Master solution of group tasks

Exercise 4 (Bayes theorem - 4 points)

Consider the joint probability $P[A, B, I]$

$$P[A, B, I] = P[A \mid B, I]P[B, I] = P[A \mid B, I]P[B \mid I]P[I].$$

Moreover,

$$P[A, B, I] = P[B \mid A, I]P[A, I] = P[B \mid A, I]P[A \mid I]P[I].$$

Thus,

$$P[A \mid B, I] = \frac{P[B \mid A, I]P[A \mid I]}{P[B \mid I]}.$$

Exercise 5 (Application of the Bayes theorem - 4 points)

We know:

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

Apply the Bayes theorem

$$P[D^- \mid T^+] = \frac{P[T^+ \mid D^-]P[D^-]}{P[T^+]} \tag{1}$$

By the law of total probabilities

$$P[D^- \mid T^+] = \frac{P[T^+ \mid D^-]P[D^-]}{P[T^+]} = \frac{P[T^+ \mid D^-]P[D^-]}{P[T^+ \mid D^-]P[D^-] + P[T^+ \mid D^+]P[D^+]}. \tag{2}$$

Remember that $P[T^+ \mid D^-] = 1 - P[T^- \mid D^-]$.

Therefore,

$$P[D^- \mid T^+] = \frac{(1-0.97)(1-0.002)}{(1-0.97)(1-0.002)+0.960.002} = 0.9397. \tag{3}$$

Given a positive test, the probability of being healthy is very high, because the prevalence is very small.

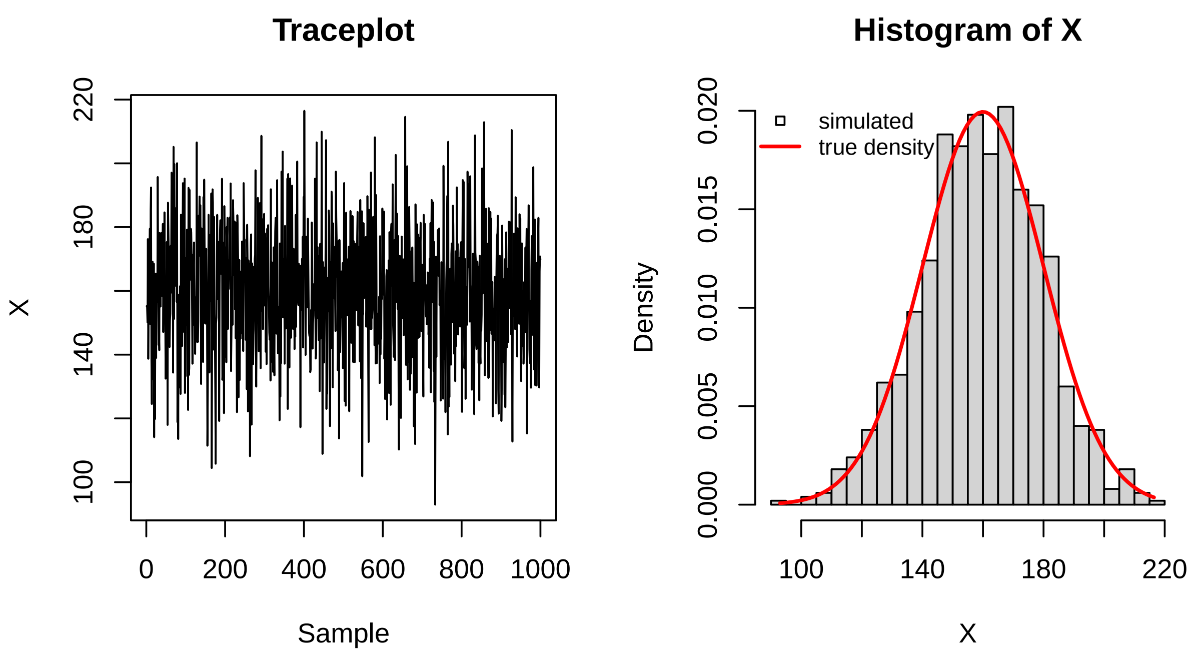
Exercise 6 (Monte Carlo: random sample vs the true distribution - 8 points)

Let the random variable $X \sim \mathcal{N}(\mu, \sigma^2)$, with $\mu = 160$ and $\sigma = 20$. Generate Monte Carlo sample X of size $M = 1000$.

```
M <- 1000
mu <- 160
sigma <- 20
set.seed( 44566 )
X <- rnorm(M, mu, sigma)
```

```
#true mean mu
mu

## [1] 160

#true standard deviation
sigma

## [1] 20

#true variance
sigma^2

## [1] 400

#true median
(true_median <- qnorm(0.5, mu, sigma))

## [1] 160

#true quantiles
(true_q0.025 <- qnorm(0.025, mu, sigma))

## [1] 120.8007

(true_q0.975 <- qnorm(0.975, mu, sigma))

## [1] 199.1993
```

```r
par(mfrow = c(1, 2), cex = 0.9, pty = "s")
plot(seq_along(X), X, type = "l", xlab = "Sample", ylab = "X",
     main = "Traceplot", cex = 0.5)

hist(X, freq = FALSE, ylim = c(0, 0.02), breaks = 20)
#lines(density(X))

curve(dnorm(x, mu, sigma), min(X), max(X), add = TRUE, col = "red", lwd = 2)
legend("topleft", c( "simulated", "true density"), pch = c(22, NA), lwd = c(NA, 2),
       bty = "n", col = c("black", "red"), cex = 0.8)
```



```r
#sample mean
mean(X)

## [1] 159.6127

# standard deviation
sd(X)

## [1] 19.81549

#variance
var(X)

## [1] 392.6537

#quantiles
quantile(X, c(0.025, 0.5, 0.975))

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

summary(X)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   92.95  146.58  159.80  159.61  173.28  216.48

#boxplot(X)
```

True probability that $P[X > 175]$

```r
1 - pnorm(175, mu, sigma)

## [1] 0.2266274
```

and the estimate of probability that $P[X > 175]$ can be calculated as

```r
mean( X > 175 )

## [1] 0.225
```

True probability that $P[150 < X < 180]$

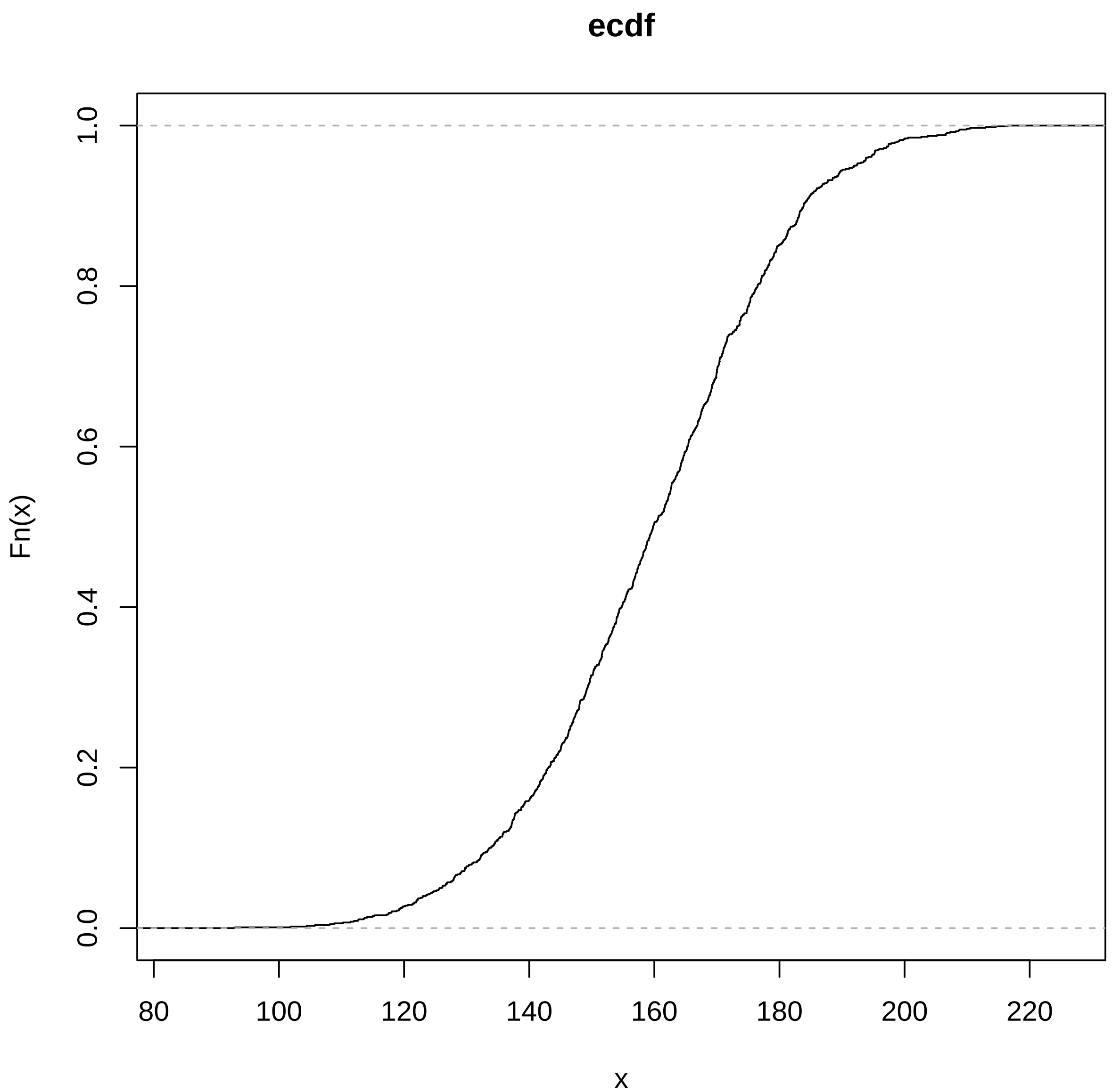
```r
pnorm(180, mu, sigma) - pnorm(150, mu, sigma)

## [1] 0.5328072
```

and the estimate of the probability that $P[150 < X < 180]$ is

```
mean( X < 180 & X > 150 )

## [1] 0.536
```

Estimates of the probabilities $P[X > 175]$ and $P[150 < X < 180]$ can also be computed by the empirical cumulative distribution function.

```
P<- ecdf(X)
plot(P, main = "ecdf") # empirical cumulative distribution function
```



```
print(1 - P(175))       # empirical values

## [1] 0.225

print(P(180)- P(150))

## [1] 0.536
```

We can conclude that estimates based on a random sample of size $M = 1000$ are close to the true probabilities.

Exercise 7 (Bayes Factor - 10 points)

a) Let $Y \mid \mu \sim \mathcal{N}(\mu, \kappa^{-1})$ with known κ^{-1}.

- $H_0 : \mu = \mu_0$
- H_1 : suppose that μ is known with prior distribution $\mu \sim \mathcal{N}(\nu, \lambda^{-1})$

$$f(y \mid H_1) = \int f(y \mid \theta) f(\theta \mid H_1) d\theta \tag{4}$$

First approach

$$\begin{aligned}
f(y \mid H_1) &= \int \sqrt{\frac{\kappa}{2\pi}} \exp\left(-\frac{\kappa}{2}(y-\mu)^2\right) \sqrt{\frac{\lambda}{2\pi}} \exp\left(-\frac{\lambda}{2}(\mu-\nu)^2\right) d\mu \\
&= \int \sqrt{\frac{\kappa\lambda}{4\pi^2}} \exp\left(-\frac{1}{2}\left(\kappa y^2 + \kappa\mu^2 - 2\kappa\mu y + \lambda\mu^2 + \lambda\nu^2 - 2\mu\nu\lambda\right)\right) d\mu \\
&= \int \sqrt{\frac{\kappa\lambda}{4\pi^2}} \exp\left(-\frac{1}{2}\left((\kappa+\lambda)(\mu^2 - 2\mu\frac{\kappa y + \lambda\nu}{\kappa+\lambda} + \left(\frac{\kappa y + \lambda\nu}{\kappa+\lambda}\right)^2 - \left(\frac{\kappa y + \lambda\nu}{\kappa+\lambda}\right)^2) + \kappa y^2 + \lambda\nu^2\right)\right) d\mu \\
&= \int \sqrt{\frac{\kappa\lambda}{2\pi}} \sqrt{\frac{\kappa+\lambda}{2\pi}} \frac{1}{\sqrt{\kappa+\lambda}} \exp\left(-\frac{1}{2}(\kappa+\lambda)\left(\mu - \frac{\kappa y + \lambda\nu}{\kappa+\lambda}\right)^2\right) \\
&\qquad \exp\left(\frac{(\kappa y + \lambda\nu)^2 - (\kappa y^2 + \lambda\nu^2)(\kappa+\lambda)}{2(\kappa+\lambda)}\right) d\mu \\
&= \sqrt{\frac{\kappa\lambda}{2\pi(\kappa+\lambda)}} \exp\left(\frac{-\kappa\lambda(y-\nu)^2}{2(\kappa+\lambda)}\right) \underbrace{\int \sqrt{\frac{\kappa+\lambda}{2\pi}} \exp\left(-\frac{1}{2}(\kappa+\lambda)\left(\mu - \frac{\kappa y + \lambda\nu}{\kappa+\lambda}\right)^2\right) d\mu}_{=1}
\end{aligned} \tag{5}$$

Second approach, consider the distribution of $Y \mid H_1$

$$Y \mid H_1 \sim \mathcal{N}(0, \kappa^{-1}) + \mathcal{N}(\nu, \lambda^{-1}) = X_1 + X_2.$$

X_1, X_2 are both Gaussian and independent of each other, then $X_1 + X_2$ is also Gaussian.

$$\mathrm{E}(Y \mid H_1) = \mathrm{E}(X_1 + X_2) = \mathrm{E}(X_1) + \mathrm{E}(X_2) = 0 + \nu = \nu,$$

$$\mathrm{Var}(Y \mid H_1) = \mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = \kappa^{-1} + \lambda^{-1} = \frac{1}{\kappa} + \frac{1}{\lambda} = \frac{\kappa + \lambda}{\kappa\lambda},$$

Therefore,

$$Y \mid H_1 \sim \mathcal{N}\left(\nu, \frac{\kappa + \lambda}{\kappa\lambda}\right),$$

Finally,

$$f(y \mid H_1) = \frac{1}{\sqrt{2\pi \frac{\kappa+\lambda}{\kappa\lambda}}} \exp\left\{-\frac{(y-\nu)^2}{2\frac{\kappa+\lambda}{\kappa\lambda}}\right\} = \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\kappa\lambda}{\kappa+\lambda}} \exp\left\{-\frac{\kappa\lambda(y-\nu)^2}{2(\kappa+\lambda)}\right\}.$$

b) Bayes factor $\mathrm{BF}_{01}(y)$ can be determined analytically

$$\mathrm{BF}_{01}(y) = \frac{f(y \mid H_0)}{f(y \mid H_1)}$$

$$f(y \mid H_0) = \sqrt{\frac{\kappa}{2\pi}} \exp\left(-\frac{\kappa}{2}(y - \mu_0)^2\right)$$

$$f(y \mid H_1) = \frac{1}{\sqrt{2\pi}}\sqrt{\frac{\kappa\lambda}{\kappa+\lambda}} \exp\left(-\frac{\kappa\lambda}{2(\kappa+\lambda)}(y-\nu)^2\right)$$

Hence,

$$\mathrm{BF}_{01}(y) = \sqrt{\frac{\kappa+\lambda}{\lambda}} \exp\left(\frac{\kappa\lambda}{2(\kappa+\lambda)}(y-\nu)^2 - \frac{\kappa}{2}(y-\mu_0)^2\right) \tag{6}$$

c)

$$\mathrm{BF}_{01}(y) = \sqrt{\frac{\kappa+\lambda}{\lambda}} \exp\left(\frac{\kappa\lambda}{2(\kappa+\lambda)}(y-\nu)^2 - \underbrace{\frac{\kappa}{2}(y-\mu_0)^2}_{\text{indep. of } \lambda}\right),$$

$$\frac{\kappa\lambda}{\kappa+\lambda} = \frac{\kappa}{\frac{\kappa}{\lambda}+1} \xrightarrow{\lambda\to 0} 0$$

and

$$\frac{\kappa+\lambda}{\lambda} = \frac{\kappa}{\lambda} + 1 \xrightarrow{\lambda\to 0} \infty,$$

then

$$\mathrm{BF}_{01}(y) \xrightarrow{\lambda\to 0} \infty.$$

d) Compute $P[H_0 \mid y]$ if $P[H_0] = P[H_1] = 0.5$, $\mu_0 = 0$, $\kappa = 1$, $\nu = 2$, $\lambda = 1/2$ and $y = 1$.

$$P[H_0 \mid y] = \frac{\mathrm{BF}_{01}(y)\frac{P[H_0]}{P[H_1]}}{1 + \mathrm{BF}_{01}(y)\frac{P[H_0]}{P[H_1]}} = \frac{\mathrm{BF}_{01}(y)}{1 + \mathrm{BF}_{01}(y)},$$

because $P[H_0] = P[H_1] = 0.5$. Given (6) we get

$$P[H_0 \mid y] = \frac{\sqrt{3}\exp(-\frac{1}{3})}{1+\sqrt{3}\exp(-\frac{1}{3})} = 0.55.$$

In our example $\mathrm{BF}_{01}(y) = 1.24$. This means that there is slight evidence in favor of H_0 against H_1. The probability that H_0 is correct given the data is 0.55. Prior $P[H_0] = 0.5$ has increased to posterior $P[H_0 \mid y] = 0.55$ after seeing one observation $y = 1$.

Exercise 8 (Calibration of p-values: pCalibrate - 4 points)

```
library(pCalibrate)
#?twoby2Calibrate
# re-formatted data of Table 1
table_2by2 <- matrix(c(9, 5, 14, 1), ncol = 2,
        dimnames = list("Intervention group" = c("Secukinumab", "Placebo"),
        "Result" = c("Non-Responders", "Responders")))
table_2by2

##                    Result
## Intervention group Non-Responders Responders
##        Secukinumab              9         14
##        Placebo                  5          1

(BF_pv <- twoby2Calibrate(table_2by2, type = "two.sided", alternative = "simple",
                direction = NULL, transform.bf = "id"))

## $minBF
## [1] 0.2937708
##
## $p.value
##       p.pb       p.ce       p.bl      p.mid      p.lie
## 0.08007663 0.13908046 0.08007663 0.07586207 0.06580151
```

Bayes factor is decreasing the prior odds because the minimum Bayes factor is smaller than 1 (0.29).

```
#?formatBF
formatBF(BF_pv$minBF)

## [1] "1/3.4"
```

```
#?BF2pp
(BF2pp( BF = BF_pv$minBF, prior.prob = 0.5 ))

## [1] 0.2270656
```

Prior $P[H_0] = 0.5$ has decreased to posterior $P[H_0 \mid y] = 0.23$ given the data.