

Please provide complete and well-written solutions to the following exercises.

Due September 21, 12PM noon PST, to be uploaded as a single PDF document to Gradescope.

Homework 3

Exercise 1. Let X be a random variable depending on parameters $\alpha, \beta > 0$. Suppose X satisfies

$$\mathbf{E}X = \alpha\beta, \quad \mathbf{E}X^2 = \alpha\beta^2.$$

Find the method of moments estimator for α , and also find the method of moments estimator for β .

Exercise 2. Let $\sigma > 0$, and suppose a random variable X has density

$$f(x) := \frac{1}{2\sigma} e^{-|x|/\sigma}, \quad \forall x \in \mathbf{R}.$$

Find the method of moments estimator for σ .

Exercise 3. Suppose you know that the following list of numbers is a random sample of size 20 from a Gaussian distribution with mean 1 and unknown variance $\sigma^2 > 0$.

2.0753 4.6678 -3.5177 2.7243 1.6375 -1.6154 0.1328 1.6852 8.1568 6.5389
-1.6998 7.0698 2.4508 0.8739 2.4295 0.5901 0.7517 3.9794 3.8181 3.8344.

- Using a method of moments estimator, estimate the value of σ^2 for this data. (Hint: Since the mean is 1, the variance σ^2 is equal to the second moment minus 1.)
- Denote your method of moments estimator for σ^2 as Z . Is Z unbiased?
- We know for sure that $\sigma^2 > 0$. Is it possible that Z could take negative values? If so, then perhaps Z is not the best way to estimate σ^2 .
- The Delta Method suggests that $1/Z$ could be a good estimate for $1/\sigma^2$. What estimate of $1/\sigma^2$ do you get from the data above? Is $\mathbf{E}|1/Z|$ finite? If not, then we cannot even compute the bias of this estimator. (Note that the distribution of Z should be closely related to a chi-squared distribution.) (Optional: if you use the fact that $\lim_{\varepsilon \rightarrow 0^+} \int_{\varepsilon < |t| < 1} \frac{1}{t} dt = 0$, then you should be able to estimate $\mathbf{E}(1/Z)$ as the number of samples n goes to infinity.)
- The Delta Method also suggests that Z^2 could be a good estimate for σ^4 . What estimate of σ^4 do you get from the data above? Is Z^2 an unbiased estimate of σ^4 ?
- Is Z^2 an asymptotically unbiased estimate of σ^4 ? That is, as the number of samples n goes to infinity, does $\mathbf{E}Z^2$ converge to σ^4 ?

Exercise 4 (Conditional Expectation as a Random Variable). Let $X, Y, Z: \Omega \rightarrow \mathbf{R}$ be discrete or continuous random variables. Let A be the range of Y . Define $g: A \rightarrow \mathbf{R}$ by $g(y) := \mathbf{E}(X|Y = y)$, for any $y \in A$. We then define the **conditional expectation** of X given Y , denoted $\mathbf{E}(X|Y)$, to be the random variable $g(Y)$.

- (i) Let X, Y be random variables such that (X, Y) is uniformly distributed on the triangle $\{(x, y) \in \mathbf{R}^2: x \geq 0, y \geq 0, x + y \leq 1\}$. Show that

$$\mathbf{E}(X|Y) = \frac{1}{2}(1 - Y).$$

- (ii) Prove the following version of the Total Expectation Theorem

$$\mathbf{E}(\mathbf{E}(X|Y)) = \mathbf{E}(X).$$

- (Optional) If X is a random variable, and if $f(t) := \mathbf{E}(X - t)^2$, $t \in \mathbf{R}$, then the function $f: \mathbf{R} \rightarrow \mathbf{R}$ is uniquely minimized when $t = \mathbf{E}X$. A similar minimizing property holds for conditional expectation. Let $h: \mathbf{R} \rightarrow \mathbf{R}$. Show that the quantity $\mathbf{E}(X - h(Y))^2$ is minimized among all functions $h: \mathbf{R} \rightarrow \mathbf{R}$ when $h(Y) = \mathbf{E}(X|Y)$. (Hint: use the previous item.)

- (iv) Show the following

$$\mathbf{E}(X|X) = X.$$

$$\mathbf{E}(X + Y|Z) = \mathbf{E}(X|Z) + \mathbf{E}(Y|Z).$$

- (v) If Z is independent of X and Y , show that

$$\mathbf{E}(X|Y, Z) = \mathbf{E}(X|Y).$$

(Here $\mathbf{E}(X|Y, Z)$ is notation for $\mathbf{E}(X|(Y, Z))$ where (Y, Z) is interpreted as a random vector, so that X is conditioned on the random vector (Y, Z) .)

Exercise 5 (Sunspot Data, Version 2). This exercise deals with sunspot data from the following files (the same data appears in different formats)

[txt file](#) [csv \(excel\) file](#)

These files are taken from <http://www.sidc.be/silso/datafiles#total>

To work with this data, e.g. in Matlab you can use the command

```
x=importdata('SN_d_tot_V2.0.txt')
```

to import the .txt file.

The format of the data is as follows.

- Columns 1-3: Gregorian calendar date (Year, Month, then Day)
- Column 4: Date in fraction of year
- Column 5: Daily total number of sunspots observed on the sun. A value of -1 indicates that no number is available for that day (missing value).
- Column 6: Daily standard deviation of the input sunspot numbers from individual stations.
- Column 7: Number of observations used to compute the daily value.
- Column 8: Definitive/provisional indicator. A blank indicates that the value is definitive. A '*' symbol indicates that the value is still provisional and is subject to a possible revision (Usually the last 3 to 6 months)

In a previous Exercise, we examined the number of sunspots U_t versus time t , where the units of t in the data are in integers divided by 365 (or by 365.25). You should have observed that the sunspots had a roughly 11-year periodicity. To make this more precise, we will use an estimator that checks for frequencies present in the data.

Denote $i := \sqrt{-1}$. For any real number r , consider the following estimator

$$\widehat{U}(r) := \sum_{t \in \mathbf{Z}/365} U_t e^{2\pi i t r}.$$

This estimator measures the “amount” of frequency r that the number of sunspots has. (As usual \mathbf{Z} denotes the set of integers.)

Plot $|\widehat{U}(r)|$ versus r , where $r \in [-1, 1]$. Do you observe any large absolute values of $\widehat{U}(r)$ for any values of r near $1/11$?

You should observe some large values of $\widehat{U}(r)$ when r takes the values: .0842, .0921, and .0995, corresponding to frequencies of 11.87, 10.858, and 10.05, respectively. This large signal should correspond to $r \in [.08, .105]$ (and to $r \in [-.105, -.08]$).

Exercise 6. Let $\theta \in \mathbf{R}$ be an unknown parameter. Consider the density

$$f_\theta(x) := \begin{cases} e^{-(x-\theta)}, & \text{if } x \geq \theta \\ 0, & \text{if } x < \theta. \end{cases}$$

Suppose X_1, \dots, X_n is a random sample of size n , such that X_i has density f_θ for all $1 \leq i \leq n$.

Show that $X_{(1)} = \min_{1 \leq i \leq n} X_i$ is a sufficient statistic for θ .