Basics of Statistical Inference and Linear Regression

Fall 2022, MATH8050: Homework 3
Your Name, Section XXX

Due September 21, 12:00 PM

General instructions for homeworks: Please follow the uploading file instructions according to the syllabus. Each answer must be supported by written statements as well as any code used. Your code must be completely reproducible and must compile. For writing mathematical expressions in R Markdown, refer to the homework template posted on Canvas, a 30-minute tutorial, or LaTeX/Mathematics.

Advice: Start early on the homeworks and it is advised that you not wait until the last day. While the professor and the TA's check emails, they will be answered in the order they are received and last minute help will not be given.

No late homework's will be accepted.

R Working Environment

Please load all the packages used in the following R chunk before the function `sessionInfo()`

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
# load packages


sessionInfo()
```

```
## R version 4.1.3 (2022-03-10)
## Platform: aarch64-apple-darwin20 (64-bit)
## Running under: macOS Monterey 12.3.1
##
## Matrix products: default
## LAPACK: /Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] scico_1.3.1     patchwork_1.1.2 forcats_0.5.1   stringr_1.4.0
```

```
##  [5] dplyr_1.0.8     purrr_0.3.4     readr_2.1.2     tidyr_1.2.0
##  [9] tibble_3.1.6    ggplot2_3.3.5   tidyverse_1.3.1
##
## loaded via a namespace (and not attached):
##  [1] tidyselect_1.1.2 xfun_0.30        haven_2.5.0      colorspace_2.0-3
##  [5] vctrs_0.4.1      generics_0.1.2   htmltools_0.5.2  yaml_2.3.5
##  [9] utf8_1.2.2       rlang_1.0.2      pillar_1.7.0     glue_1.6.2
## [13] withr_2.5.0      DBI_1.1.2        dbplyr_2.1.1     modelr_0.1.8
## [17] readxl_1.4.0     lifecycle_1.0.1  munsell_0.5.0    gtable_0.3.0
## [21] cellranger_1.1.0 rvest_1.0.2      evaluate_0.15    knitr_1.38
## [25] tzdb_0.3.0       fastmap_1.1.0    fansi_1.0.3      broom_0.8.0
## [29] backports_1.4.1  scales_1.2.0     jsonlite_1.8.0   fs_1.5.2
## [33] hms_1.1.1        digest_0.6.29    stringi_1.7.6    grid_4.1.3
## [37] cli_3.2.0        tools_4.1.3      magrittr_2.0.3   crayon_1.5.1
## [41] pkgconfig_2.0.3  ellipsis_0.3.2   xml2_1.3.3       reprex_2.0.1
## [45] lubridate_1.8.0  assertthat_0.2.1 rmarkdown_2.13   httr_1.4.2
## [49] rstudioapi_0.13  R6_2.5.1         compiler_4.1.3
```

Total points on assignment: 10 (reproducibility) + 6 (Q1) + 49 (Q2) + 21 (Q3) + 14 (Q4)

Reproducibility component: 10 points.

1. (6pts total) Consider the function

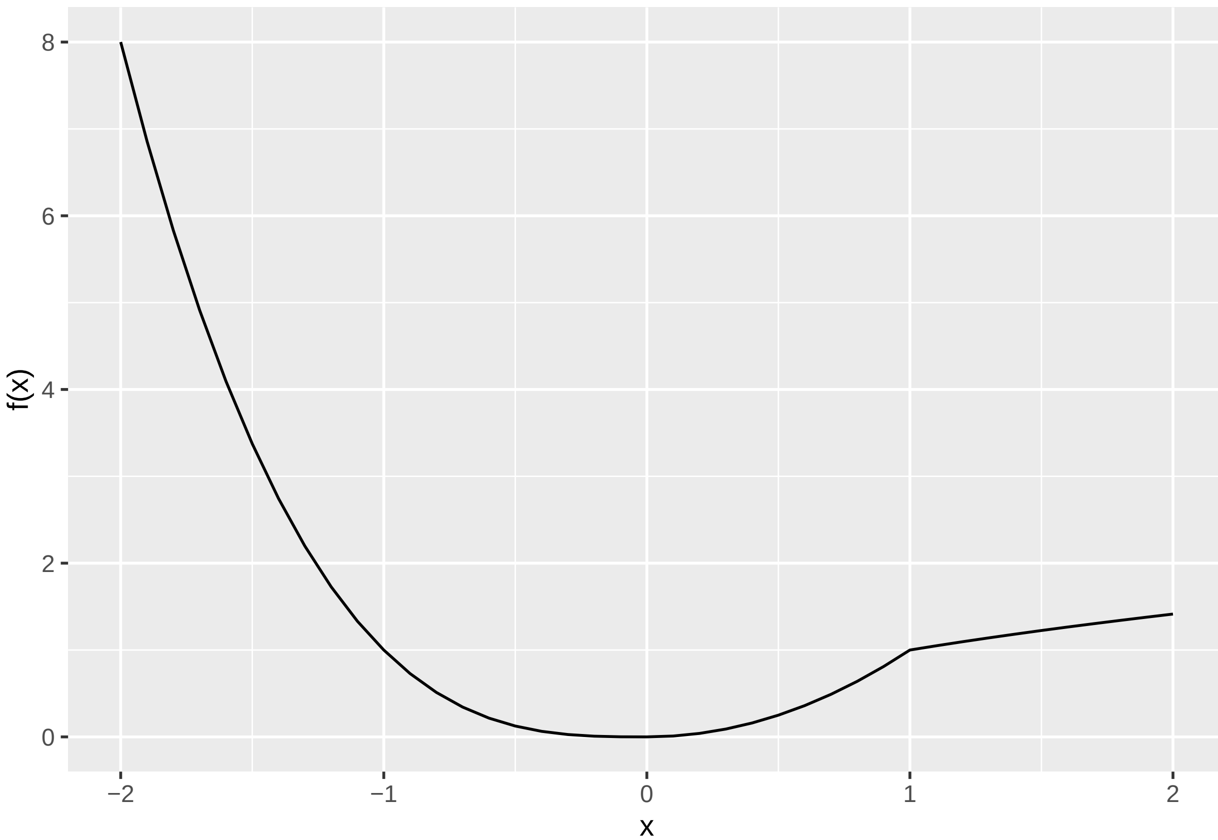
$$f(x) = \begin{cases} -x^3 & x \leq 0, \\ x^2 & x \in (0,1], \\ \sqrt{x} & x > 1. \end{cases}$$

Write an R function called `f()` that can take a numeric vector as input argument to implement this function using `if` (or `ifelse`) statements, and plot the function in the interval $[-2, 2]$.

Solution:

```r
x = seq(-2, 2, by=0.1)

f = function(x){
  n = length(x)
  y = rep(NA, n)
  for(i in 1:n){
    if(x[i] <= 0){
      y[i] = - x[i]^3
    }else if(x[i] > 0 && x[i] <= 1){
      y[i] = x[i]^2
    }else{
      y[i] = sqrt(x[i])
    }
  }

  return(y)

}
```

```r
df = data.frame(x=x, y=f(x))

ggplot(df) +
  geom_line(aes(x,y)) +
  ylab("f(x)")
```



2. (49 pts total, equally weighted) We write $X \sim \text{Poisson}(\theta)$ if X has the Poisson distribution with rate $\theta > 0$, that is, its p.m.f. is

$$p(x|\theta) = \text{Poisson}(x|\theta) = \exp(-\theta)\frac{\theta^x}{x!},$$

for $x \in \{0, 1, 2, \ldots\}$ (and is 0 otherwise). Suppose $X_1, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\theta)$ given θ. Please work on the following problems.

a. Find the likelihood function and the maximum likelihood estimator of θ.

Solution:

Since the data is independent given θ, the likelihood factors and we get

$$\begin{aligned} L(\theta \mid X_1, X_2, \ldots, X_n) &:= p(X_1, X_2, \ldots, X_n \mid \theta) \\ &= \prod_{i=1}^{n} p(X_i \mid \theta) \\ &= \prod_{i=1}^{n} \exp(-\theta) \frac{\theta^{X_i}}{X_i!} \\ &= \exp(-n\theta) \frac{\theta^{\sum_{i=1}^{n} X_i}}{\prod_{i=1}^{n} X_i!} \end{aligned}$$

Taking the log of the likelihood yields

$$\log L(\theta | X_{1:n}) = -n\theta + (\sum_{i=1}^{n} X_i) \log(\theta) + C,$$

where C is a constant (that does not depend on θ).

Now we take the derivative of the likelihood function with respect to θ, and set it to zero

$$0 = \frac{\partial \log L(\theta | X_{1:n})}{\partial \theta} = -n + \frac{\sum_{i=1}^{n} X_i}{\theta}.$$

We obtain a maximum likelihood estimator $\hat{\theta} = \frac{\sum_{i=1}^{n} X_i}{n} =: \bar{X}$.

b. Show that the $\hat{\theta}$ is the unique maximum likelihood estimator (MLE).

Solution:

It suffices to show that the log-likelihood function is concave. Indeed, taking the second derivative of the log-likelihood yields

$$\frac{\partial^2 \log L(\theta | X_{1:n})}{\partial \theta^2} = -\frac{\sum_{i=1}^{n} X_i}{\theta^2} < 0,$$

which suggests that the log-likelihood is concave and the MLE is unique.

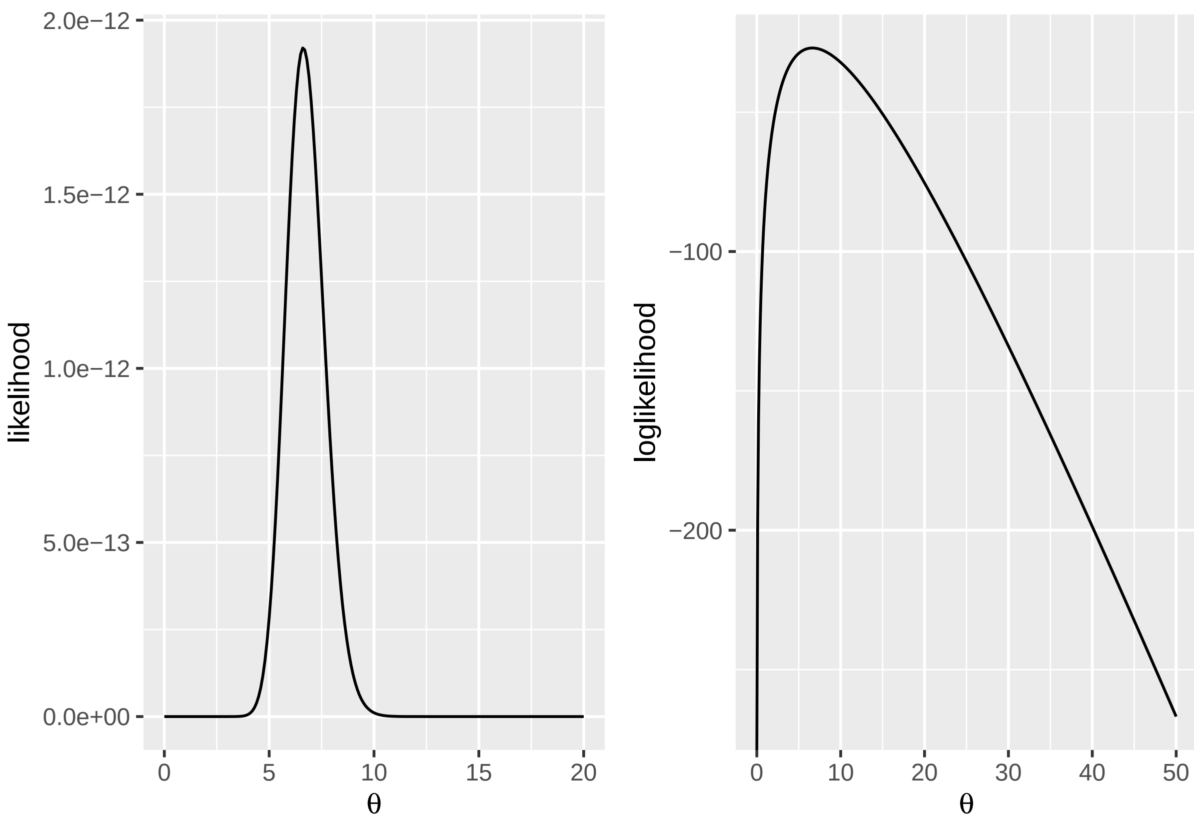
c. Further assume that the observations are $\{0, 2, 5, 6, 10, 15, 8, 7\}$. Plot the likelihood function as a function of θ, and maximize the likelihood function numerically using the `optim` function from the base package. Denote the numerical solution to the MLE by $\hat{\theta}_1$. Is the numerical solution $\hat{\theta}_1$ the same as the true MLE $\hat{\theta} = \bar{x} = 6.625$?

Solution:

Plots of the likelihood function or the log-likelihood function are both acceptable.

```
PoissonLoglik = function(x, theta){
  n = length(x)

  l = -n*theta + sum(x) * log(theta) - sum(lfactorial(x))
```

```
  return(l)
}

theta = seq(0, 20, by=0.1)
x = c(0, 2, 5, 6, 10, 15, 8, 7)
loglik = PoissonLoglik(x, theta)
df = data.frame(theta=theta, loglik=loglik)
p1 = df %>%
  ggplot() +
  geom_line(aes(x=theta, y=exp(loglik))) +
  labs(x=expression(theta), y="likelihood")
theta = seq(0, 50, by=0.1)
loglik = PoissonLoglik(x, theta)
df = data.frame(theta=theta, loglik=loglik)
p2 = df %>%
  ggplot() +
  geom_line(aes(x=theta, y=(loglik))) +
  labs(x=expression(theta), y="loglikelihood")
p1 + p2
```



Now let's find the numerical solution using `optim`.

```
init = 1

# by default, optim performs minimization
# so we need to set the argument fnscale to be negative
optim(par=init, fn=PoissonLoglik, x=x,
      control=list(fnscale=-1))
```

```
## Warning in optim(par = init, fn = PoissonLoglik, x = x, control = list(fnscale = -1)): one-dimensiona
## use "Brent" or optimize() directly
```

```
## $par
## [1] 6.625
##
## $value
## [1] -26.97827
##
## $counts
## function gradient
##       36       NA
##
## $convergence
## [1] 0
##
## $message
## NULL
```

d. Assume that we want to put a prior distribution on θ and perform Bayesian estimation. Your prior on θ is a gamma distribution

$$p(\theta) = \mathrm{Ga}(\theta|a, b) = \frac{b^a}{\Gamma(a)}\theta^{a-1}\exp(-b\theta)\mathbb{1}(\theta > 0).$$

Derive the posterior distribution of θ.

Solution:

According the part (a), the likelihood is

$$p(x_{1:n}|\theta) \underset{\theta}{\propto} e^{-n\theta}\theta^{\sum x_i}.$$

Thus, using Bayes' theorem,

$$\begin{aligned}
p(\theta|x_{1:n}) &\propto p(x_{1:n}|\theta)p(\theta) \\
&\propto e^{-n\theta}\theta^{\sum x_i}\theta^{a-1}e^{-b\theta}\mathbb{1}(\theta > 0) \\
&\propto e^{-(b+n)\theta}\theta^{a+\sum x_i - 1}\mathbb{1}(\theta > 0) \\
&\propto \mathrm{Gamma}\left(\theta \mid a + \sum x_i,\, b + n\right).
\end{aligned}$$

Therefore, since the posterior density must integrate to 1, we have

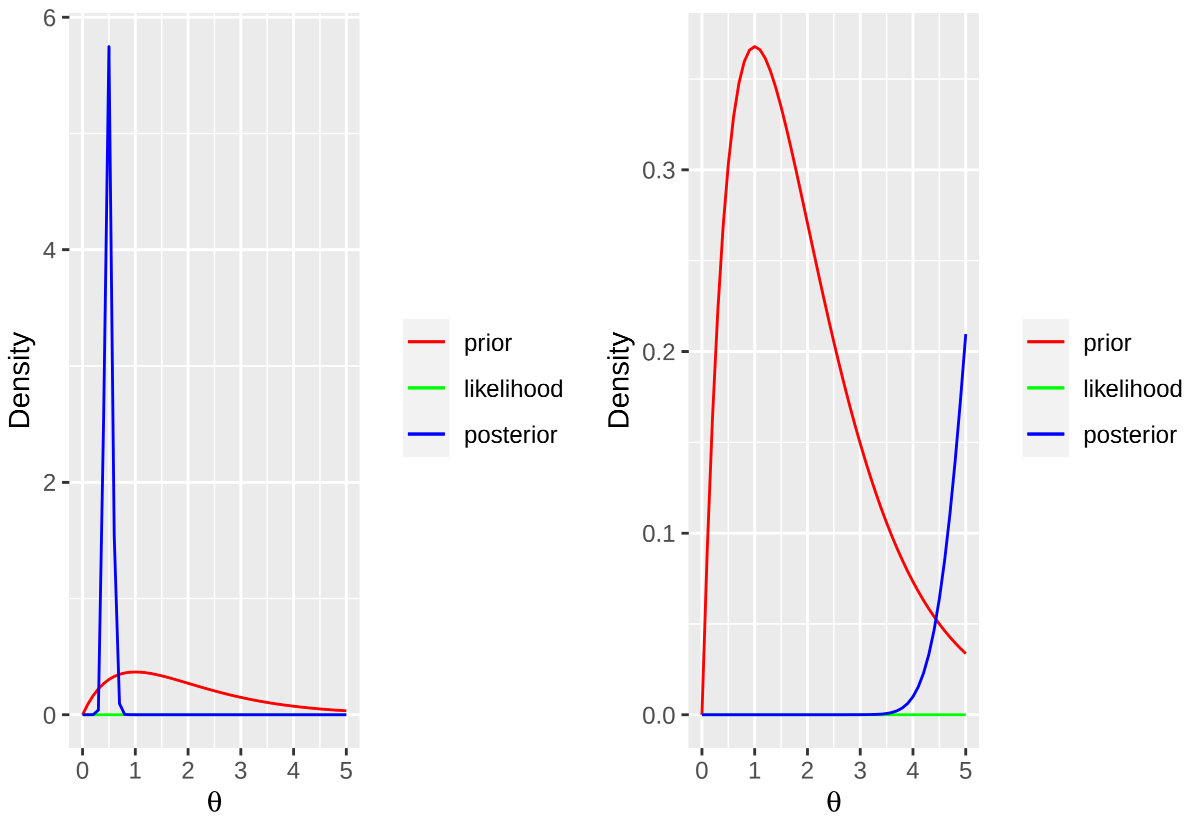
$$p(\theta|x_{1:n}) = \mathrm{Ga}\big(\theta \mid a + \sum x_i,\, b + n\big).$$

e. Consider two sets of values for a, b: (1) $a = 0.1, b = 0.1$, which corresponds to a non-informative prior, and (2) $a = 2, b = 5$, which corresponds to an informative prior. Plot the prior, likelihood, and posterior, and arrange the plots from (1) and (2) in two panels in a row.

Solution:

```r
a = .01
b = 100
theta = seq(0,5,by=0.1)
prior = dgamma(theta, shape=a, rate=b)
likelihood = rep(NA, length(theta))
for(i in 1:length(theta)){
  likelihood[i] = exp(PoissonLoglik(x, theta[i]))
}
post = dgamma(theta, shape=a+sum(x), rate=b + length(x))
df = data.frame(theta=theta, likelihood=likelihood,
                post=post)
g1 = df %>%
  ggplot() +
  geom_line(aes(x=theta, y=prior, color="prior")) +
  geom_line(aes(x=theta, y=likelihood, color="likelihood")) +
  geom_line(aes(x=theta, y=post, color="posterior")) +
  labs(
    x = expression(theta),
    y = "Density"
  ) +
  scale_color_manual(name=NULL,
                     values=c(
                       "prior"="red",
                       "likelihood"="green",
                       "posterior"="blue"
                     ))

a = 2
b = 1
theta = seq(0,5,by=0.1)
prior = dgamma(theta, shape=a, rate=b)
likelihood = rep(NA, length(theta))
for(i in 1:length(theta)){
  likelihood[i] = exp(PoissonLoglik(x, theta[i]))
}
post = dgamma(theta, shape=a+sum(x), rate=b + length(x))
df = data.frame(theta=theta, likelihood=likelihood,
                post=post)
g2 = df %>%
  ggplot() +
  geom_line(aes(x=theta, y=prior, color="prior")) +
  geom_line(aes(x=theta, y=likelihood, color="likelihood")) +
  geom_line(aes(x=theta, y=post, color="posterior")) +
  labs(
    x = expression(theta),
    y = "Density"
  ) +
  scale_color_manual(name=NULL,
                     values=c(
                       "prior"="red",
                       "likelihood"="green",
                       "posterior"="blue"
                     ))
```

```
g1 + g2
```



f. Derive the expression of the marginal likelihood with the $\text{Ga}(a, b)$ prior on θ.

Solution:

$$
\begin{aligned}
p(x_{1:n}) &= \int p(x_{1:n}|\theta)p(\theta)d\theta \\
&= \int_0^\infty \exp(-n\theta)\frac{\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!} \times \frac{b^a}{\Gamma(a)}\theta^{a-1}\exp(-b\theta)\mathbb{1}(\theta > 0)d\theta \\
&= \int_0^\infty \exp\{-(n+b)\theta\} \times \theta^{\sum_{i=1}^n x_i + a - 1} \times \frac{b^a}{\Gamma(a)\prod_{i=1}^n x_i!}d\theta \\
&= \frac{\Gamma(\sum_{i=1}^n x_i + a)}{(n+b)^{\sum_{i=1}^n x_i + a}} \times \frac{b^a}{\Gamma(a)\prod_{i=1}^n x_i!}
\end{aligned}
$$

g. Derive the expression of the posterior predictive distribution with the $\text{Ga}(a, b)$ prior on θ.

Solution:

Let x_{n+1} be a new data point. Then

$$\begin{aligned}
p(x_{n+1}|x_{1:n}) &= \int p(x_{n+1}|\theta, x_{1:n}) p(\theta|x_{1:n}) d\theta \\
&= \int p(x_{n+1}|\theta) p(\theta|x_{1:n}) d\theta \\
&= \int_0^\infty \exp(-\theta) \frac{\theta^{x_{n+1}}}{x_{n+1}!} \times \frac{(b+n)^{a+\sum x_i}}{\Gamma(a+\sum x_i)} \theta^{a+\sum x_i - 1} \exp\{-(b+n)\theta\} d\theta \\
&= \frac{(b+n)^{a+\sum x_i}}{\Gamma(a+\sum x_i) x_{n+1}!} \int_0^\infty \exp\{-(b+n+1)\theta\} \theta^{a+\sum_{i=1}^n x_i + x_{n+1} - 1} d\theta \\
&= \frac{(b+n)^{a+\sum_{i=1}^n x_i}}{\Gamma(a+\sum_{i=1}^n x_i) x_{n+1}!} \times \frac{\Gamma(a+\sum_{i=1}^{n+1} x_i)}{(b+n+1)^{a+\sum_{i=1}^{n+1} x_i}} \\
&= \frac{\Gamma(a+\sum_{i=1}^{n+1} x_i)}{\Gamma(a+\sum_{i=1}^n x_i) x_{n+1}!} \times \frac{(b+n)^{a+\sum_{i=1}^n x_i}}{(b+n+1)^{a+\sum_{i=1}^{n+1} x_i}}
\end{aligned}$$

3. (21pts total, equally weighted) Refer to the Lecture 4, and prove the following statements:

a. Show that the hat matrix is symmetric and idempotent: i.e., $H^\top = H$ and $HH = H$.

Solution:

Proof. By definition, $H = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top$. Then $H^\top = \{\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top\}^\top = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top$.

$$\begin{aligned}
HH &= \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top * \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \\
&= \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}(\mathbf{X}^\top \mathbf{X})(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \\
&= \mathbf{X} I (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top = H
\end{aligned}$$

□

b. Show that the residual vector $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ has mean zero, i.e., $E(\mathbf{e}) = 0$ and variance $Var(\mathbf{e}) = \sigma^2(I - H)$.

Solution:

Proof. Note that $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = H\mathbf{y}$ and $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = (I-H)\mathbf{y}$. We also know that $E[\mathbf{y}] = \boldsymbol{X\beta}$. Thus,

$$\begin{aligned}
E(\mathbf{e}) &= (I-H)E(\mathbf{y}) = (I-H)\boldsymbol{X\beta} \\
&= \boldsymbol{X\beta} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \boldsymbol{X\beta} = \boldsymbol{X\beta} - \boldsymbol{X\beta} = \mathbf{0}.
\end{aligned}$$

To prove the variance, it follows that

$$Var(\mathbf{e}) = Var((I-H)\boldsymbol{y}) = (I-H)Var(\boldsymbol{y})(I-H)^\top = (I-H)\sigma^2(I-H) = \sigma^2(I-H).$$

□

c. Show that the MSE is an unbiased estimator of σ^2, i.e., $E(MSE) = \sigma^2$.

Solution:

Proof. We first write MSE in a matrix/vector notation:

$$MSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2/(n-p) = \sum_{i}^{n} e_i^2/(n-p) = \mathbf{e}^\top \mathbf{e}/(n-p).$$

Taking the expectation yields that

$$\begin{aligned}
E(MSE) &= \frac{E(\boldsymbol{e}^\top \boldsymbol{e})}{n-p} = \frac{E(\mathrm{tr}(\boldsymbol{e}^\top \boldsymbol{e}))}{n-p} = \frac{E(\mathrm{tr}\{\boldsymbol{e}\boldsymbol{e}^\top\})}{n-p} \\
&= \frac{\mathrm{tr}(E(\boldsymbol{e}\boldsymbol{e}^\top))}{n-p} \overset{\text{From Part b}}{=} \frac{\mathrm{tr}\{\sigma^2(I-H)\}}{n-p} = \frac{\sigma^2 \mathrm{tr}(I_{n\times n}) - \mathrm{tr}(H)}{n-p} = \frac{\sigma^2(n-\mathrm{tr}(H))}{n-p}.
\end{aligned}$$

Note that $\mathrm{tr}(H) = \mathrm{tr}\{\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\} = \mathrm{tr}\{(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{X}\} = \mathrm{tr}(I_{p\times p}) = \mathrm{p}$.

Therefor, $E(MSE) = \frac{\sigma^2(n-\mathrm{tr}(H))}{n-p} = \sigma^2$, as desired. □

4. (14pts total, equally weighted) Refer to the Lecture 4. For the normal simple linear regression model

$$y_i | x_i, \boldsymbol{\beta}, \sigma^2 \overset{ind}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2)$$

with $i = 1, \ldots, n$.

a. Show that

$$Var(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1} = \begin{bmatrix} \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i-\bar{x})^2}\right) & \frac{-\sigma^2\bar{x}}{\sum_{i=1}^n (x_i-\bar{x})^2} \\ \frac{-\sigma^2\bar{x}}{\sum_{i=1}^n (x_i-\bar{x})^2} & \frac{\sigma^2}{\sum_{i=1}^n (x_i-\bar{x})^2} \end{bmatrix},$$

Solution:

Proof. Note that for simple linear regression model, the design matrix is

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}.$$

Then it follows that

$$\boldsymbol{X}^\top\boldsymbol{X} = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{bmatrix},$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$.

Recall that if $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. Applying this formula yields that

$$\begin{aligned}
ad - bc &= n\sum_{i=1}^n x_i^2 - (n\bar{x})^2 = n\left(\sum_{i=1}^n x_i^2 - \bar{x}^2\right) = n\sum_{i=1}^n (x_i - \bar{x})^2 \\
\frac{d}{ad-bc} &= \frac{\sum_{i=1}^n x_i^2}{ad-bc} = \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2 + n\bar{x}^2}{n\sum_{i=1}^n (x_i-\bar{x})^2} = \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i-\bar{x})^2}.
\end{aligned}$$

Thus,

$$Var(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^\top \mathbf{X})^{-1} = \begin{bmatrix} \sigma^2 \left(\frac{1}{n} + \frac{\overline{x}^2}{\sum_{i=1}^n (x_i - \overline{x})^2} \right) & \frac{-\sigma^2 \overline{x}}{\sum_{i=1}^n (x_i - \overline{x})^2} \\ \frac{-\sigma^2 \overline{x}}{\sum_{i=1}^n (x_i - \overline{x})^2} & \frac{\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})^2} \end{bmatrix}.$$

□

b. Assume that σ^2 is given. Let $\hat{y}_n := \hat{\beta}_0 + \hat{\beta}_1 x_h$ denote the estimator of y_h at new covariate x_h. Show that the posterior predictive distribution of the new response y_h at new covariate x_h is

$$y_h \mid \mathbf{y}, \mathbf{X}, x_h, \sigma^2 \sim \mathcal{N}\left(\hat{y}_h, \sigma^2 \left\{ 1 + \frac{1}{n} + \frac{(x_h - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right\} \right).$$

Solution: ===

For this problem, as long as students can show the following relationships for expectation and variance, full credit can be given.

Note that $y_h = \beta_0 + \beta_1 x_h + \epsilon_h$ and

$$E[\hat{y}_h] = E[\hat{\beta}_0 + \hat{\beta}_1 x_h] = \beta_0 + \beta_1 x_h = E[y_h].$$

Let $\mathbf{f}_h := (1, x_h)^\top$. Then $\hat{y}_h = \mathbf{f}_h^\top \hat{\boldsymbol{\beta}}$.

$$\begin{aligned} Var(\hat{y}_h) &= \mathbf{f}_h^\top Var(\hat{\boldsymbol{\beta}}) \mathbf{f}_h = \mathbf{f}_h^\top \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \mathbf{f}_h \\ &= (1, x_h) \begin{bmatrix} \sigma^2 \left(\frac{1}{n} + \frac{\overline{x}^2}{\sum_{i=1}^n (x_i - \overline{x})^2} \right) & \frac{-\sigma^2 \overline{x}}{\sum_{i=1}^n (x_i - \overline{x})^2} \\ \frac{-\sigma^2 \overline{x}}{\sum_{i=1}^n (x_i - \overline{x})^2} & \frac{\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})^2} \end{bmatrix} (1, x_h)^\top \\ &= \sigma^2 \left(\frac{1}{n} + \frac{\overline{x}^2 - x_h \bar{x}}{\sum_{i=1}^n (x_i - \overline{x})^2}, \frac{-\overline{x} + x_h}{\sum_{i=1}^n (x_i - \overline{x})^2} \right) (1, x_h)^\top = \sigma^2 \left(\frac{1}{n} + \frac{(\overline{x} - x_h)^2}{\sum_{i=1}^n (x_i - \overline{x})^2} \right) \end{aligned}$$

Note that

$$Var(\hat{\beta}_0 + \hat{\beta}_1 x_h + \epsilon_h) = Var(\hat{y}_h) + Var(\epsilon_h) = \sigma^2 \left(1 + \frac{1}{n} + \frac{(\overline{x} - x_h)^2}{\sum_{i=1}^n (x_i - \overline{x})^2} \right)$$