STAT 543/641 – WINTER 2019 – HOMEWORK #1

DUE FEBRUARY 11, 2019

(1) Let $X_1, \dots, X_n$ be a random sample from $N(\mu, \sigma)$ and let S^2 be the associated unbiased estimator of σ^2:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

Show that

$$\operatorname{Var} S^2 = \frac{2\sigma^4}{n-1}.$$

Feel free to "cheat" and use the fact that

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

(Can you do it without "cheating"?)

Solution: The distribution χ^2_{n-1} has variance $2(n-1)$. Therefore,

$$2(n-1) = \operatorname{Var}\left[\frac{(n-1)S^2}{\sigma^2}\right] = \frac{(n-1)^2}{\sigma^4} \operatorname{Var} S^2.$$

Solving for $\operatorname{Var} S^2$, we get

$$\operatorname{Var} S^2 = \frac{2\sigma^4}{n-1}.$$

(2) (a) Let $\widetilde{x}$ be the median of $x_1, \dots, x_n$, n odd. Prove that the identity

$$\sum_{i=1}^{n} |x_i - z| = \min_{y \in \mathbb{R}} \sum_{i=1}^{n} |x_i - y|$$

holds if and only if $z = \widetilde{x}$.

Solution: Let

$$f(z) = \sum_{i=1}^{n} |x_i - z| = \sum_{i=1}^{n} \operatorname{sgn}(x_i - z)(x_i - z).$$

Suppose $z \notin \{x_1, \dots, x_n\}$. Then, for each i, $\operatorname{sgn}(x_i - z)$ is constant in a neighborhood U_z of z. Thus, f is differentiable in U_z for and

$$f'(w) = \sum_{i=1}^{n} \operatorname{sgn}(x_i - z)(-1)$$

for all $w \in U_z$. This expression for $f'(w)$ is a sum of n terms, each of which is ± 1. Since n is odd, this sum cannot be 0. Thus, $f'(w) \neq 0$ for all $w \in U_z$.

Therefore, f has no local extrema in U_z. In particular, f can't achieve its global minimum at z. It follows that f must achieve its minimum value on the set $\{x_1, \ldots, x_n\}$.
Reindexing if necessary, assume

$$x_1 \leq x_2 \leq \cdots \leq x_n.$$

I claim that f takes on its minimum value at $z = x_m$. If $n = 2m-1$, and $\ell \leq m$, then

$$\begin{aligned}
\sum_{i=1}^{n} |x_i - x_\ell| &= \sum_{i=1}^{\ell-1} |x_i - x_\ell| + \sum_{i=1}^{\ell-1} |x_{n-i+1} - x_\ell| + \sum_{i=\ell}^{n-\ell} (x_{n-i+1} - x_\ell) \\
&= \sum_{i=1}^{\ell-1} (x_\ell - x_i) + \sum_{i=1}^{m-1} (x_{n-i+1} - x_\ell) + \sum_{i=\ell}^{n-\ell} (x_{n-i+1} - x_\ell) \\
&= \sum_{i=1}^{\ell-1} (x_{n-i+1} - x_i) + \sum_{i=\ell}^{n-\ell} (x_{n-i+1} - x_\ell)
\end{aligned}$$

In particular, if $\ell \leq m$, then

$$\begin{aligned}
\sum_{i=1}^{n} |x_i - x_m| &= \sum_{i=1}^{m-1} (x_{n-i+1} - x_i) \\
&= \sum_{i=1}^{\ell-1} (x_{n-i+1} - x_i) + \sum_{i=\ell}^{m-1} (x_{n-i+1} - x_i) \\
&\leq \sum_{i=1}^{\ell-1} (x_{n-i+1} - x_i) + \sum_{i=\ell}^{m-1} (x_{n-i+1} - x_\ell) && (i \geq \ell \Longrightarrow x_i \geq x_\ell) \\
&\leq \sum_{i=1}^{\ell-1} (x_{n-i+1} - x_i) + \sum_{i=\ell}^{n-\ell} (x_{n-i+1} - x_\ell) && (n - \ell = (m-1) + (m-\ell) \geq m-1) \\
&= \sum_{i=1}^{n} |x_i - x_\ell|.
\end{aligned}$$

Now suppose $\ell > m$. We consider the sequence $y_i := -x_{n-i+1}$. By the above, if $n - \ell + 1 \leq m$, then

$$\begin{aligned}
\sum_{i=1}^{n} |x_{n-i+1} - x_m| &= \sum |-x_{n-i+1} - (-x_m)| \\
&= \sum_{i=1}^{n} |y_i - y_m| \\
&\leq \sum_{i=1}^{n} |y_i - y_{n-\ell+1}| \\
&= \sum_{i=1}^{n} |-x_{n-i+1} - (-x_{n-(n-\ell+1)+1})| \\
&= \sum_{i=1}^{n} |x_{n-i+1} - x_\ell| \\
&= \sum_{i=1}^{n} |x_i - x_\ell|,
\end{aligned}$$

as was to be shown.

(b) Let $X_1, \ldots, X_n$ be a random sample from $\mathscr{L}(\mu, b)$, where $\mathscr{L}(\mu, b)$ is the *Laplace distribution* with density

$$f(x|\mu, b) = \frac{1}{2b} e^{-|x-\mu|/b}.$$

Assuming that b is known and that n is odd, Show that the MLE of μ is the sample median, $\widetilde{X}$. (Hint: Use (a).)

Solution: We minimize the negative log-likelihood function,

$$h(\mu) = \log 2 + \log b + \frac{1}{b} \sum_{i=1}^{n} |x - \mu|.$$

For every $b > 0$,

$$\operatorname*{argmin}_{\mu} h(\mu) = \operatorname*{argmin}_{\mu} \sum_{i=1}^{n} |x_i - \mu|.$$

By (a),

$$\operatorname*{argmin}_{\mu} \sum_{i=1}^{n} |x_i - \mu| = \widetilde{x}.$$

(3) [2, Exercise 7.1.3] Let $Y_1 < Y_2 < Y_3$ be the order statistics of a random sample of size three drawn from the uniform distribution having density function

$$f(x|\theta) = \begin{cases} 1/\theta & \text{if } 0 < x < \theta \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$. Show that $4Y_1$, $2Y_2$, and $\frac{4}{3}Y_3$ are all unbiased estimators of θ. Find the variance of each of these estimators.

Solution: Let $Z_i = \theta^{-1}Y_i$. Then Z_1, Z_2, and Z_3 are the order statistics of *standard* uniform random variables and, thus, have densities

$$3(1-z)^2, \quad 6z(1-z), \quad \text{and} \quad 3z^2,$$

respectivelz. Therefore,

$$\begin{aligned}
\mathbb{E}\, Z_1 &= \int_0^1 3z(1-z)^2\, dz = \frac{1}{4}, \\
\operatorname{Var} Z_1 &= \mathbb{E}[(Z_1 - \tfrac{1}{4})^2] = \int_0^1 3(z-\tfrac{1}{4})^2(1-z)^2\, dz = \frac{3}{80} \\
\mathbb{E}\, Z_2 &= \int_0^1 6z^2(1-z)\, dz = \frac{1}{2}, \\
\operatorname{Var} Z_2 &= \mathbb{E}[(Z_1 - \tfrac{1}{2})^2] = \int_0^1 3(z-\tfrac{1}{4})^2 z(1-z)\, dz = \frac{9}{80} \\
\mathbb{E}\, Z_3 &= \int_0^1 6z^2(1-z)\, dz = \frac{3}{4}, \\
\operatorname{Var} Z_3 &= \mathbb{E}[(Z_1 - \tfrac{1}{2})^2] = \int_0^1 3(z-\tfrac{3}{4})^2 z^2\, dz = \frac{3}{32}
\end{aligned}$$

If follows that

$$\mathbb{E}[4Y_1] = 4\,\mathbb{E}[\theta Z_1] = 4\theta\frac{1}{4} = \theta, \quad \operatorname{Var}[4Y_1] = 16\operatorname{Var}(\theta Z_1) = 16\theta^2\frac{3}{80} = \frac{3\theta^2}{5},$$

$$\mathbb{E}[2Y_2] = 2\,\mathbb{E}[\theta Z_2] = 2\theta\frac{1}{2} = \theta, \quad \operatorname{Var}[2Y_2] = 4\operatorname{Var}(\theta Z_2) = 4\theta^2\frac{9}{80} = \frac{9\theta^2}{20},$$

and

$$\mathbb{E}[\tfrac{4}{3}Y_3] = \frac{4}{3}\,\mathbb{E}[\theta Z_3] = \frac{4}{3}\theta\frac{3}{4} = \theta, \quad \operatorname{Var}[\tfrac{4}{3}Y_3] = \frac{16}{9}\operatorname{Var}(\theta Z_3) = \frac{16}{9}\theta^2\frac{3}{32} = \frac{\theta^2}{6}.$$

In particular, these are all unbiased estimators of θ.

(4) Suppose that

$$(X, Y) \sim N((\mu_X, \mu_Y), \Sigma), \quad \text{where} \quad \Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}.$$

(a) Write down the conditional density of Y given X.

Solution: The conditional distribution of Y given X is the quotient of the joint distribution of X and Y by the marginal distribution of X:

$$f_(y|x) = \frac{f(x,y)}{f(x)}$$

Setting

$$u = \frac{x - \mu_X}{\sigma_X}, \quad v = \frac{y - \mu_Y}{\sigma_Y},$$

we have

$$f(x, y) = c \exp\left\{-\frac{1}{2}\frac{1}{1-\rho^2}\left(u^2 - 2\rho uv + v^2\right)\right\}$$

Completing the square in v,

$$v^2 - 2\rho uv + u^2 = (v - \rho u)^2 + u^2(1 - \rho^2)$$

Thus,

$$\begin{aligned} f(x, y) &= C \exp\left\{-\frac{1}{2}\frac{(v - \rho u)^2}{1-\rho^2} - \frac{1}{2}u^2\right\} \\ (*) \qquad &= C \exp\left\{-\frac{1}{2}\frac{(v - \rho u)^2}{1-\rho^2}\right\} \exp\left\{-\frac{1}{2}u^2\right\}, \end{aligned}$$

where $C = (2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2})^{-1}$. Therefore,

$$\begin{aligned} f(x) &= \int_{-\infty}^{\infty} C \exp\left\{-\frac{1}{2}\frac{(v - \rho u)^2}{1-\rho^2}\right\} dy \\ &= C \exp\left\{-\frac{1}{2}u^2\right\} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\frac{(v - \rho u)^2}{1-\rho^2}\right\} dy \end{aligned}$$

By the translation-invariance of the Gaussian integral,

$$\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\frac{(v - \rho u)^2}{1-\rho^2}\right\} dy = \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\frac{v^2}{1-\rho^2}\right\} dy = \text{constant}.$$

It follows that

$$(**) \qquad f(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left\{-\frac{1}{2}\left(\frac{x - \mu_X}{\sigma_x^2}\right)^2\right\}$$

Thus, by $(*)$ and $(**)$,

$$\begin{aligned} f(y|x) &= \frac{\sqrt{2\pi}\sigma_X}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}\frac{(v - \rho u)^2}{1-\rho^2}\right\} \\ &= \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2}\frac{\left(y - \left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)\right)\right)^2}{\sigma_Y^2(1-\rho^2)}\right\} \end{aligned}$$

.

This final expression is the density of the univariate normal distribution

$$N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X), \sigma_Y^2(1-\rho^2)\right).$$

In other words, the marginal distribution of X is just the density of the univariate Gaussian distribution with mean μ_X and variance σ_X^2.

(b) Show that $\mathbb{E}[Y|X]$ is has the form $a + bX$. Express a and b in terms of μ_X, μ_Y, σ_X, σ_Y, and ρ. (Hint: Use (a).)

Solution: Since

$$Y|X = x \sim N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X), \sigma_Y^2(1-\rho^2)\right),$$

by (a),

$$\mathbb{E}[Y|X] = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X - \mu_X) = \underbrace{\rho\frac{\sigma_Y}{\sigma_X}}_{a} X + \underbrace{\mu_Y - \rho\frac{\sigma_Y}{\sigma_X}\mu_X}_{b}.$$

(c) Confirm your answer to (b) experimentally by finding the least-squares line for data sampled from a bivariate normal distribution with randomly generated mean and covariance matrix.

Solution: Something like this:

```
library(MASS)
library(GetoptLong)

rho <- -0.6
mu1 <- 1; s1 <- 2
mu2 <- 1; s2 <- 8

data <- mvrnorm(1e6, mu = c(mu1,mu2), Sigma = matrix(c(s1^2, s1*s2*rho, s1*s2*
f <- lm(formula = data[,2] ~ data[,1])

print(qq("predicted: (a, b) = (@{mu2 - rho*s2*mu1/s1}, @{rho*s2/s1})"))
print(qq("computed: (a, b) = (@{f$coefficients[1]}, @{f$coefficients[2]})"))
```

And here's the output:

```
predicted: (a, b) = (3.4, -2.4)
computed: (a, b) = (3.41071960965298, -2.40549630753275)
```

(5) Let $x_0, x_1, \ldots, x_n \in \mathbb{R}$, let $\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_n$ be independent normally distributed random variables with common mean 0 and common variance σ^2, and suppose

$$Y_i = a + bx_i + \varepsilon_i, \quad i = 0, 1, \ldots, n.$$

Recall our notation:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$S_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2, \quad S_{xy} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}), \quad S_{xY} = \sum_{i=1}^{n} (x_i - \overline{x})(Y_i - \overline{Y})$$

Let $\widehat{b}$, $\widehat{a}$, and $\widehat{\sigma}^2$ be the maximum likelihood estimators of b, a, and σ^2, respectively:

$$\begin{aligned}
\widehat{b} = \widehat{b}(Y_1, \ldots, Y_n) &= \frac{S_{xY}}{S_{xx}}, \\
\widehat{a} = \widehat{a}(Y_1, \ldots, Y_n) &= \overline{Y} - \widehat{b}\,\overline{x}, \\
\widehat{\sigma}^2 = \widehat{\sigma}^2(Y_1, \ldots, Y_n) &= \frac{1}{n} \sum_{i=1}^{n} (Y_i - \widehat{a} - \widehat{b}x_i)^2.
\end{aligned}$$

Note that these expressions involve only the *training data* $(x_1, Y_1), \ldots, (x_n, Y_n)$. They omit the *test data* (x_0, Y_0).

The training error of our regression model is

$$\text{MSE}_{\text{train}} = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} \left(Y_i - (\widehat{a} + \widehat{b}x_i)\right)^2\right],$$

while its test (prediction) error is

$$\text{MSE}_{\text{test}} = \mathbb{E}\left[\left(Y_0 - (\widehat{a} + \widehat{b}x_0)\right)^2\right].$$

We know that

$$\text{MSE}_{\text{train}} = \mathbb{E}\left[\widehat{\sigma}^2\right] = \frac{n-2}{n}\sigma^2.$$

In this exercise, we prove

$$\text{MSE}_{\text{test}} = \left(1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right)\sigma^2.$$

Note that

$$\text{MSE}_{\text{train}} \leq \text{MSE}_{\text{test}},$$

as one would expect (why?).

(a) Show that

$$\widehat{b} = \sum_{i=1}^{n} d_i Y_i \quad \text{and} \quad \widehat{a} = \sum_{i=1}^{n} c_i Y_i,$$

where

$$d_i = \frac{(x_i - \overline{x})}{S_{xx}} \quad \text{and} \quad c_i = \frac{1}{n} - \frac{\overline{x}(x_i - \overline{x})}{S_{xx}}.$$

(b) Prove that $\widehat{b}$ and $\widehat{a}$ are unbiased estimators of b and a, respectively. (Hint: Use (5a).)

(c) Establish the following identities:

$$\operatorname{Var}\widehat{b} = \frac{1}{S_{xx}}\sigma^2, \quad \operatorname{Var}\widehat{a} = \left(\frac{1}{nS_{xx}} \sum_{i=1}^{n} x_i^2\right)\sigma^2, \quad \operatorname{Cov}(\widehat{a}, \widehat{b}) = -\frac{\overline{x}}{S_{xx}}\sigma^2$$

(Hint: Use (5a) and the independence of $Y_1, \ldots, Y_n$.)

(d) What are the distributions of $\widehat{b}$ and $\widehat{a}$? (Hint: Use (5b) and (5c).)

(e) Establish the following identities:

$$\mathbb{E}[\widehat{a} + \widehat{b}x_0], \quad \mathrm{Var}(\widehat{a} + \widehat{b}x_0) = \left(\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right)\sigma^2.$$

What is the distribution of $\widehat{a} + \widehat{b}x_0$? (Hint: For the variance, use (5c). The calculation is a bit tricky; if you get stuck, see [1, §11.3.5].)

(f) Prove that

$$\mathbb{E}\left[\left(Y_0 - \widehat{a} - \widehat{b}x_0\right)^2\right] = \left(1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}\right)\sigma^2.$$

(Hint: Use the fact that Y_0 and $\widehat{a} + \widehat{b}x_0$ are independent (why?) and (5f).)

References

[1] Casella, Bergger, *Statistical Inference (2nd ed.)*, Duxbury, 2002.
[2] Hogg, McKean, Craig, *Introduction to Mathematical Statistics (7th ed.)*, Pearson, 2013.