

Worksheet1

Foundations of Bayesian Methodology

Group: Goliath

Spring Semester 2022

Exercise 4 (Bayes theorem)

$$\begin{aligned}P[A \mid B, I] &= \frac{P[A, B, I]}{P[B, I]} \\&= \frac{P[B \mid A, I] \cdot P[A, I]}{P[B, I]} \\&= \frac{P[B \mid A, I] \cdot P[A \mid I] \cdot P[I]}{P[B, I]} \\&= \frac{P[B \mid A, I] \cdot P[A \mid I] \cdot P[I]}{P[B \mid I] \cdot P[I]} \\&= \frac{P[B \mid A, I] \cdot P[A \mid I]}{P[B \mid I]}\end{aligned}$$

Exercise 5 (Application of the Bayes theorem)

Given information:

- Sensitivity: $P[T^+ \mid D^+] = 0.96$
- Specificity: $P[T^- \mid D^-] = 0.97$
- Prior: $P[D^+] = 0.002$

From given information, we can derive:

- $P[T^- \mid D^+] = 1 - P[T^+ \mid D^+] = 0.04$
- $P[T^+ \mid D^-] = 1 - P[T^- \mid D^-] = 0.03$
- $P[D^-] = 1 - P[D^+] = 0.998$

$$\begin{aligned}P[D^- \mid T^+] &= \frac{P[D^-, T^+]}{P[T^+]} \\&= \frac{P[T^+ \mid D^-] \cdot P[D^-]}{P[T^+ \mid D^-] \cdot P[D^-] + P[T^+ \mid D^+] \cdot P[D^+]} \\&= \frac{0.03 \times 0.998}{0.03 \times 0.998 + 0.96 \times 0.002} \\&\approx 0.94\end{aligned}$$

This diagnostic test seems reliable at the first glance since the sensitivity is 0.96 and the specificity is 0.97. From a Bayesian perspective, however, we see that the probability that someone is healthy given he or she is tested positive is 0.94. In other words, if someone is tested positive, he or she is actually healthy with a probability of 0.94. This diagnostic test in this sense is not reliable at all.

Exercise 6 (Monte Carlo: random sample vs the true distribution)

```
set.seed(44566) # set seed for reproducibility

#  $X \sim N(160, 20^2)$ 
mu <- 160
sigma <- 20

# Generate a Monte Carlo sample of size 1000
mc.samples <- rnorm(1000, mean = mu, sd = sigma)
```

1

Given that the random variable X follows a normal distribution with $\mu = 160$ and $\sigma = 20$, we can know that:

- Expectation of X : 160
- Standard deviation of X : 20
- Variance of X : 400
- Median of X : 160

```
# Report the (0.025, 0.5, 0.975) quantiles of X
qnorm(c(0.025, 0.5, 0.975), mean = 160, sd = 20)
```

```
## [1] 120.8007 160.0000 199.1993
```

2

```
# Plot the traceplot of the MC sample for X
traceplot(mcmc(mc.samples))
```

3

```
hist(mc.samples, freq = FALSE, breaks = 20) # generate a histogram
lines(density(mc.samples), col = "red", lwd = 2) # add empirical density
lines(seq(100, 220), dnorm(seq(100, 220), mean = 160, sd = 20),
      col = "blue", lwd = 2) # add true density
legend(x = "topright", legend = c("empirical density", "true density"),
      col = c("red", "blue"), lwd = 2) # add legend
```

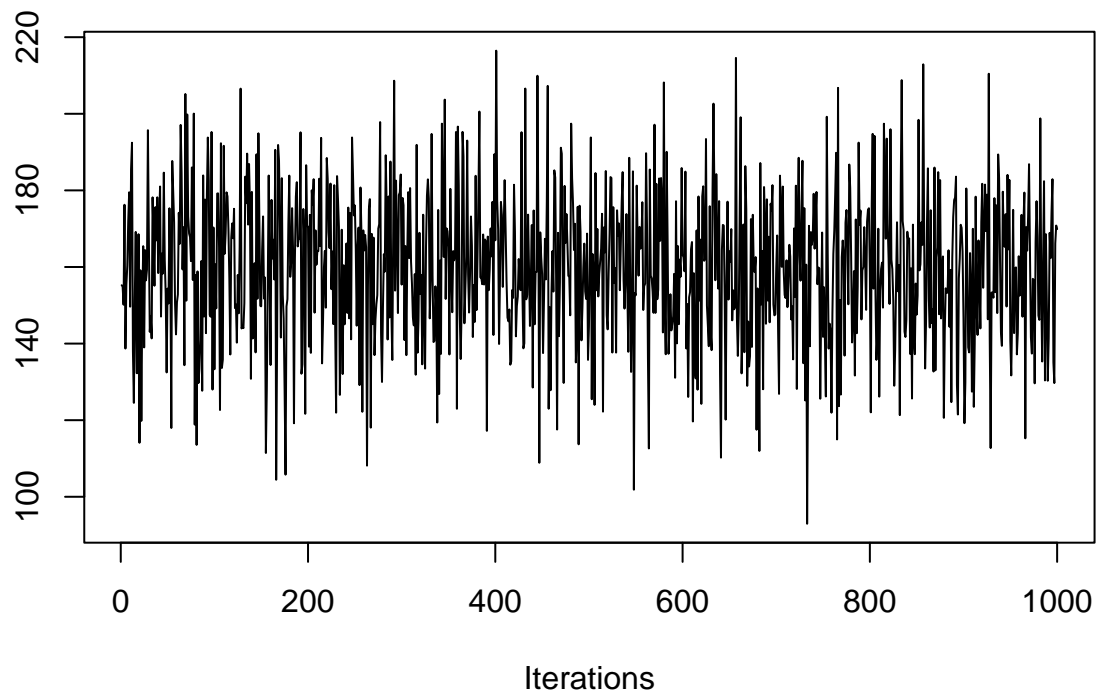


Figure 1: Traceplot of the MC sample for X

4

```
# Summary statistics of MC sample
sample.mean <- mean(mc.samples)
sample.sd <- sd(mc.samples)
sample.var <- var(mc.samples)

cat(sprintf("The sample mean is %.4f
            \nThe sample standard deviation is %.4f
            \nThe sample variance is %.4f\n\n",
            sample.mean, sample.sd, sample.var))
```

```
## The sample mean is 159.6127
##
## The sample standard deviation is 19.8155
##
## The sample variance is 392.6537
```

```
# Sample quantiles
quantile(mc.samples, probs = c(0.025, 0.5, 0.975))
```

```
##      2.5%      50%      97.5%
## 119.6477 159.8035 197.4104
```

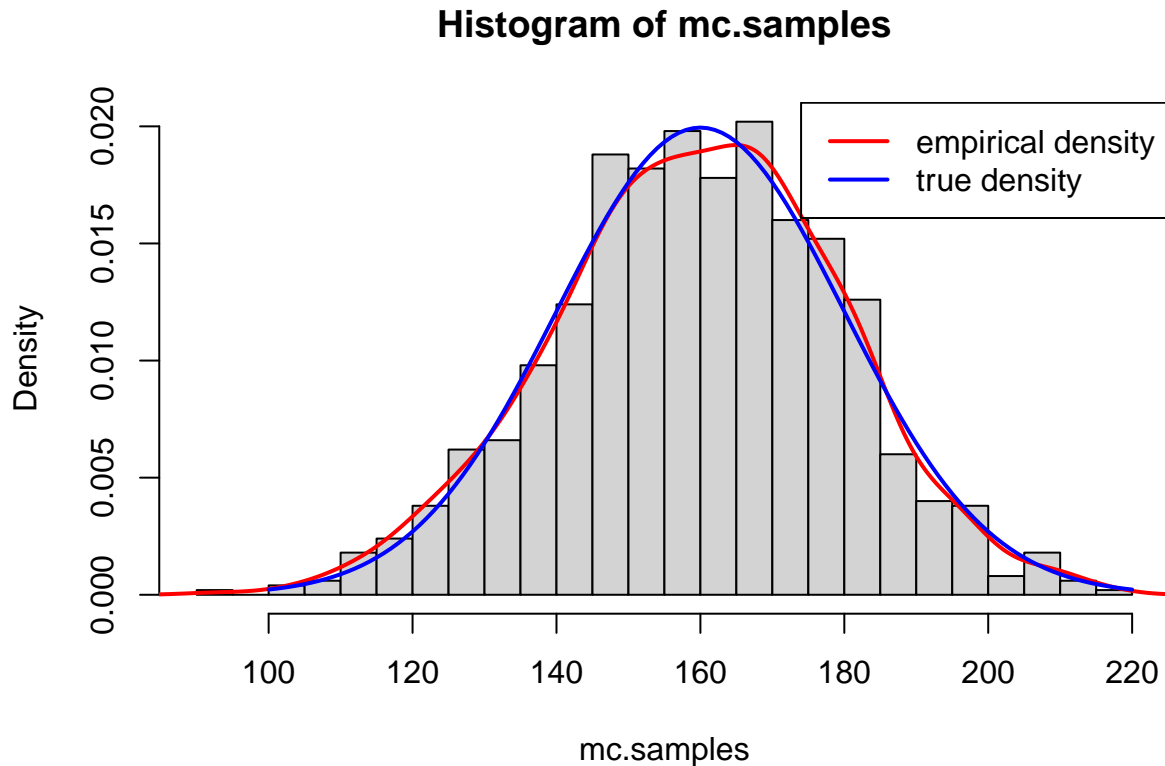


Figure 2: Histogram of the MC sample for X

5

$$P[X > 175] = 1 - P[X \leq 175]$$

$$P[150 < X < 180] = P[X \leq 180] - P[X \leq 150]$$

```
# Compute the empirical cumulative distribution function for MC sample
Fn <- ecdf(mc.samples)
```

```
# Calculate P[X>175] from empirical CDF
1 - Fn(175)
```

```
## [1] 0.225
```

```
# Calculate P[150<X<180] from empirical CDF
Fn(180) - Fn(150)
```

```
## [1] 0.536
```

```
# Calculate P[X>175] from true CDF
pnorm(175, mean = 160, sd = 20, lower.tail = FALSE)
```

```
## [1] 0.2266274
```

```
# Calculate  $P[150 < X < 180]$  from true CDF
```

```
pnorm(180, mean = 160, sd = 20) - pnorm(150, mean = 160, sd = 20)
```

```
## [1] 0.5328072
```