
```

```
sd_post = (n*kappa+lambda)^-1+kappa^-1
sd_post
```

```
## [1] 934.8066
```

Derived with the help of the formula above we obtain an expected value for $y_{n+1} = 164.558$ and a standard deviation of 934.80663.

With

- $n = 13$
- $\kappa = \frac{1}{900}$
- $\lambda = \frac{1}{70}$
- $\mu = 161$
- $\bar{y} = E[y_1, ..., y_n] = 168.0769$

Estimate for $P[y_{n+1} > 200|y_1, ..., y_n]$ for one future observation $y_n + 1$:

```
pnorm(200, mean=164.558, sd=sqrt(934.80663), lower.tail=F)
```

```
## [1] 0.1231879
```

As a result we obtain a probability of 12.3% that a future observation of height will be larger than 200 cm.

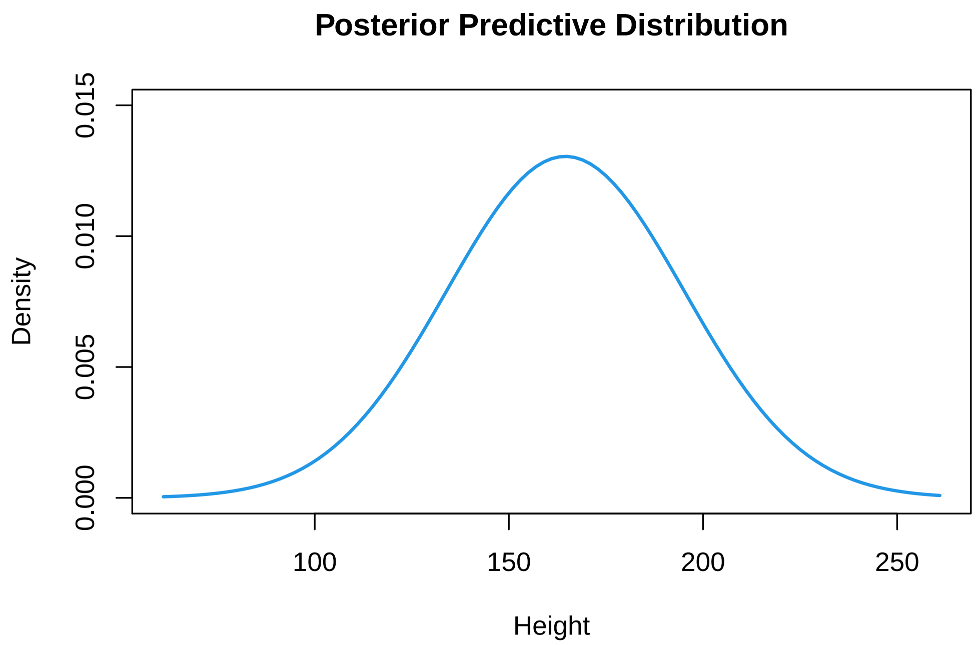


Figure 2: Posterior Predictive Distribution

4(c) Comparison between posterior, prior predictive, and posterior distributions

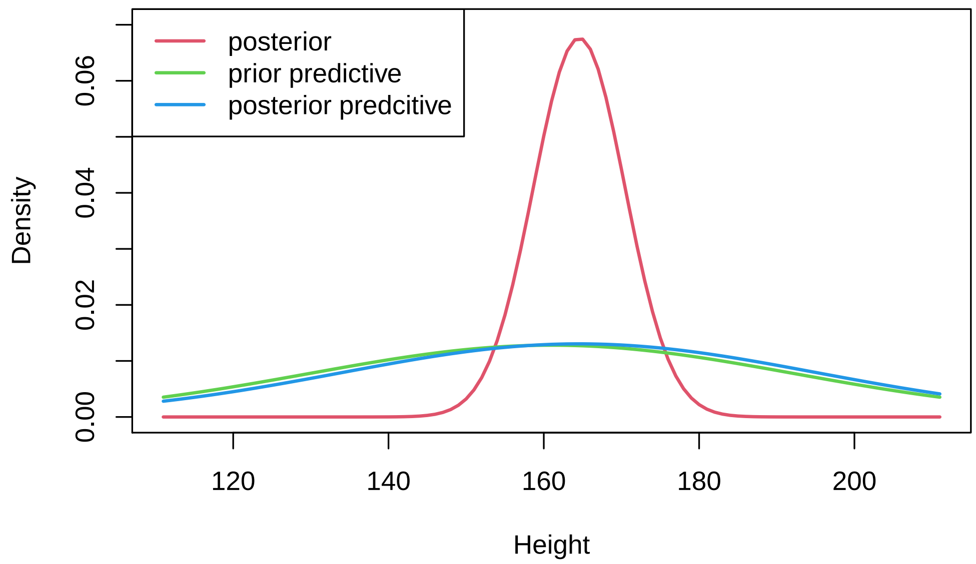


Figure 3: Comparison of Posterior, Prior Predictive and Posterior Predictive Distribution for Height.

ADD DISCUSSION, INTERPRETATION

Exercise 5 (The change-of-variables formula)

$$X \sim \text{Gamma}(a, b)$$

$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx)$$

$$\begin{aligned} P(Y \leq y) &= P(g(X) \leq y) \\ &= P(X \leq g^{-1}(y)) \\ F_Y(y) &= F_X(g^{-1}(y)) \end{aligned}$$

By differentiating the CDFs on both sides w.r.t. y, we can get the PDF of Y.

If the function $g(\cdot)$ is monotonically increasing:

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y)$$

If the function $g(\cdot)$ is monotonically decreasing:

$$f_Y(y) = -f_X(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y)$$

Therefore:

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right|$$

$$Y = \frac{1}{X} \implies X = \frac{1}{Y}$$

$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx)$$

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right| \\ &= \frac{b^a}{\Gamma(a)} \left(\frac{1}{y} \right)^{a-1} \exp(-\frac{b}{y}) \cdot \left| -\frac{1}{y^2} \right| \\ &= \frac{b^a}{\Gamma(a)} \left(\frac{1}{y} \right)^{a+1} \exp(-\frac{b}{y}) \end{aligned}$$

$$Z = \sqrt{\frac{1}{X}} \implies X = \frac{1}{Z^2}$$

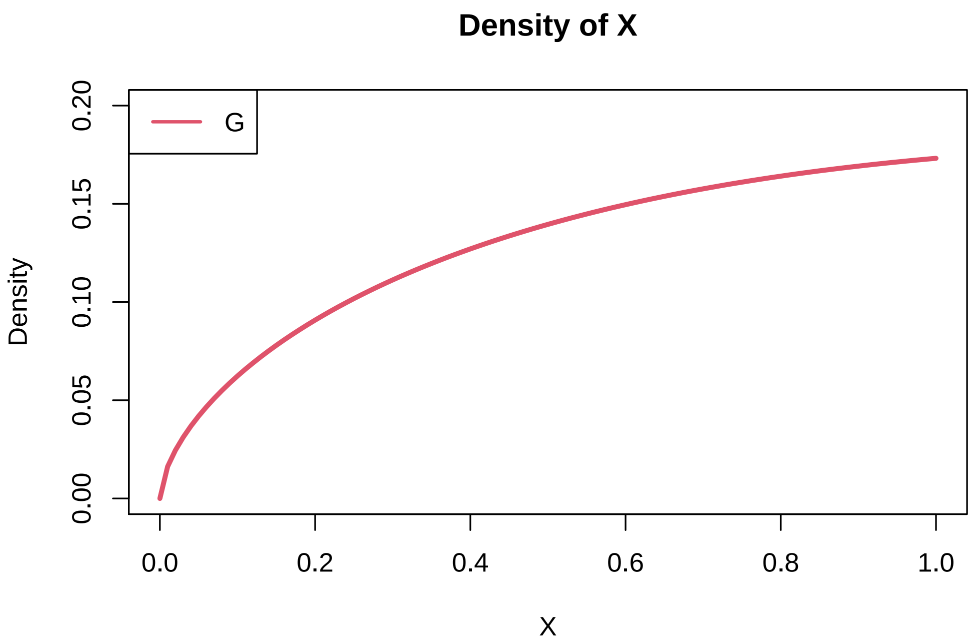
$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx)$$

$$\begin{aligned} f_Z(z) &= f_X(g^{-1}(z)) \cdot \left| \frac{d}{dz} g^{-1}(z) \right| \\ &= \frac{b^a}{\Gamma(a)} \left(\frac{1}{z^2} \right)^{a-1} \exp(-\frac{b}{z^2}) \cdot \left| -\frac{2}{z^3} \right| \\ &= \frac{b^a}{\Gamma(a)} 2 \left(\frac{1}{z} \right)^{2a+1} \exp(-\frac{b}{z^2}) \end{aligned}$$

```r
## Define inverse-gamma distribution function
dinvgamma <- function(x, a, b) {
  return(
    (b^a)/gamma(a) * (1/x)^(a+1) * exp(-b/x)
  )
}

## Define square root inverse-gamma distribution function
dsqrtinvgamma <- function(x, a, b) {
  return(
    2 * (b^a)/gamma(a) * (1/x)^(2*a+1) * exp(-b/x^2)
  )
}
```

```r
a <- 1.6
b <- 0.4
```

```r
curve(dgamma(x, shape=a, rate=b), ylim=c(0, 0.2), col=2, lwd=3,
      xlab="X", ylab="Density", main="Density of X")
legend("topleft", legend="G", col=2, lwd=2)

curve(dinvgamma(x, a, b), xlim=c(0, 0.5), col=3, lwd=2,
      main="Densities of Y and Z", y="Density", xlab="Variable")
curve(dsqrtinvgamma(x, a, b), add=TRUE, col=4, lwd=2)
legend("topleft", legend=c("IG", "SIG"), col=c(3, 4), lwd=2)
```



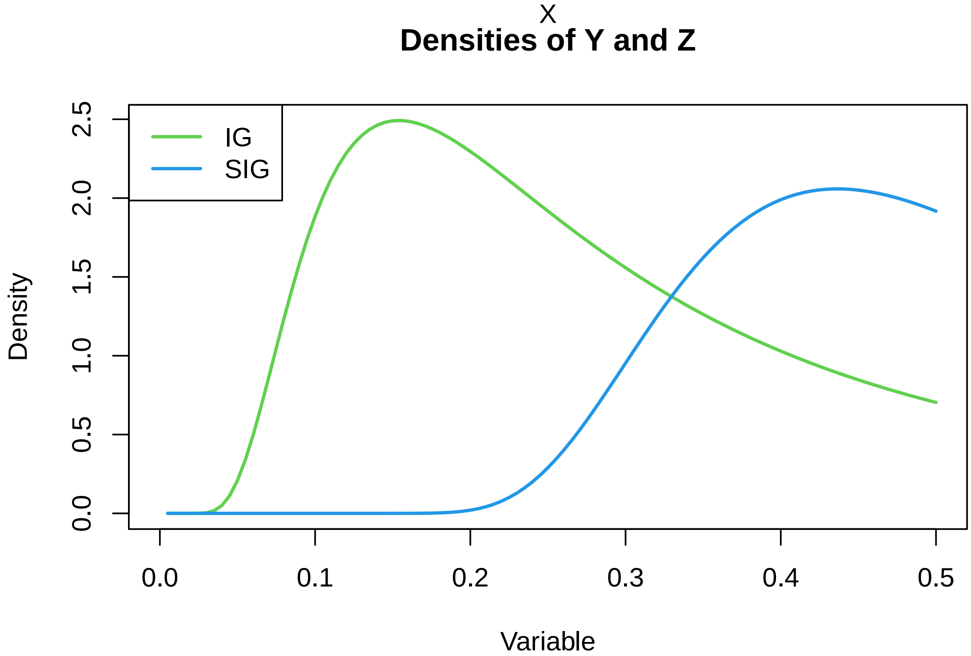


Figure 4: Density Plots.

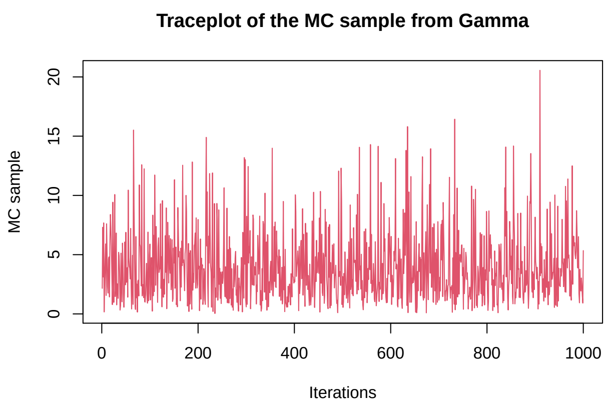
ADD INTERPRETATION: Interpret the shape of densities of Y and Z close to 0.

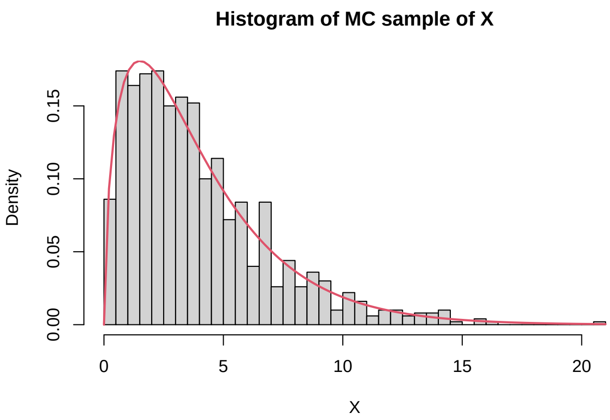
Exercise 6 (Monte Carlo: transformations of random variables)

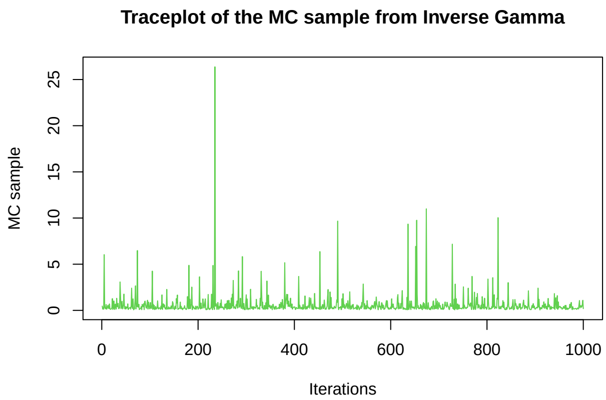
```r
## Set seed for reproducible results
set.seed(44566)

## Parameters for Gamma
a <- 1.6 # shape
b <- 0.4 # rate (inverse of scale)

## MC sample size
M <- 1000
```

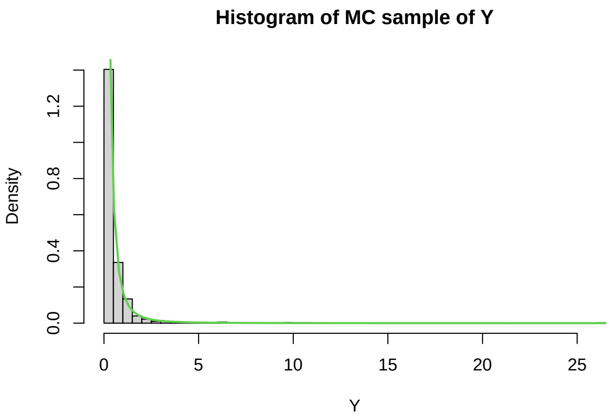
```r
## Generate a MC sample of size 1000 from Gamma
mc.G <- rgamma(M, shape=a, rate=b)

## Generate a MC sample of size 1000 from Inverse Gamma
mc.IG <- 1 / mc.G

## Generate a MC sample of size 1000 from Square root Inverse Gamma
mc.SIG <- sqrt(1/mc.G)
```

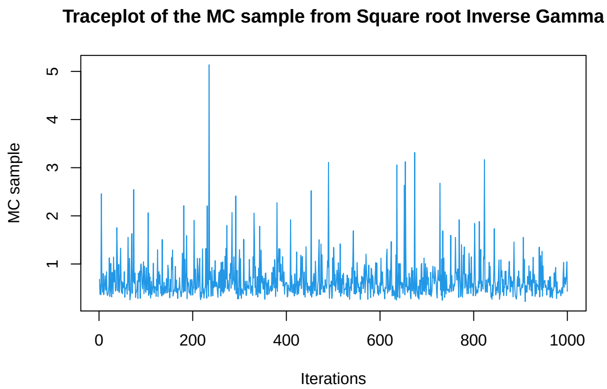
```r
plot(1:M, mc.G, type="l", col=2, xlab="Iterations", ylab="MC sample",
     main="Traceplot of the MC sample from Gamma")
hist(mc.G, breaks=50, freq=FALSE, xlab="X", main="Histogram of MC sample of X")
curve(dgamma(x, a, b), add=TRUE, col=2, lwd=2)

plot(1:M, mc.IG, type="l", col=3, xlab="Iterations", ylab="MC sample",
     main="Traceplot of the MC sample from Inverse Gamma")
hist(mc.IG, breaks=50, freq=FALSE, xlab="Y", main="Histogram of MC sample of Y")
curve(dinvgamma(x, a, b), add=TRUE, col=3, lwd=2)

plot(1:M, mc.SIG, type="l", col=4, xlab="Iterations", ylab="MC sample",
     main="Traceplot of the MC sample from Square root Inverse Gamma")
hist(mc.SIG, breaks=50, freq=FALSE, ylim=c(0, 2), xlab="Z",
     main="Histogram of MC sample of Z")
curve(dsqrtinvgamma(x, a, b), add=TRUE, col=4, lwd=2)
```

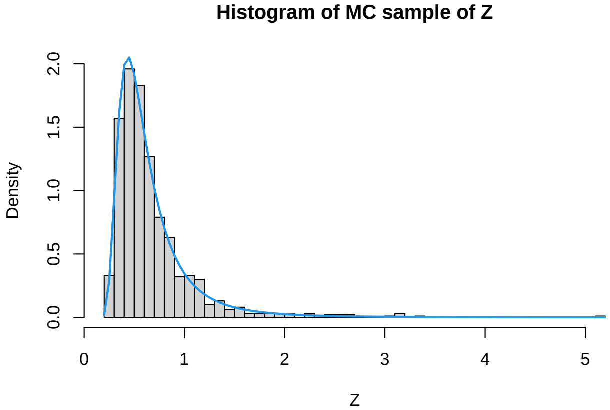













```r
## Gamma
meanG <- mean(mc.G)
medG <- median(mc.G)

## Inverse Gamma
meanIG <- mean(mc.IG)
medIG <- median(mc.IG)

## Square root Inverse Gamma
meanSIG <- mean(mc.SIG)
medSIG <- median(mc.SIG)
```

```r
df <- data.frame(
  c(meanG, meanIG, meanSIG),
  c(medG, medIG, medSIG)
)
colnames(df) <- c("Sample Mean", "Sample Median")
rownames(df) <- c("G", "IG", "SIG")

knitr::kable(df, caption="Summary statistics", align="c")
```

Table 1: Summary statistics

	Sample Mean	Sample Median
G	3.9667004	3.2625686
IG	0.6143637	0.3065072
SIG	0.6672328	0.5536309