ESTIMATION BASICS

Mauro Díaz

Exercise 13.4 (Estimating the median)

a. Show that the sample median is an unbiased estimator of the median for $N(\mu, \sigma^2)$.

b. Show that the sample median is an unbiased estimator of the mean for any distribution with symmetric density.

Hint 1: The pdf of an order statistic is $f_{X_{(k)}}(x) = \frac{n!}{(n-k)!(k-1)!} f_X(x)\left(F_X(x)^{k-1}(1-F_X(x))^{n-k}\right)$.

Hint 2: A distribution is symmetric when X and $2a - X$ have the same distribution for some a.

(b)

Let Z_i, $i=1,\dots,n$ i.i.d variables with a symmetric distribution Z with $\mathbb{E}[Z]=\mu$ and let's order the data $Z_{1:n} \le Z_{2:n} \le \dots \le Z_{n:n}$.

The median is defined as the number m s.t. $P(Z \le m) = P(Z \ge m) = 1/2$ being X a RV.
In the case that our distribution is defined by the ECDF, having the Z_i ordered, the median will be M s.t. $F_n(M) = 1 - F_n(M) = 1/2$.
So, if:

$$\begin{cases} n = 2m+1 & \Rightarrow M = Z_{m+1:n} \\ n = 2m & \Rightarrow M = \dfrac{Z_{m:n} + Z_{m+1:n}}{2} \end{cases}$$

Also, if a RV is simmetric $Z = 2a - Z \Rightarrow F_Z(z) = P(Z \le z) = P(2a - Z \le z) = 1 - F_Z(2a-z)$ and if the Z is continuous, taking derivatives we can get $f_Z(z) = f_Z(2a-z)$.

<u>Case $n = 2m+1$</u>

In this case, $M = Z_{m+1:n}$ and $f_M(z) = \overbrace{C_{m+1}\binom{2m+1}{m}}^{C} f_Z(z)\left(F_Z(z)^m(1-F_Z(z))^m\right)$

and using $a=\mu$:

$$f_M(2\mu - z) = C \cdot f_Z(2\mu - z)\left(F_Z(2\mu-z)^m(1-F_Z(2\mu-z))^m\right)$$

and using the simmetry of $Z \Rightarrow f_M(z) = f_M(2\mu - z)$

Therefore, $\mathbb{E}[M] = \mathbb{E}[2\mu - M] \Rightarrow \mathbb{E}[M] = \mu \Rightarrow M$ is unbiased !

<u>Case $n = 2m$</u>:

In this case $M = \frac{Z_{min} + Z_{max:n}}{2}$ and can be shown that the joint distribution is also symmetric:

$$(Z_{min}, Z_{max:n}) = (2\mu - Z_{min}, 2\mu - Z_{max:n})$$ with the same proof and then

$$\Rightarrow M \overset{d}{=} \frac{(2\mu - Z_{min}) + (2\mu - Z_{max:n})}{2} = 2\mu - M$$

Therefore: $$E[M] = \frac{1}{2}\Big(\overset{\mu}{\overset{\shortparallel}{E[Z_{min}]}} + \overset{\mu}{\overset{\shortparallel}{E[Z_{max:n}]}}\Big) = \mu$$

(a) In particular $N(\mu, \sigma^2)$ is symmetric and the same proof can be applied. Then, the median and μ of $N(\mu, \sigma^2)$ have the same value.

13.3

Exercise 13.3 (Consistent but biased estimator)

a. Show that sample variance (the plug-in estimator of variance) is a biased estimator of variance.
b. Show that sample variance is a consistent estimator of variance.
c. Show that the estimator with $(N-1)$ (Bessel correction) is unbiased.

(a) Let's study the expectancy of the sample variance.
To define the sample variance, we define first the functional for the plug-in estimator, that in this case will be:

$$Var(F) = \int (x - \mu_F)^2 \, dF(x), \qquad \mu_F = \int x \, dF(x)$$

Now, the sample variance, for $x_1, \dots, x_n \overset{iid}{\sim} F$: $T(F_n) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

$$E\Big[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\Big] \overset{\text{linearity}}{=} \frac{1}{n}\sum_{i=1}^{n} E[(x_i-\bar{x})^2] \overset{\text{solving squares}}{=} \frac{1}{n}\sum_{i=1}^{n} E[x_i^2 - 2x_i\bar{x} + \bar{x}^2] =$$

decomposing $\bar{x}$

$$= \frac{1}{n}\sum_{i=1}^{n} E\Big[x_i^2 - 2x_i\Big(\frac{1}{n}\sum_{j=1}^{n} x_j\Big) + \Big(\frac{1}{n}\sum_{j=1}^{n} x_j\Big)^2\Big] =$$

$$= \frac{1}{n}\sum_{i=1}^{n} E\Big[x_i^2 - \frac{2}{n}x_i^2 - \frac{2}{n}\sum_{j\neq i}^{n} x_i x_j + \frac{1}{n^2}\sum_j x_j^2 + \frac{1}{n^2}\sum_j\sum_{k\neq j} x_j x_k\Big]$$ expectancies

$$= \frac{1}{n}\sum_{i=1}^{n} \frac{n-2}{n}(\sigma^2+\mu^2) - \frac{2}{n}(n-1)\mu^2 + \frac{1}{n^2}n(\sigma^2+\mu^2) + \frac{1}{n^2}n(n-1)\mu^2$$

$$= \frac{n-1}{n}\sigma^2 \neq \sigma^2 \Rightarrow \text{Is biased!}$$

(b) If we denote as S_n the sample variance:

$$S_n = \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu+\mu-\overline{X})^2 =$$

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

Now, using WLLN, $\overline{X} \xrightarrow{P} \mu$ and $\mathbb{E}[(X_i-\mu)^2]=\sigma^2$ so $\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 \xrightarrow{P} \sigma^2$

therefore:

$$S_n = \underbrace{\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2}_{\downarrow P\ \sigma^2} + \underbrace{(\mu-\overline{X})^2}_{\downarrow P\ 0} + \underbrace{2(X_i-\mu)(\mu-\overline{X})}_{\downarrow P\ 0} \xrightarrow[P]{} \sigma^2$$

It's consistent!!

(c) The estimator $S_{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})$ is unbiased because

$$S_{n-1} = \frac{n}{n-1}S_n \Rightarrow \mathbb{E}[S_{n-1}] = \mathbb{E}\left[\frac{n}{n-1}S_n\right] = \frac{n}{n-1}\mathbb{E}[S_n] = \frac{\cancel{n}}{\cancel{n-1}}\cdot\frac{\cancel{n-1}}{\cancel{n}}\sigma^2 = \sigma^2$$