DS 203: Assignment 3

~ Shubham Lohiya, 18D100020

Ex-1 To show : $Cov(X_1+X_2, X_3) = Cov(X_1, X_3) + Cov(X_2, X_3)$

LHS: $Cov(X_1+X_2, X_3) = E[(X_1+X_2)(X_3 - E[X_3])]$

$= E[X_1(X_3 - E[X_3]) + X_2(X_3 - E[X_3])]$

$= E[X_1(X_3-E[X_3])] + E[X_2(X_3-E[X_3])] = \boxed{Cov(X_1,X_3)+Cov(X_2,X_3) = RHS}$

Ex-2 given: $X_1 \dots X_n \sim$ i.i.d. s.t. $\mu = E[X_i]$ & $\sigma^2 = Var(X_i)$

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

$$E[Y] = E\left[\sum_{i=1}^{n}\frac{(X_i-\mu)}{\sqrt{n\sigma^2}}\right] = \frac{1}{\sqrt{n\sigma^2}}\sum_{i=1}^{n}(E[X_i]-\mu)$$

$$= \frac{1}{\sqrt{n\sigma^2}}\sum_{i=1}^{n}(\mu-\mu) = \boxed{0}$$

$$Var(Y) = Var\left(\frac{\sum_{i=1}^{n}X_i - n\mu}{\sqrt{n\sigma^2}}\right)$$

$$= \frac{Var\left(\sum_{i=1}^{n}X_i\right)}{n\sigma^2} = \frac{\sum_{i=1}^{n}Var(X_i)}{n\sigma^2}$$

$$= \frac{n\sigma^2}{n\sigma^2} = \boxed{1}$$

useful Props:

$Var(X+a) = Var(X)$

$Var(X/a) = \frac{1}{a^2}Var(X)$

$Var(X_1+X_2+\dots X_n) = Var(X_1)+\dots Var(X_n)$ as $X_1, X_2 \dots X_n$ are i.i.d.

Ex-3 given 25 observations from a population of 25 values.

Results of analysis in Excel

Original Data					
	104	109	111	109	87
	86	80	119	88	122
	91	103	99	108	96
	104	98	98	83	107
	79	87	94	92	97
Sorted					
	79	80	83	86	87
	87	88	91	92	94
	96	97	98	98	99
	103	104	104	107	108
	109	109	111	119	122
Sample Mean	98.04				
Sample Variance	133.707				
Sample Std. Dev	11.5632				
Range	43				
Median	98				
Lower Quartile	87.5				
Upper Quartile	107.5				

Formulas & Calculation

1) Mean $= \sum_{i=1}^{25} X_i/25 = \frac{2451}{25} = 98.04$

2) S. Variance $= \sum_{i=1}^{25}\frac{(x_i-98.04)^2}{24} = \frac{3208.96}{24}$

$= 133.707$

3) S. Dev $= \sqrt{\text{S.Variance}} = 11.5632$

4) Range $=$ Max$-$Min $= 122-79$

$= 43$

5) 6) 7) Calculation of Quartiles using convention is displayed above. Based on formulas in slides, Lower Quartile is 87 & Upper Quartile is 108.

8) 95% confidence intervals for population mean

→ Using Student's t- Statistic (Assuming data is sampled from Gaussian dist.)

(Not using CLT here as sample size is less than 30)

We have $T_n = \frac{\hat{\mu} - \mu}{\tilde{S}_n} \sim t_{n-1} = t_{24}$

$\hat{\mu}$ – sample mean
$\tilde{S}_n$ – sample std. dev
t_{n-1} – T-dist with n-1 degrees of freedom.

Here T_n is called the t-statistic

for a ≥95% confidence interval, we want

$$P(|T_n| > \varepsilon)) \leq 0.05$$

By looking up the t-table, we get $\varepsilon \geq 2.064$

Thus for tighter bound: $\varepsilon = 2.064$

⇒ Confidence interval: $|\hat{\mu} - \mu| \leq 2.064 \times \tilde{S}_n = 2.064 \times \frac{11.5623}{\sqrt{25}}$

$$\simeq 4.773$$

⇒ $\mu \in [93.267, 102.813]$ with at least 95% confidence

Source for lookup table: https://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf

Ex-4 Given, Population I : Sample Size n_1, unknown mean μ_1, Variance σ_1^2
Population I : Sample Size n_2, unknown mean μ_2, Variance σ_2^2

We know from CLT that Sample means $\bar{X}_1$ & $\bar{X}_2$ follow:

$$\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \quad \& \quad \bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

But we have $n_2 = 2n_1$ & $\sigma_2 = 2\sigma_1$

$$\Rightarrow \bar{X}_2 \sim N\left(\mu_2, \frac{4\sigma_1^2}{2n_1}\right) = N\left(\mu_2, \frac{2\sigma_1^2}{n_1}\right)$$

We expect the sample mean to be a more accurate population mean estimator when it comes from a distribution with lower variance.

Since we have seen that $Var(\bar{X}_1) < Var(\bar{X}_2)$, we expect the estimate for $\bar{X}_1$ to be more accurate.

Ex-5 We have X_1, X_2 sampled from $N(0, \sigma^2)$

First order statistic $X_{(1)} = \min(X_1, X_2)$

Let $Z = \min(X_1, X_2)$. Let's look at the CDF of Z

$$P(Z \leq z) = 1 - P(Z > z) = 1 - P(X_1 > z)P(X_2 > z) = 1 - (1 - \Phi(z))^2$$

where $\Phi(x) = P(X_i \leq x)$. Thus we have $F_Z = 1 - (1 - \Phi(z))^2$

$$f_Z = \frac{\partial}{\partial z} F_Z = 2(1 - \Phi(z))\Phi'(z)\,dz = 2f_X(z)(1 - \Phi(z))$$

Now that we have f_Z, We know $E[Z] = \int_{-\infty}^{\infty} x f_Z(x)\,dx$

$$= \int_{-\infty}^{\infty} x \; 2f_X(x)(1 - \Phi(x))\,dx \quad ①$$

Substitute $p = \Phi(x)$

$dp = f_X(x)\,dx$

Limits: lower $\rightarrow 0$

upper $\rightarrow 1$

$$= \int_0^1 2\Phi^{-1}(p)(1 - p)\,dp$$

We have $f_{Beta(1,2)}(x) = \frac{(1-x)}{0!\,1!/2!}$

$= 2(1-x)$

Thus $E_Z[Z] = E_{p \sim Beta(1,2)}\left[\Phi^{-1}(p)\right]$

Crude Approximation:

$$\Rightarrow E_Z[Z] \simeq \Phi^{-1}\left(E_{p \sim Beta(1,2)}[p]\right)$$

We know from delta method, $E[f(x)] \simeq f(E(x))$ for a "well-behaved" function. (which Φ^{-1} is)

$$= \Phi^{-1}\left(\frac{1}{3}\right) = \sigma \cdot F^{-1}_{X \sim N(0,1)}(1/3) \simeq -0.431\sigma$$

Since $X \sim N(0, \sigma^2)$, We have the relation $\Phi^{-1}(p) = \sigma \cdot F^{-1}_{N(0,1)}(p)$

This is only an approximation, and a better estimate can be achieved using a Monte Carlo simulation

```python
from tqdm import tqdm
import numpy as np
from scipy.stats import norm

expectation = norm.ppf(np.random.beta(1,2))
for i in tqdm(range(1, 1000000)):
    expectation = (expectation*i + norm.ppf(np.random.beta(1,2)))/(i+1)

expectation
```

```
100%|██████████████████████████████| 999999/999999 [02:38<00:00, 6309.06it/s]
```

```
-0.5649986107955756
```

Thus by doing a monte-carlo simulation & looking at 1 million observations, we estimate the following

$$\boxed{E[Z] \simeq -0.565\sigma}$$

If one does the integral in ①, it yields $-\frac{\sigma}{\sqrt{\pi}}$ which is very close to our approximation

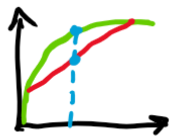
Ex-6 $\quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$ is an unbiased estimator of σ^2

Sample std. dev $= \sqrt{S^2} = S$. To check bias in S:

$\text{Bias}(S) = E[S] - \sigma$

Let's assume $\sigma > 0$ as $\sigma = 0$ case is trivial & all sample-based methods will yield exact 0 estimates.

As $f(x) = \sqrt{x}$ is a strictly concave function, we have using the Jensen's inequality that:



$$E[\sqrt{S^2}] < \sqrt{E[S^2]}$$

but we know $E[S^2] = \sigma^2$ as it is an unbiased estimator.

$$\Rightarrow \quad E[S] < \sigma$$

$$\boxed{\begin{array}{l} \text{i.e. } E[S] - \sigma < 0 \Rightarrow \text{Bias}(S) < 0 \\ \text{As the bias is non-zero, } S \text{ is a biased estimator of} \\ \text{sample standard deviation.} \end{array}}$$

Ex-7 $\quad P = \begin{cases} \frac{3}{5}\theta & , x=0 \\ \frac{2}{5}\theta & , x=1 \\ \frac{3}{5}(1-\theta) & , x=2 \\ \frac{2}{5}(1-\theta) & , x=3 \end{cases}$

Observation counts (total 10):
$n_0 = 2$
$n_1 = 3$
$n_2 = 3$
$n_3 = 2$

As the observations are independent, we have the likelihood:

$$L(\theta) = \left(\frac{3}{5}\right)^{n_0+n_2}\left(\frac{2}{5}\right)^{n_1+n_3}\theta^{n_0+n_1}(1-\theta)^{n_2+n_3}$$

$$= \boxed{\left(\frac{6}{25}\right)^5 \theta^5 (1-\theta)^5}$$

MLE: $\quad \ell(\theta) = \ln L(\theta) = 5\ln\left(\frac{6}{25}\right) + 5\ln\theta + 5\ln(1-\theta)$

$$\ell'(\hat{\theta}) = \frac{5}{\hat{\theta}} - \frac{5}{1-\hat{\theta}} = 0 \Rightarrow \boxed{\hat{\theta} = \frac{1}{2}}$$

$$\ell''(\hat{\theta}) = -\frac{5}{\hat{\theta}^2} - \frac{5}{(1-\hat{\theta})^2} < 0 \Rightarrow \hat{\theta} \text{ is indeed a maxima}$$

Ex-8 $f(x|\theta) = \begin{cases} \frac{\theta}{(1+x)^{\theta+1}} & , 0 < x < \infty \\ 0 & o/w \end{cases} \quad , \quad \theta > 0$

For the generalized case, assume n independent observations from this distribution: $X_1, X_2 \cdots X_n$

$$L(\theta) = \prod_{i=1}^{n} \frac{\theta}{(1+X_i)^{\theta+1}} \mathbb{1}(X_i > 0)$$

Assuming $\min_i (X_i) > 0$:

$$\Rightarrow \ell(\theta) = n \ln \theta - (\theta+1) \sum_{i=1}^{n} \ln(1+X_i)$$

MLE: $0 = \ell'(\hat{\theta}) = \frac{n}{\hat{\theta}} - \sum_{i=1}^{n} \ln(1+X_i)$

$$\Rightarrow \boxed{\hat{\theta} = \frac{n}{\sum_{i=1}^{n} \ln(1+X_i)}} \; ; \; \ell''(\hat{\theta}) = -\frac{n}{\hat{\theta}^2} < 0$$

$\Rightarrow$ maxima

for $n=1$ case (single observation):

$L(\theta) = \frac{\theta}{(1+x)^{\theta+1}} \mathbb{1}(x > 0)$ & MLE: $\hat{\theta} = \frac{1}{\ln(1+x)}$

Ex-9 $f(x|\theta) = \theta x^{(\theta-1)} \mathbb{1}(0 \leq x \leq 1)$

For n observations $X_1, X_2 \cdots X_n$:

$$\ell(\theta) = \ln(L(\theta)) = n \ln \theta + (\theta-1) \sum_{i=1}^{n} \ln(x_i)$$

MLE: $0 = \ell'(\hat{\theta}) = \frac{n}{\hat{\theta}} + \sum_{i=1}^{n} \ln(X_i)$

$$\Rightarrow \boxed{\hat{\theta} = \frac{-n}{\sum_{i=1}^{n} \ln(X_i)}}$$

To calculate estimator variance, we can use Fisher information:

Fisher Information $(I(\theta)) = -E[\ell''(\theta)]$

where log likelihood is w.r.t. 1 observation.

$$\Rightarrow \ell(\theta) = \ln(\theta) + (\theta-1) \ln(X)$$

$$\ell'(\theta) = \frac{1}{\theta} + \ln(X) \quad \& \quad \ell''(\theta) = -\frac{1}{\theta^2}$$

Thus $I(\theta) = E[-l''(\theta)] = \frac{1}{\theta^2}$

We know that MLE is an asymptotically normal estimator with asymptotic variance equal to inverse Fisher info

ie. $\sqrt{n}\,(\hat{\theta} - \theta) \sim N(0, \frac{1}{I(\theta)})$, for large n

$$\Rightarrow \quad \hat{\theta} \sim N\left(\theta, \frac{\theta^2}{n}\right), \text{ for large } n$$

Thus for large n, $Var(\hat{\theta}) \simeq \frac{\theta^2}{n}$

$$\Rightarrow \boxed{\lim_{n\to\infty} Var(\hat{\theta}) = \lim_{n\to\infty} \frac{\theta^2}{n} = 0}$$

Thus with more number of samples, our variance converges to 0 in the limit, & hence MLE is a consistent estimator of θ.

Ex-10 Given $Y_1, \dots Y_m \sim$ Bernoulli (θ)

Hypotheses: $H_0: \theta \leq \theta_0 \qquad H_1: \theta > \theta_0$

Likelihood ratio test: Calc $\lambda = \frac{\sup_{\theta\in\Theta_0}(L(\theta|x))}{\sup_{\theta\in\{\Theta_0,\Theta_0^c\}}(L(\theta|x))}$

We know $L(\theta)$ for bernoulli trials has single peak at MLE.

$\hat{\theta}_{MLE} = \frac{\Sigma Y_i}{m}$ is a well known result

We know $\sup_{\theta\in\{\Theta_0,\Theta_0^c\}}(L(\theta|x)) = L(\hat{\theta}_{MLE}|x)$

Case 1: $\hat{\theta}_{MLE} \leq \theta_0 \Rightarrow \sup_{\theta\in\Theta_0}(L(\theta|x)) = L(\hat{\theta}_{MLE}|x)$

Thus $\lambda = 1$ & since we reject when $\exists x$ s.t. $\lambda(x) \leq c$ for some $c \in (0,1)$, we can never reject in this case.

Case 2: $\theta_0 < \hat{\theta}_{MLE} \Rightarrow \sup_{\theta\in\Theta_0}(L(\theta|x)) = L(\theta_0|x)$

Let $y = \Sigma Y_i$

Thus, $\lambda(x) = \left(\frac{\theta_0}{y/m}\right)^{y} \left(\frac{1-\theta_0}{1-y/m}\right)^{m-y}$

For rejection we need $\lambda(x) \leq c$ for some threshold c

substitute $t = \frac{y}{m\theta_0}$. Thus we are inspecting only for $t > 1$ in this case.

$$\lambda(t) = \left(\frac{1}{t}\right)^{m\theta_0 t} \cdot \left(\frac{1-\theta_0}{1-\theta_0 t}\right)^{m-m\theta_0 t}$$

Let's consider function $f(t) = \ln(\lambda(t))$

$$f(t) = -m\theta_0 t \ln t + (m - m\theta_0 t)(\ln(1-\theta_0) - \ln(1-\theta_0 t))$$

$$f'(t) = -m\theta_0(1+\ln t) - m\theta_0(\ln(1-\theta_0) - \ln(1-\theta_0 t)) + (m - m\theta_0 t)\left(\frac{\theta_0}{1-\theta_0 t}\right)$$

$$= \underbrace{-m\theta_0 \ln t}_{<0\ \forall t>1} + \underbrace{m\theta_0 \ln\left(\frac{1-\theta_0 t}{1-\theta_0}\right)}_{<0\ \forall t>1}$$

$\forall t > 1$, $\frac{1-\theta_0 t}{1-\theta_0} < 1$

$\Rightarrow f'(t) < 0 \quad \forall t > 1 \Rightarrow \lambda'(t) < 0 \quad \forall t > 1$

Thus $\lambda(t)$ is a strictly decreasing func^n for $t > 1$

Thus for every $c \in (\theta_0^m, 1)$, we can find a positive value t' s.t. $t' > 1$ & $\lambda(t') = c$ (using IVT)

$\inf_y(\lambda(y)) = \theta_0^m$ as $\frac{1-\theta_0}{1-y/m} > 1$ for this case

Thus we'll have $\lambda(t) \leq c \quad \forall t \geq t'$. So, we end up rejecting H_0 for all $t \geq t'$

$$t \geq t' \Rightarrow \frac{y}{m\theta_0} \geq t' \Rightarrow y \geq m\theta_0 t'$$

as $t' > 1$ we have $y \geq m\theta_0 t' > m\theta_0 + k$, for some $k > 0$

i.e. $y > m\theta_0 + k = b$ (, say)

as $y = \sum_{i=1}^{m} Y_i$, we have rejection region of the form:

$$\boxed{\sum_{i=1}^{m} Y_i > b}$$