PROBABILITY AND STATISTICS - 2
ASSIGNMENT - 1

2. (i) Sample mean is $\bar{X} = \frac{\sum x_i}{n}$

$$\Rightarrow E[\bar{X}] = E\frac{[\sum X_i]}{n}$$
$$= \frac{\sum E[X_i]}{n}$$
$$= \frac{\sum \mu}{n}$$
$$= \frac{n\mu}{n}$$
$$= \boxed{\mu}$$

(ii) For $\sum X_i$, $var[\sum X_i] = \sum var[X_i]$
$$= \sum \sigma^2$$
$$= n\sigma^2$$

(Since the variables are independent and identically distributed)

$$\rightarrow var[\bar{X}] = var\left[\frac{\sum X_i}{n}\right]$$
$$= \frac{1}{n^2} var[\sum X_i]$$
$$= \frac{1}{n^2}(n\sigma^2)$$
$$= \boxed{\frac{\sigma^2}{n}}$$

(ii) $S^2 = \frac{1}{n-1} \Sigma (x_i - \bar{x})^2$

$\Rightarrow S^2 = \frac{1}{n-1} [\Sigma x_i^2 - n\bar{x}^2]$

→ Taking expectations on both sides, noting that for a random variable Y,
$E[Y^2] = var[Y] + (E[Y])^2$

$\Rightarrow E[S^2] = \frac{1}{n-1} (\Sigma E[x_i^2] - nE[\bar{x}^2])$

$= \frac{1}{n-1} \left\{ \Sigma (\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right) \right\}$

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

$= \frac{1}{n-1} \{ (n-1)\sigma^2 \}$

$= \boxed{\sigma^2}$

(iii) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

→ Using the χ^2 results to investigate first and second order moments of S^2 when sampling from a normal popn, and the mean and variance of χ^2_k are k and $2k$ respectively:

$E\left(\frac{(n-1)s^2}{\sigma^2}\right) = n-1$

$\Rightarrow \boxed{E[S^2] = \sigma^2}$

$$\Rightarrow \quad var\left[\frac{(n-1)s^2}{\sigma^2}\right] = var\left[\chi^2_{n-1}\right]$$

$$= 2(n-1)$$

$$\Rightarrow \quad var[s^2] = \left[\frac{\sigma^2}{(n-1)}\right]^2 \cdot 2(n-1)$$

$$= \boxed{\frac{2\sigma^4}{n-1}}$$

(iv) $n = 10$

$\sigma^2 = 100$

$$P(-0.05(100) < s^2 < 0.05(100)) = ? \quad [\because E(s^2) = \sigma^2]$$

$$\Rightarrow \quad P(-50 < s^2 < 50) = P\left(\frac{(9)\,-0.5\,\sigma^2}{\sigma^2} < \frac{9s^2}{\sigma^2} < \frac{(9)\,0.5\,\sigma^2}{\sigma^2}\right)$$

$$= P\left(-0.5 < \chi^2_9 < 0.5\right)$$

$$= P\left(\chi^2_9 < 0.5\right) - P\left(\chi^2_9 < -0.5\right)$$

19 $X \sim Poisson(\theta)$

$$\frac{X - \theta}{\sqrt{\theta}} \sim \text{approx. } N(0,1)$$

$$\rightarrow \quad P(X = 200) = P(199.5 < X < 200.5]$$

$$= P\left[\frac{199.5-\theta}{\sqrt{\theta}} < z < \frac{200.5-\theta}{\sqrt{\theta}}\right]$$

$$\Rightarrow \quad 95\% \quad CI = \theta \pm 1.96\sqrt{\theta}$$

Date

3. let the number of claims be X.

→ $X \sim Bin(n, p)$
$n = 180$

(i) $L(p) = \prod_{i=0}^{4} {}^{180}C_0 p^0 q^{180} \times {}^{180}C_1 p^1 q^{179}$
$\times {}^{180}C_2 p^2 q^{178} \times {}^{180}C_3 p^3 q^{177}$
$\times {}^{180}C_4 p^4 q^{176}$

$= \text{constant} \times p^0 q^{180} \times p^1 q^{179} \times p^2 q^{178}$
$\times p^3 q^{177} \times p^4 q^{176}$

→ $\log L(P) = \text{cons.} + 0 + 180 \log p + 1 \log p + 179 \log q$
$+ 2 \log p + 178 \log q + 3 \log p + 177 \log q$
$+ 4 \log p + 176 \log q$

$= c + 10 \log p + 890 (1-p)$

→ $\frac{d}{dp} \log L[P] = \frac{-890}{\log p}$

Date

4. P.D.F of random variable X,

$$f(x) = \frac{c\beta^3}{(x+\beta)^4} \quad : x > 0, \beta > 0$$

(i) $E(X) = \beta/2$

$E(\hat{X}) = \bar{X}$

$\therefore \boxed{\hat{\beta} = 2\bar{x}}$

(ii) Bias $= E(\hat{\beta}) - \beta = E[2\bar{x}] - \beta$

$= 2E[\bar{x}] - \beta$

$= \frac{2}{n} \frac{n\beta}{2} - \beta$

$= \beta - \beta = 0.$

$\rightarrow$ MSE $= V(\hat{\beta}) + (\text{Bias})^2$

$= V(\hat{\beta})$

$= V(2\bar{x})$

$= 4V(\bar{x})$

$= 4V\left[\frac{\Sigma x_i}{n}\right]$

$= \frac{4}{n^2} V[\Sigma x_i]$

$= \frac{4}{n^2} \times \frac{3\beta^2 n}{4}$

$= \frac{3\beta^2}{n}$

Here, $\hat{\beta}$ is constant.

(iii) $\hat{\beta} = b\bar{X}_n$

$$E(\hat{\beta}_2) = E(b\bar{X})$$
$$= b\,E[\bar{X}]$$
$$= \frac{b}{n}\,\frac{\beta}{2} \times n$$
$$= \frac{\beta b}{2}$$

$\rightarrow$
$$V(\hat{\beta}_2) = V(b\bar{X})$$
$$= b^2\,V(\bar{X})$$
$$= \frac{b^2}{n^2} \times \frac{3}{n}\,\beta^2 \times n$$
$$= \boxed{\frac{3}{n}\,\frac{b^2\beta^2}{n}}$$

$\rightarrow$ $MSE = V(\beta_2) + (Bias)^2$
$$= \frac{3b^2\beta^2}{4n} + \left(\frac{b\beta - \beta}{2}\right)^2$$

5. For $U(a,b)$, $\hat{a} = \bar{x} - \sqrt{3}s$

$$f(x) = \frac{1}{b-a}$$

$$E(x) = \int_a^b x \cdot f(x) \cdot dx$$

$$= \int_a^b \frac{x}{b-a} dx$$

$$= \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b$$

$$= \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2} \quad - (1)$$

$\rightarrow$ $$E(X^2) = \int_a^b x^2 \frac{1}{b-a} dx$$

$$= \left[\frac{x^3}{3}\right]_a^b \frac{1}{b-a}$$

$$= \frac{b^3-a^3}{3(b-a)}$$

$$= \frac{(b-a)(b^2+ab+a^2)}{3(b-a)}$$

$$= \frac{a^2+b^2+ab}{3} \rightarrow (2)$$

(ii) $\rightarrow$ $\frac{a+b}{2} = \bar{x}$ From (1)

$\Rightarrow$ $a+b = 2\bar{x}$

$\Rightarrow$ $(2\bar{x}-b)^2 + b^2 + (2\bar{x}-b)b = 3\bar{x}_2$ (From (2))

$\rightarrow b^2 + 4\bar{x}^2 - 4b\bar{x} + 4\bar{x}^2 - 3\bar{x}_2 = 0$

$$\Rightarrow \quad \hat{b} = -2\bar{x} \pm \sqrt{2(\bar{x})^2 - 4(4\bar{x}^2 - 3\bar{x}_2)}$$

$$\therefore \hat{b} = \bar{x} + \frac{1}{2}\sqrt{4x^2 - 4.4\bar{x}^2 + 4\cdot 3\bar{x}_2{}^2}$$

$$\therefore \hat{b} = \bar{x} + \sqrt{3(\bar{x}_2 - \bar{x}^2)}$$

(iii) $n = 1, 2, 3, 4, 5 \rightarrow U(1,5)$

$$\bar{x} = \frac{1+5}{2} = 3$$

$$S^2 = \frac{(5-1)^2}{12} = \frac{16}{12} = 1.333.$$

(iv) $f(x) = \dfrac{1}{b-a}$

$$\Rightarrow \quad L(b) = \prod_{i=1}^{n} \frac{1}{b-a}$$

$$= \cancel{\frac{1}{b^n - a^n}}$$

$$= \frac{1}{(b-a)^n}$$

$$= (b-a)^{-n}$$

$$\Rightarrow \quad \log L(b) = -n \log(b-a)$$

$$\Rightarrow \quad \frac{d}{db} \log L(b) = -\frac{n}{b-a}(-a)$$

$$= \boxed{\frac{na}{b-a}}$$

6 (i) $H_0 : p = 0.1$
$H_1 : p > 0.1$

$p(\text{Type I error}) = p(\text{rejecting } H_0 / H_0 \text{ is true})$

$$P(\text{Type I error}) = P(X \geq 3 / p = 0.1) \times P(X = 0.1, p = 0.1)$$
$$P(Y \geq 5 / p = 0.1) + P(X = 1 \mid p = 0.1)$$
$$P(Y \geq 4 \mid p = 0.1) + P(X = 2 \mid p = 0.1)$$
$$+ (Y \geq 3 \mid p = 0.1)$$

$$= 0.14727$$

~~(ii) Probability of type II error~~

(ii) Probability of ~~1 - 0.9~~
rejecting null Hypothesis H_0 when $p = 0.3$

$$= P(X \geq 3 \mid p = 0.3) + P(X = 0 \mid p = 0.3)\, P(X > 5 \mid p = 0.3)$$
$$+ p(X = 1 \mid p = 0.3) \cdot p(X \geq 4 \mid p = 0.3) + P(X = 2 \mid p = 0.3)$$
$$p(Y \geq 3 \mid p = 0.3)$$

$$= 0.9125$$

(iii) Probability of type II error $= 1 - 0.91253$
$= 0.08747$
$= ~~9~~ \approx 0.09$

(iv) $$\frac{X/n - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

$$\hat{p} - 12.816 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.27822$$

(V) $p > 0.278$ at 10% LOS

$H_0 : p = 0.278$ against $H_1 : p > 0.278$

$$\frac{X - nP_0}{\sqrt{nP_0 \%}} \sim N(0,1)$$

$$\frac{39.5 - 120(0.278)}{\sqrt{120 \times 0.278 \times 0.722}} = 1.2311$$

(vii) Test statistic $\frac{(\bar{X}_1 - \bar{X}_2)(\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1 + n_2 - 2}$

$$\Rightarrow \frac{85 - 7}{65.451 \sqrt{\frac{1}{12} + \frac{1}{15}}} = -0.2034$$

7.

(i) Sample Mean, $X_1 = 30$

$X_2 = 35.0556$

Variance, $\sigma_1^2 = 57.1667$

$\sigma_2^2 = 59.91358$

(iii) 95% of CI = $\mu \pm z_{\alpha/2} \sqrt{\sigma^2/n}$

$= 35.0556 \pm 1.96 \sqrt{\frac{59.91358}{9}}$

$= (29.9985, 40.11265)$

8. (i) If we have a random sample from a distribution with an unknown parameter θ and $g(\underline{X})$ is an estimater of θ, It seems desirable that $E(g(\underline{X})) = \theta$

(ii) When the expected value and true value is not equal is referred to as a bias.

(iii) The $MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$

(iv) An estimator with lower value of MSE is more efficient. Thus, θ is better than $\hat{\theta}$.

(V) They will both be considered consistent when

(vi) We can estimate θ by MOM & MLE

9.(i) ~~The sampling distribution will be~~
The t_k will be defined as

$$t_k = \frac{N(0,1)}{\sqrt{\frac{\chi^2_k}{k}}} \quad \text{where } N(0,1) \text{ \& } \chi^2_k \text{ are iid.}$$

(ii) Mean = 0
var = $\frac{k}{k-2}$

(iii) $n = 10$
$\bar{x} = 50$
$\sigma^2 = 48.667$

$$\bar{X} \sim N(\mu, \sigma^2) \Rightarrow \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

→ 99% of CI $= \bar{x} \pm t_{\alpha/2, n-1} \sqrt{\frac{\sigma^2}{n}}$

$= 50 \pm 3.250 \sqrt{\frac{48.667}{10}}$

$= (142.83, 57.169)$

11. $n = 12$

$\Sigma x = 213,200$

$\Sigma x^2 = 4,919,860,000$

$\bar{x} = 17,767$

$S = 10,144$

Temp employees' salary - 11,000 & 48,000

Permanent employees' salary - 27,500 & 31,500

→ ~~Revise~~

Revised $\Sigma x_i = 213200 - 11000 - 48000$
$+ 27500 + 31500$
$= 213200$

→ Revised mean, $= \frac{213200}{12} = 17767.$

→ Revised $s^2 = \frac{1}{11}\left[4919860000 - 12(17767)^2\right]$

$= 447240617.8$

$\Rightarrow S = 21148.06466.$

12. (i) The cramer-rao lower bounds estimators does not apply here because of the range being discontinuous.

(ii) P.D.F of Y

$P[Y < y] = P[\max X_i < y]$

$$= P\left[\prod_{i=1}^{n} (x_i < y)\right]$$

$$= \prod_{i=1}^{n} P(x_i < y)$$

$$= \prod_{i=1}^{n} \left(\int_0^y dy/\theta \right]$$

$$= [y/\theta]^n$$

$$\Rightarrow \quad g_Y(y) = \frac{d}{dy} P(x < y)$$

$$= \boxed{\frac{n}{\theta^n} y^{n-1}}$$

$$\rightarrow \quad E[Y^k] = \int_0^\theta y^k \frac{n}{\theta^n} y^{n-1} \, dy$$

$$= \frac{n}{n+k} \int_0^\theta \frac{n+k}{n} y^{n+k-n} \, dy$$

$$= \frac{n}{n+k} \frac{\theta^{n+k} - 0}{\theta^n}$$

$$= \frac{n\theta^k}{n+k}$$

(iii) ~~Biased estimator~~

$$\text{Bias}\,[\hat{\theta}(c)] = E[\hat{\theta}(c) - \theta]$$

$$= E[c^y - \theta]$$

$$= c\,E[Y] - \theta$$

$$= \frac{cn\theta}{n+1} - \theta$$

$$= \left(\frac{cn - 1}{n+1}\right)\theta$$

Date

$\rightarrow \quad MSE[\hat{\theta}(c)] = E[(\hat{\theta}(c) - \theta)^2]$

$= E[(cy - \theta y^2)]$

$= E[c^2y^2 - 2c\theta y + \theta^2]$

$= c^2 \; E[y^2] - 2c\theta E[y] + \theta^2$

$= c^2 \dfrac{n\theta^2}{n+2} - c\,\dfrac{2n\theta^2}{n+1} + \theta$

(iv) For unbiasedness, the bias should result to 0

$Bias(\hat{\theta}\, c(n)) = 0$

$\left(\dfrac{cn}{n+1} - 1\right)\theta = 0$

$\rightarrow c_n = \dfrac{n+1}{n}$

(v) $\quad \dfrac{d}{dc} \; MSE(\hat{\theta}(c)) = 0$

$\Rightarrow \quad 0 = \dfrac{d}{dc}\left[c^2 \dfrac{n\theta^2}{n+2} - \dfrac{c^2 n\theta^2}{n+1} + \theta^2\right]$

$= \dfrac{2cn\theta^2}{n+2} - \dfrac{2n\theta^2}{n+1}$

$\Rightarrow \quad C_m = \dfrac{2n\theta^2}{n+1} \cdot \dfrac{n+2}{2n\theta^2} = \dfrac{n+2}{n+1}$

vi) The estimator with lower MSE should be chosen for estimating θ.