Stat 135 Section 1
06/23/2021

## Lab 2 Practice problems

1. SD of combined list

A population consists of $n$ men and $n$ women. The heights of the men have an average of $\mu_m$ and an SD of $\sigma_m$. The heights of the women have an average of $\mu_w$ and an SD of $\sigma_w$. Find a formula for the SD of the heights of all $2n$ people, in terms of $\mu_m, \mu_w, \sigma_m$, and $\sigma_w$.

**Solution:** Let $X_1, X_2, \ldots, X_n$ be the heights of men, and $X_{n+1}, \ldots, X_{2n}$ be the heights of women. $\therefore \mu_{M+W} = \frac{n\mu_M + n\mu_W}{2n}$. Let $\sigma_{M+W}$ be the SD of the combined list of men and women. Then,

$$\begin{aligned}
\sigma^2_{M+W} &= E(X^2) - E(X)^2 = \frac{1}{2n}\sum_{i=1}^{2n} X_i^2 - \mu^2_{M+W} \\
&= \frac{1}{2n}(\sum_{i=1}^{n} X_i^2 + \sum_{i=n+1}^{2n} X_i^2) - \mu^2_{M+W} \text{ where} \\
\sum_{i=1}^{n} X_i^2 &= n\sigma_M^2 + n\mu_M^2 \text{ and } \sum_{i=n+1}^{2n} X_i^2 = n\sigma_W^2 + n\mu_W^2 \\
\therefore \sigma^2_{M+W} &= \frac{1}{2n}(n(\sigma_M^2 + \mu_M^2) + n(\sigma_W^2 + \mu_W^2)) - (\frac{1}{2}(\mu_M + \mu_W))^2 \\
&= \frac{\sigma_M^2 + \mu_M^2}{2} + \frac{\sigma_W^2 + \mu_W^2}{2} - \frac{(\mu_M + \mu_W))^2}{4}
\end{aligned}$$

Two populations are surveyed with simple random samples. A sample of size $n_1$ is used for population I, which has a population standard deviation $\sigma_1$; a sample of size $n_2 = 2n_1$ is used for population II, which has a population standard deviation $\sigma_2 = 2\sigma_1$. Ignoring finite population corrections, in which of the two samples would you expect the estimate of the population mean to be more accurate?

**Solution:** Since we are ignoring finite population corrections, we have

$$Var(\bar{X}_1) \approx \frac{\sigma_1^2}{n_1} \quad \text{and} \quad Var(\bar{X}_2) \approx \frac{\sigma_2^2}{n_2} = \frac{4\sigma_1^2}{2n_1} = 2Var(\bar{X}_1)$$

$$\therefore SD(\bar{X}_2) = \sqrt{2}SD(\bar{X}_1) > SD(\bar{X}_1).$$

Therefore we expect the estimate of Population I to be more accurate because it has a smaller standard deviation.

2. Method of Moments (MoM)

Consider the i.i.d. random variables $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$ with both $\mu$ and $\sigma^2$ unknown. Find $\hat{\mu}_{MM}$ and $\hat{\sigma}^2_{MM}$, i.e. the MoM estimators of the mean and the variance.
**Solution:**

$$\begin{cases} \mu_1 = \mu \\ \mu_2 = E(X_i^2) = \sigma^2 + \mu^2 \end{cases} \overset{\text{Rearrange}}{\Longrightarrow} \begin{cases} \mu = \mu_1 \\ \sigma^2 = \mu_2 - \mu_1^2 \end{cases}$$

The stimators of $\mu_1$ and $\mu_2$:

$$\hat{\mu_1} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}, \ \hat{\mu_2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Plug in and get

$$\hat{\mu}_{MM} = \bar{X}, \ \ \hat{\sigma}^2_{MM} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 \text{ (which in class is proved to be a biased estimator)}$$

3. Expectation and p.d.f of a *function* of a random variable, Y = g(X)

Consider a random variable $X \sim \text{Uniform}(e, e^2)$ and define the new random variable $Y = \ln X$
(a): Compute $E(Y) = E(g(X))$

**Solution:** By uniform distribution we get

$$f_X(x) = \begin{cases} \frac{1}{e^2-e} & x \in (e, e^2) \\ 0 & x \notin (e, e^2) \end{cases}$$

Then we have

$$\begin{aligned} E(g(X)) &= \int g(x) f_X(x) dx = \int_e^{e^2} \ln x * \frac{1}{e^2 - e} dx \\ &\overset{\text{by parts}}{=} \frac{1}{e^2 - e}[x \ln x - x]_{x=e}^{x=e^2} \\ &= \frac{[(2e^2 - e^2) - (e - e)]}{e^2 - e} = \frac{e^2}{e^2 - e} \end{aligned}$$

(b): Calculate $E(Y)$ by computing the p.d.f of Y. To do so, the initial step is to figure out the c.d.f of Y, by noticing that:

$$F_Y(y) = P(Y \leq y) = P(\ln X \leq y) \overset{(*)}{=} P(X \leq e^y) = F_X(e^y). \tag{1}$$

(*) By applying the exponential function, which is *monotone increasing*.

**Solution:** by (1), using chain rule, we take the derivative of $F_X(e^y)$

$$f_Y(y) = \frac{d}{dy}[F_X(e^y)] = F'_X(e^y) * e^y = f_X(e^y) * e^y \quad \text{where}$$

$$f_X(x) = \begin{cases} \frac{1}{e^2-e} & x \in (e, e^2) \\ 0 & x \notin (e, e^2) \end{cases} \quad \Rightarrow f_X(e^y) = \begin{cases} \frac{1}{e^2-e} & e^y \in (e, e^2) \\ 0 & e^y \notin (e, e^2) \end{cases} \quad \Rightarrow \begin{cases} \frac{1}{e^2-e} & y \in (1,2) \\ 0 & y \notin (1,2) \end{cases}$$

$$\therefore f_Y(y) \begin{cases} \frac{e^y}{e^2-e} & y \in (1,2) \\ 0 & y \notin (1,2) \end{cases}$$

$E(Y) = \frac{1}{e^2-e} \int_1^2 ye^y dy$ and by chain rule this is equal to $\frac{1}{e^2-e}(2e^2 - e - (e^2 - e)) = \frac{e^2}{e^2-e}$