Solutions to Assignment 1

Rongfei Jin

January 16, 2025

Exercise 1

Solution. The minimization is found by differentiation,

$$d(\sum_{j=1}^{n}(X_j-\theta)^2) = \sum_{j=1}^{n} -2X_j + 2\theta = -2(\sum_{j=1}^{n} X_j - \sum_{j=1}^{n}\theta)$$

We check that the point is indeed a minima by deriving the 2nd derivative

$$d^2(\sum_{j=1}^{n}(X_j-\theta)^2) = 2n > 0$$

We set the gradient to 0 and get the θ that minimize the expression

$$\sum_{j=1}^{n}\theta = \sum_{j=1}^{n} X_j$$
$$\hat{\theta} = \frac{1}{n}\sum_{j=1}^{n} X_j$$

For part 2, we note that

$$|X_j - \eta| = \begin{cases} X_j - \eta & \text{if } X_j >= \eta \\ -X_j + \eta & \text{if } X_j < \eta \end{cases}$$

Now we define the order statistics by sorting the realization value of $X_1, X_2, \ldots, X_n$ in increasing order as such

$$X_{(1)}, X_{(2)}, \ldots, X_{(n)}$$

where $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$

When n is even Let $k \in \mathbb{Z}^+$ such that $X_{(k)} \leq \eta \leq X_{(k+1)}$ where $0 < k < n$

$$\begin{aligned}
\arg\min_{\eta\in\mathbb{R}} \sum_{j=1}^{n} |X_j - \eta| &= \arg\min_{\eta\in\mathbb{R}, k\in\mathbb{Z}^+} (\sum_{j=1}^{k}(-X_{(j)}+\eta) + \sum_{j=k+1}^{n}(X_{(j)}-\eta)) \\
&= \arg\min_{\eta\in\mathbb{R}, k\in\mathbb{Z}^+} \sum_{j=1}^{k}(-X_{(j)}) + \sum_{j=k+1}^{n}(X_{(j)}) + (k-(n-(k+1)+1))\eta \\
&= \arg\min_{\eta\in\mathbb{R}, k\in\mathbb{Z}^+} \sum_{j=1}^{k}(-X_{(j)}) + \sum_{j=k+1}^{n}(X_{(j)}) + (2k-n)\eta \\
&= \arg\min_{\eta\in\mathbb{R}, k\in\mathbb{Z}^+} (2k-n)\eta
\end{aligned}$$

Now we differentiate and find the minima

$$\frac{\partial}{\partial\eta}((2k-n)\eta) = 2k - n$$

$$k = n/2$$

if n is even, then we can let $k = n/2$ However, if n is odd, then the minima $n/2$ is not defined, therefore, we must find the next best value.

Since n is odd, we let $n = 2m - 1$ where m is any positive integer, if $k > m$ then $2k - n > 2m - (2m - 1) = 1$, if $k \leq m - 1$ then $2k - n \leq 2m - 2 - (2m - 1) = -1$, therefore, the best k must be in the range $m - 1 < k \leq m$, and the only possible value is $k = m$ since m is only integer in this range

we found that the change of the expression with respect to η is constant, which means the order which η locates in the sequenence A_j matters rather that the value of η.

The expression $\sum_{j=1}^{n} |A_j - \eta|$ is minimized when $k = n/2$ if n is even and $k = (n+1)/2$ if n is odd, which means $\hat{\eta} =$ median of A_j ∎

Exercise 2

Solution.

$$
\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathrm{E}((X - E(X))(Y - E(Y)) \\
&= E(XY - XE(Y) - XE(Y) + E(X)E(Y)) \\
&= E(XY) - E(XE(Y)) - E(XE(Y)) + E(E(X)E(Y)) && \text{(Expectation Linearity)[3]} \\
&= E(XY) - E(X)E(Y)
\end{aligned}
$$

■

Exercise 3

Solution.

$$L(\lambda|x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$$

$$\begin{aligned} l(\lambda|x_1, x_2, \ldots, x_n) &= \log(\prod_{i=1}^{n} \lambda e^{-\lambda x_i}) \\ &= \sum_{i=1}^{n} \log(\lambda) + (-\lambda x_i) \\ &= n\log(\lambda) - \lambda \sum_{i=1}^{n} x_i \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \lambda}(n\log(\lambda) - \lambda \sum_{i=1}^{n} x_i) &= \frac{n}{\lambda} - \sum_{i=1}^{n} x_i \\ \frac{\partial}{\partial^2 \lambda}(n\log(\lambda) - \lambda \sum_{i=1}^{n} x_i) &= -\frac{n}{\lambda^2} < 0 && \text{(The point is the maximum)} \end{aligned}$$

$$\begin{aligned} \frac{\partial}{\partial \lambda} &= 0 \\ \frac{n}{\lambda} - \sum_{i=1}^{n} x_i &= 0 \end{aligned}$$

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}$$

■

Exercise 4

Solution.

$$L(\boldsymbol{\mu};\boldsymbol{\Sigma}|\boldsymbol{x_1},\ldots,\boldsymbol{x_n}) = \prod_{i=1}^{n}\frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}}\exp\{-\frac{1}{2}(\boldsymbol{x}_i-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x_i}-\boldsymbol{\mu})\}$$

$$l(\boldsymbol{\mu};\boldsymbol{\Sigma}|\boldsymbol{x_1},\ldots,\boldsymbol{x_n}) = \log(\prod_{i=1}^{n}\frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}}\exp\{-\frac{1}{2}(\boldsymbol{x}_i-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})\})$$

$$l(\boldsymbol{\mu};\boldsymbol{\Sigma}|\boldsymbol{x_1},\ldots,\boldsymbol{x_n}) = n\log(\frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}}) - \frac{1}{2}\sum_{i=1}^{n}((\boldsymbol{x}_i-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}))$$

$$l(\boldsymbol{\mu};\boldsymbol{\Sigma}|\boldsymbol{x_1},\ldots,\boldsymbol{x_n}) = n\log(\frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}}) - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x_i}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_i - 2\boldsymbol{x}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})$$

$$(\boldsymbol{x}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \boldsymbol{\mu}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}, \text{ are scalars, so they're the same})$$

$$l(\boldsymbol{\mu};\boldsymbol{\Sigma}|\boldsymbol{x_1},\ldots,\boldsymbol{x_n}) = n\log(\frac{1}{2\pi|\boldsymbol{\Sigma}|^{1/2}}) + \sum_{i=1}^{n}(-\frac{1}{2}\boldsymbol{x_i}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_i + \boldsymbol{x}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})$$

$$l(\boldsymbol{\mu};\boldsymbol{\Sigma}|\boldsymbol{x_1},\ldots,\boldsymbol{x_n}) = -n\log(2\pi) - \frac{n}{2}\log(|\boldsymbol{\Sigma}|) + \sum_{i=1}^{n}(-\frac{1}{2}\boldsymbol{x_i}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}_i + \boldsymbol{x}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})$$

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\mu}} &= \sum_{i=1}^{n}\boldsymbol{x}_i^T\boldsymbol{\Sigma}^{-1} - \sum\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\\ &= \sum_{i=1}^{n}\boldsymbol{x}_i^T\boldsymbol{\Sigma}^{-1} - n\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1}\end{aligned}$$

$$\begin{aligned}n\boldsymbol{\mu}^T\boldsymbol{\Sigma}^{-1} &= \sum_{i=1}^{n}\boldsymbol{x}_i^T\boldsymbol{\Sigma}^{-1} && \text{(Set gradient to 0)}\\ n\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} &= \boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}\boldsymbol{x}_i && \text{(Note that co-variance is symmetric)}\\ \boldsymbol{\mu} &= \overline{\boldsymbol{x}_i}\\ \hat{\boldsymbol{\mu}} &= \begin{pmatrix}\overline{x_{1i}}\\ \overline{x_{2i}}\end{pmatrix}\\ \begin{pmatrix}\hat{\mu}_1\\ \hat{\mu}_2\end{pmatrix} &= \begin{pmatrix}\frac{1}{n}\sum_{i=1}^{n}x_{i1}\\ \frac{1}{n}\sum_{i=1}^{n}x_{i2}\end{pmatrix}\end{aligned}$$

We break down the partial differentiation of Σ^{-1} into 2 parts,

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\Sigma}^{-1}}(-\frac{n}{2}\log(|\boldsymbol{\Sigma}|)) &= \frac{\partial}{\partial\boldsymbol{\Sigma}^{-1}}(\frac{n}{2}\log(|\boldsymbol{\Sigma}^{-1}|))\\ &= \frac{n}{2}\frac{1}{|\boldsymbol{\Sigma}^{-1}|}|\boldsymbol{\Sigma}^{-1}|(\boldsymbol{\Sigma})^T && \text{Determinant Derivative[1]}\\ &= \frac{n}{2}\boldsymbol{\Sigma}\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\Sigma}}(-\frac{1}{2}\sum_{i=1}^{n}((\boldsymbol{x}_i-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}))) &= -\frac{1}{2}\frac{\partial}{\partial\boldsymbol{\Sigma}}\sum_{i=1}^{n}\mathrm{Tr}((\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1})\\ &= -\frac{1}{2}\sum_{i=1}^{n}\mathrm{Tr}((\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^T)\\ &= -\frac{1}{2}\sum_{i=1}^{n}((\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})^T)\end{aligned}$$

We then add part 1 and 2 together and set this gradient to 0

$$\frac{n}{2}\boldsymbol{\Sigma} - \frac{1}{2}\sum_{i=1}^{n}((\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T) = 0$$

$$\begin{aligned}
\boldsymbol{\Sigma} &= \frac{\sum_{i=1}^{n}((\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T)}{n} \\
&= \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i\boldsymbol{x}_i^T - \boldsymbol{x}_i\boldsymbol{\mu}^T - \boldsymbol{\mu}\boldsymbol{x}_i^T + \boldsymbol{\mu}\boldsymbol{\mu}^T) \\
&= \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix}\begin{pmatrix} x_{i1} & x_{i2} \end{pmatrix} - \begin{pmatrix} x_{i1} \\ x_{i2} \end{pmatrix}\begin{pmatrix} \mu_1 & \mu_2 \end{pmatrix} - \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}\begin{pmatrix} x_{i1} & x_{i2} \end{pmatrix} + \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}\begin{pmatrix} \mu_1 & \mu_{i2} \end{pmatrix} \\
&= \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} x_{i1}^2 & x_{i1}x_{i2} \\ x_{i1}x_{i2} & x_{i2}^2 \end{pmatrix} - \begin{pmatrix} x_{i1}\mu_1 & x_{i1}\mu_2 \\ x_{i2}\mu_1 & x_{i2}\mu_2 \end{pmatrix} - \begin{pmatrix} x_{i1}\mu_1 & x_{i2}\mu_1 \\ x_{i1}\mu_2 & x_{i2}\mu_2 \end{pmatrix} + \begin{pmatrix} \mu_1^2 & \mu_1\mu_2 \\ \mu_1\mu_{i2} & \mu_2^2 \end{pmatrix} \\
&= \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} x_{1i}^2 - 2x_{i1}\mu_1 + \mu_1^2 & x_{i1}x_{i2} - x_{i1}\mu_2 - x_{i2}\mu_1 + \mu_1\mu_2 \\ x_{i1}x_{i2} - x_{i1}\mu_2 - x_{i2}\mu_1 + \mu_1\mu_2 & x_{2i}^2 - 2x_{i1}\mu_2 + \mu_2^2 \end{pmatrix} \\
&= \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} (x_{i1} - \mu_1)^2 & (x_{i1} - \mu_1)(x_{i2} - \mu_2) \\ (x_{i1} - \mu_1)(x_{i2} - \mu_2) & (x_{i2} - \mu_2)^2 \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n}(x_{i1} - \mu_1)^2 & \frac{1}{n}\sum_{i=1}^{n}(x_{i1} - \mu_1)(x_{i2} - \mu_2) \\ \frac{1}{n}\sum_{i=1}^{n}(x_{i1} - \mu_1)(x_{i2} - \mu_2) & \frac{1}{n}\sum_{i=1}^{n}(x_{i2} - \mu_2)^2 \end{pmatrix}
\end{aligned}$$

$$\frac{1}{n}\sum_{i=1}^{n}(x_{i1} - \mu_1)(x_{i2} - \mu_2) = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_{i1} - \mu_1)(x_{i2} - \mu_2)}{\sqrt{\sum_{i=1}^{n}(x_{i1} - \mu_1)^2 \cdot \sum_{i=1}^{n}(x_{i2} - \mu_2)^2}}\sqrt{\sum_{i=1}^{n}(x_{i1} - \mu_1)^2 \cdot \sum_{i=1}^{n}(x_{i2} - \mu_2)^2}$$

Now it is clear that we have the following estimation:

$$\hat{\sigma}_1^2 = \frac{1}{n}\sum_{i=1}^{n}(x_{i1} - \mu_1)^2$$

$$\hat{\sigma}_2^2 = \frac{1}{n}\sum_{i=1}^{n}(x_{i2} - \mu_2)^2$$

$$\hat{\rho} = \frac{1}{\sqrt{\hat{\sigma}_1\hat{\sigma}_2}}\frac{1}{n}\sum_{i=1}^{n}(x_{i1} - \mu_1)(x_{i2} - \mu_2)$$

∎

Exercise 5

Solution.

(i)
```
generate_n_gamma <- function(n, alpha=1, beta=1) {
  delta = 1/beta

  gammas <- c()
  for (i in 1:n) {
    us <- sample.int(.Machine$integer.max, size = alpha, replace=TRUE)
    us <- us / .Machine$integer.max
    exp <- -log(us) / delta
    gammas <- c(gammas,sum(exp))
  }
  return (gammas)
}
```

(ii)

$$L(\alpha,\beta|x_1,x_2,\dots,x_n)=\prod_{i=1}^{n}\frac{1}{\Gamma(\alpha)\beta^{\alpha}}x^{\alpha-1}e^{-\frac{x}{\beta}}$$
$$l(\alpha,\beta|x_1,x_2,\dots,x_n)=\log(\prod_{i=1}^{n}\frac{1}{\Gamma(\alpha)\beta^{\alpha}}x^{\alpha-1}e^{-\frac{x}{\beta}})$$
$$l(\alpha,\beta|x_1,x_2,\dots,x_n)=\sum_{i=1}^{n}-\log(\Gamma(\alpha))-a\log\beta+(\alpha-1)\log(x_i)-\frac{x_i}{\beta}$$
$$l(\alpha,\beta|x_1,x_2,\dots,x_n)=-n\log(\Gamma(\alpha))-n\cdot a\log\beta+(\alpha-1)\sum_{i=1}^{n}\log(x_i)-\frac{1}{\beta}\sum_{i=1}^{n}x_i$$

$$\frac{\partial}{\partial\beta}(l(\alpha,\beta|x_1,x_2,\dots,x_n))=-\frac{n\alpha}{\beta}+\frac{1}{\beta^2}\sum_{i=1}^{n}x_i$$

$$\frac{\partial}{\partial\beta}(l(\alpha,\beta|x_1,x_2,\dots,x_n))=0$$
$$-\frac{n\alpha}{\beta}+\frac{1}{\beta^2}\sum_{i=1}^{n}x_i=0$$

$$\hat{\beta}=\frac{\sum_{i=1}^{n}x_i}{na}=\frac{\bar{x}}{a}$$

$$\begin{aligned}l(\alpha,\beta|x_1,x_2,\dots,x_n)&=-n\log(\Gamma(\alpha))-n\cdot a(\log(\bar{x})-\log(\alpha))+(\alpha-1)\sum_{i=1}^{n}\log(x_i)-\frac{a}{\bar{x}}\sum_{i=1}^{n}x_i\\&=-n\log(\Gamma(\alpha))-na\log(\bar{x})+na\log(\alpha)+\alpha\sum_{i=1}^{n}\log(x_i)-\sum_{i=1}^{n}\log(x_i)-an\end{aligned}$$

Note that we define $\psi(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$

$$\begin{aligned}
\frac{\partial}{\partial \alpha} l(\alpha, \beta | x_1, x_2, \ldots, x_n) &= -n \frac{1}{\Gamma(\alpha)} \Gamma(\alpha)\psi(\alpha) + -n\log(\bar{x}) + n(\log(\alpha) + 1) + \sum_{i=1}^{n} \log(x_i) - n \\
&= -n\psi(\alpha) + -n\log(\bar{x}) + n\log(\alpha) + \sum_{i=1}^{n} \log(x_i) \\
\psi(\alpha) &= -\log(\bar{x}) + \log(\alpha) + \frac{1}{n}\sum_{i=1}^{n} \log(x_i)
\end{aligned}$$

(iii)
```
d <- function(f, x, h = 1e-7) {
  (f(x + h) - f(x - h)) / (2*h)
}

newton <- function(f, x0, tol = 1e-9, maxiter = 100) {
  x <- x0
  for (i in seq_len(maxiter)) {
    fx <- f(x)
    dfx <- d(f, x)

    if (abs(dfx) < 1e-15) {
      stop("Numerical derivative too small. Unable to proceed")
    }

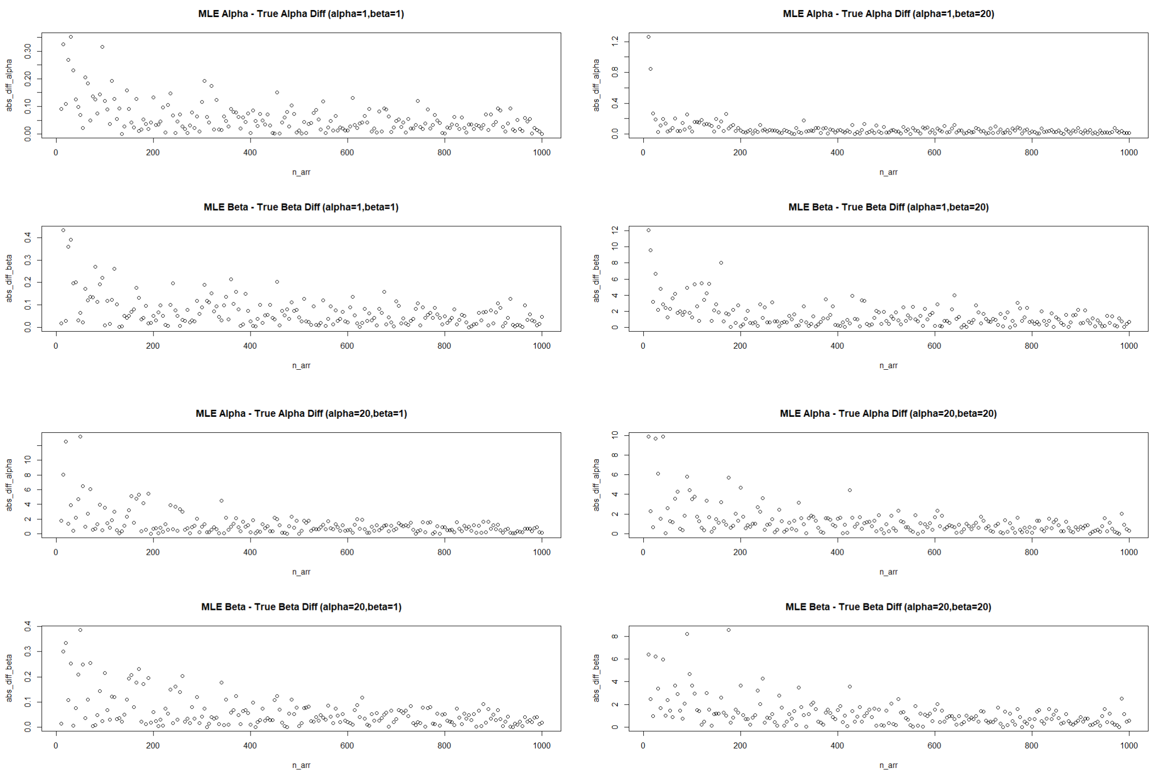
    x_new <- x - fx / dfx

    if (abs(x_new - x) < tol) {
      return(x_new)
    }
    x <- x_new
  }
  stop("Max iterations reached without convergence")
}

mle_gamma <- function(xs){
  n <- length(xs)
  x_bar <- mean(xs)
  log_x_bar <- mean(log(xs))
  f <- function(alpha) log(alpha) - digamma(alpha) - (log(x_bar) - log_x_bar)
  a <- newton(f, 1)
  return (c(a, x_bar/a))
}
```

(iv)
```
n_arr <- seq(10,2000,10)
true_alpha <- 10
true_beta <- 5
abs_diff_alpha = c()
abs_diff_beta = c()
for (n in n_arr) {
    x <- generate_n_gamma(n, true_alpha, true_beta)
    hats <- mle_gamma(x)
    abs_diff_alpha <- c(abs_diff_alpha, abs(hats[1]-true_alpha))
    abs_diff_beta <- c(abs_diff_beta,abs(hats[2]-true_beta))
}
```

Based on the simulations, I hypothesized function: $|\hat{\beta} - \beta| = \frac{\beta}{n-\beta} + \frac{\beta}{n}$ and $|\hat{\alpha} - \alpha| = \frac{\alpha}{n} + \frac{\beta}{n}$


MLE Alpha - True Alpha Diff (alpha=1,beta=1)
MLE Alpha - True Alpha Diff (alpha=1,beta=20)
MLE Beta - True Beta Diff (alpha=1,beta=1)
MLE Beta - True Beta Diff (alpha=1,beta=20)
MLE Alpha - True Alpha Diff (alpha=20,beta=1)
MLE Alpha - True Alpha Diff (alpha=20,beta=20)
MLE Beta - True Beta Diff (alpha=20,beta=1)
MLE Beta - True Beta Diff (alpha=20,beta=20)
abs_diff_alpha
abs_diff_beta
n_arr

Exercise 6

Solution. First we expand the expression

$$\begin{aligned}\sum_{i=1}^{n}(y_i - [\beta_0 + \beta_1 x_i])^2 &= \sum_{i=1}^{n}(y_i^2 - 2(\beta_0 + \beta_1 x_i)y_i + (\beta_0^2 + 2\beta_0\beta_1 x_i + \beta_1^2 x_i^2)) \\ &= \sum_{i=1}^{n}(y_i^2 - 2\beta_0 y_i - 2\beta_1 x_i y_i + \beta_0^2 + 2\beta_0\beta_1 x_i + \beta_1^2 x_i^2)\end{aligned}$$

The minimization is found by taking partial derivative and set it to 0

$$\frac{\partial}{\partial\beta_1} = -2\sum_{i=1}^{n} x_i y_i + 2\beta_0 \sum_{i=1}^{n} x_i + 2\beta_1 \sum_{i=1}^{n} x_i^2 = 0 \tag{1}$$

$$\sum_{i=1}^{n} x_i y_i - \beta_0 \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} x_i^2 = 0 \tag{2}$$

$$\frac{\partial}{\partial\beta_0} = -2\sum_{i=1}^{n} y_i + 2n\beta_0 + 2\beta_1 \sum_{i=1}^{n} x_i = 0$$

$$\beta_0 = \frac{\sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i}{n} = \bar{y} - \beta_1\bar{x}$$

Now we solve this system of equation by substituting β_0 into (2)

$$\begin{aligned}\sum_{i=1}^{n} x_i y_i - (\frac{\sum_{i=1}^{n} y_i - \beta_1 \sum_{i=1}^{n} x_i}{n})\sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} x_i^2 = 0 \\ \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n} + \frac{\beta_1(\sum_{i=1}^{n} x_i)^2}{n} - \beta_1 \sum_{i=1}^{n} x_i^2 = 0 \\ \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n} + \beta_1(\frac{(\sum_{i=1}^{n} x_i)^2}{n} - \sum_{i=1}^{n} x_i^2) = 0\end{aligned}$$

$$\beta_1 = \frac{-\frac{1}{n}\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} x_i y_i}{-\frac{1}{n}(\sum_{i=1}^{n} x_i)^2 + \sum_{i=1}^{n} x_i^2}$$

Let r be the correlation between X and Y, and r is defined as the following [2]

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

s_x are defined as

$$s_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

s_y are defined as

$$s_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}}$$

Let $S_x = \sum_{i=1}^{n} x_i$ Let $S_y = \sum_{i=1}^{n} y_i$ Let $S_{xy} = \sum_{i=1}^{n} x_i y_i$ Let $S_{xx} = \sum_{i=1}^{n} x_i^2$
We then rewrite

$$\beta_1 = \frac{S_{xy} - \frac{S_x S_y}{n}}{S_{xx} - \frac{S_x^2}{n}}$$

and expand

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) &= S_{xy} - (\frac{S_y}{n})S_x - (\frac{S_x}{n})S_y + \frac{(S_x S_y)}{n} \\
&= S_{xy} - \frac{S_x S_y}{n} \\
&= S_{xy} - \frac{\frac{S_x S_y}{n}}{n}
\end{aligned}$$

Similarly, also expand

$$(x_i - \bar{x})^2 = S_{xx} - (\frac{S_x^2}{n})$$

$$(y_i - \bar{y})^2 = S_{yy} - (\frac{S_y^2}{n})$$

Next we rewrite

$$r = \frac{S_{xy} - \frac{S_x S_y}{n}}{\sqrt{(S_{xx} - \frac{S_x^2}{n})(S_{yy} - \frac{S_y^2}{n})}}$$

Note

$$\frac{s_y}{s_x} = \frac{\sqrt{(S_{xx} - \frac{S_x^2}{n})}}{\sqrt{(S_{yy} - \frac{S_y^2}{n})}}$$

$$r\frac{s_y}{s_x} = \beta_1 = \frac{S_{xy} - \frac{S_x S_y}{n}}{S_{xx} - \frac{S_x^2}{n}}$$

$\blacksquare$

Exercise 7

1. ```
   library(alr4)

   set.seed(256)

   total_rows <- nrow(Heights)
   con_set_size <- floor(total_rows / 3)
   val_set_size <- total_rows - con_set_size

   con_indices <- sample(seq_len(total_rows), size = con_set_size)
   con_set <- Heights[con_indices, ]
   val_set <- Heights[-con_indices, ]
   ```

2. ```
   X_train <- as.matrix(cbind(Intercept = 1, con_set$mheight))
   y_train <- as.matrix(con_set$dheight)
   X_val <- as.matrix(cbind(Intercept = 1, val_set$mheight))
   y_val <- as.matrix(val_set$dheight)

   XtX <- t(X_train) %*% X_train
   XtX_inv <- solve(XtX)
   Xty <- t(X_train) %*% y_train
   beta <- XtX_inv %*% Xty

   val_set$pred <- X_val %*% beta
   val_set$resid <- val_set$dheight - val_set$pred
   mse <- mean(val_set$resid^2)
   rmse <- sqrt(mse)
   cat(rmse) # 2.205687
   ```

3. ```
   resid_train <- y_train - X_train %*% beta
   RSS <- sum(resid_train^2)
   n <- nrow(X_train)
   p <- ncol(X_train)
   sigma2 <- RSS / (n - p)
   X_val_XtX_inv <- X_val %*% XtX_inv
   val_set$SE_pred <- sqrt(sigma2 * (1 + rowSums(X_val_XtX_inv * X_val)))
   mse2 = mean(val_set$SE_pred^2)
   rmse2 = sqrt(mse2)
   cat(rmse2) # 2.392998
   ```

   The first root average squared prediction error is only .2 lower than 2nd prediction error.

References

[1] Kaare Brandt Petersen and Michael Syskind Pedersen. *The Matrix Cookbook*. N/A, 2012.

[2] Dennis D. Wackerly, William Mendenhall, and Richard L. Scheaffer. *Mathematical Statistics with Applications*. Thomson Brooks/Cole, 2008.

[3] Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer, 2004.
   ```