MECENG 276
HW3

Problem 1 - 3

Problem 1

Biased or unbiased ?

→ Unbiassed if

$$\text{Bias}[\hat{\theta}] = E[\hat{\theta}] - \theta = 0$$

for all values of θ

→ Asymptotic unbiased if

$$\lim_{N \to \infty} \text{Bias}[\hat{\theta}] = 0$$

Where $\hat{\theta}$ is the estimate of the real parameters θ

Lets see if the estimators are or not

1) $\bar{Y}_A = Y_1$

$\rightarrow \text{Bias}[\bar{Y}_A] = E[Y_1] - \mu_Y$

(all Y_i iid) $\rightarrow = \mu_Y - \mu_Y = 0$

This will (probably) not be 0, and $\bar{Y}_A$ is unbiased

2) $$\bar{Y}_B = \frac{Y_1 + 2Y_N}{3}$$

$\rightarrow \text{Bias}[\bar{Y}_B] = \frac{1}{3}(\mu_Y + 2\mu_Y) - \mu_Y$

$= 0$

$\bar{Y}_B$ is unbiased

3) $\bar{Y}_c = \frac{1}{M}\sum_{i=1}^{M} Y_i \qquad 1 < M < N$

$$\text{Bias}[\bar{Y}_c] = \frac{1}{M}\sum^{M} Y_i - \mu_Y$$

$$= \frac{\sum^{M} \mu_Y}{M} - \mu_Y = 0$$

$\bar{Y}_c$ is unbiased

4) $\bar{Y}_0 = \frac{1}{N}\sum^{N} Y_i$

$$\text{Bias}[\bar{Y}_0] = E[\bar{Y}_0] - \mu_Y = \mu_Y - \mu_Y = 0$$

$\bar{Y}_0$ is unbiased

Variance of estimators

1) $Var[\bar{Y}_A] = Var[Y_1] = \sigma_Y^2$

2) $Var[\bar{Y}_B] = \left(\frac{1}{3}\right)^2 \sigma_Y^2 + \left(\frac{2}{3}\right)^2 \sigma_Y^2 = \frac{5}{9} \sigma_Y^2$

3) $Var[\bar{Y}_C] = \frac{1}{M^2} M \cdot \sigma^2 = \frac{1}{M} \sigma^2$

4) $Var[\bar{Y}_D] = \frac{1}{N^2} N \cdot \sigma^2 = \frac{1}{N} \sigma^2$

Estimators ranked from least to highest variance :

1) $\bar{Y}_D$

2) $\bar{Y}_C$

3) $\bar{Y}_B$

4) $\bar{Y}_A$

Problem 2

Confidence interval

$\mathcal{D} \overset{iid}{=} \{Y_i\}_{100}$ $\bar{Y} = 92$ $\sigma_Y^2 = 64$

a) Find a 95% confidence interval for the yield strength

$$P(|\hat{\theta} - \theta| \leq p) = \gamma$$

where the c.i. has the form $[\hat{\theta} - p, \hat{\theta} + p]$ with certainty γ

$$\bar{Y}_{100} \sim \mathcal{N}(\mu_Y, \sigma_Y^2/100)$$

$$F_{\bar{Y}_{100}}(\mu_Y + p) - F_{\bar{Y}_{100}}(\mu_Y - p) = \gamma$$

(*) $F_{\bar{Y}_s}(p) - F_{\bar{Y}_s}(-p) = \gamma$

$\bar{Y}_s = \bar{Y}_{100} - \mu_Y$ and $\bar{Y}_s \sim \mathcal{N}(0, \sigma_Y^2/100)$

$$\rightarrow F_{\bar{Y}_s}(p) = 1 - F_{\bar{Y}_s}(-p)$$

(*) $\rightarrow 1 - 2F_{\bar{Y}_s}(-\rho) = \gamma$

$\rightarrow F_{\bar{Y}_s}(-\rho) = \frac{1-\gamma}{2}$

$\rightarrow -\rho = F_{\bar{Y}_s}^{-1}\left(\frac{1-\gamma}{2}\right)$

$\rightarrow \rho = \left| F_{\bar{Y}_s}^{-1}\left(\frac{1-\gamma}{2}\right) \right|$

$$Z = \frac{\bar{Y}_{100} - \mu_Y}{\sigma_Y / \sqrt{100}} \sim \mathcal{N}(0,1)$$

$$F_{\mathcal{N}}\left(\frac{\rho}{\sigma_Y/10}\right) - F_{\mathcal{N}}\left(-\frac{\rho}{\sigma_Y/10}\right) = \gamma$$

$$\rightarrow \rho = \frac{\sigma_Y}{10} \left| F_{\mathcal{N}}^{-1}\left(\frac{1-\gamma}{2}\right) \right|$$

$\sigma_Y = 8 \quad \gamma = 0.95 \rightarrow \rho = -1.96 \cdot \frac{\sigma_Y}{10} = -1.568$

The confidence interval of 95% is therefore $\mu_Y \pm \rho$

$[92 - 1.568,\ 92 + 1.568]$

b) Find the 99% c.i. ...

$$\rho = \frac{\sigma_Y}{\sqrt{N}} \left| F_N^{-1}\left(\frac{1-\gamma}{2}\right) \right|$$

$\rightarrow \sigma_Y = 8 \quad N = 100 \quad \gamma = .99$

$$\rho = 0.8 \cdot |(-2.576)|$$

$$[92 - 2.0608,\ 92 + 2.0608]$$

c) How many measurments to be within 0.5 with 95 confidence

$\rho = 0.5 \quad N?$

$$0.5 = \frac{\sigma_Y}{\sqrt{N}} \left| F_N^{-1}\left(\frac{1-\gamma}{2}\right) \right|$$

$$\rightarrow N = 4\sigma_Y^2 \left(F_N^{-1}(0.025)\right)^2 = 4 \cdot 8^2 \cdot 1.96^2 \approx 984$$

d) —"— 99% confidence

$$\ldots = 4 \cdot 8^2 \cdot 2.576^2 = 1\ 698.758\ 656$$

Problem 3

7 degree freedom

a) $$P = \frac{\sigma_Y}{\sqrt{N}} \left| F_\nu^{-1}\left(\frac{1-\gamma}{2}\right) \right| = \frac{0.14}{\sqrt{8}} \cdot 2.998$$

$$\approx 0.1484$$

$$[78.37 - 0.1484,\ 78.37 + 0.1484]$$

b) We assume that the true distribution is gaussian.

Problem 4

Code:

```
# Problem 4

# Get num of zeros and ones
survey = open("./survey.txt")
n = 0
y = 0
for i in survey.read().split("\n"):
    if i.__contains__("0"):
        n+=1
    else:
        y+=1

mean = (y*1+n*0)/(y+n)
std_dev = np.sqrt(((n*(0-mean)**2) + (y*(1-mean)**2))/(y+n))

print("Sample mean: ", mean)
print("Std. dev.: ", std_dev.round(4))

# We can use the CLT to make the gaussian assumption. Because the sample size is large enough.

# Find the Confidence Interval of the mean with 95% confidence

P = (std_dev/np.sqrt(n+y)) *np.abs(1.960)
confidence_interval = [(mean-P).round(4), (mean+P).round(4)]

print("Mean with c: 95%: ", confidence_interval)

N = ((std_dev/(P/2)) *np.abs(1.960))**2

print("Num of surveys to half ci: ", N)
```

Result:

```
Sample mean:  0.555
Std. dev.:  0.497
Mean with c: 95%:  [0.5063, 0.6037]
Num of surveys to half ci:  1600.0
```