INSTITUTE OF ACTUARIAL & QUANTITATIVE STUDIES

**Subject:** Probability and Statistics 1

**Chapter:** Unit 1 & 2

**Category:** Assignment Solutions

1.

**i)** $k^{th}$ moment = $E[X^k]$

**ii)** $k^{th}$ moment about $\alpha$. $= E[(X-\alpha)^k]$

**iii)** $k^{th}$ central moment $= E[(X-\mu)^k]$ where $\mu$ is the mean

**iv)** Coefficient of skewness $= \frac{E[(X-\mu)^3]}{\sigma^3}$ where $\sigma$ is the std. deviation

2.

**i)** $\text{P[Withdrawal]} = \text{P[Withdrawal | Agency]. P[Agency]} +$
$\text{P[Withdrawal | Bank]. P[Bank]} + \text{P[Surrender | Online]. P[Online]}$
$= 0.05\left(\frac{2000}{10000}\right) + 0.08\left(\frac{3500}{10000}\right) + 0.14\left(\frac{4500}{10000}\right) = 0.101$

**ii)** $\text{P[Withdrawal} \cap \text{Agency]} = \text{P[Withdrawal | Agency]. P[Agency]} = 0.05 * 0.20 = 0.01$
$\text{P[Withdrawal} \cap \text{Bank]} = \text{P[Withdrawal | Bank]. P[Bank]} = 0.08 * 0.35 = 0.028$
$\text{P[Withdrawal} \cap \text{Not online]} = 0.01 + 0.028 = 0.038$
P[Withdrawal | Not Online] = P[Withdrawal ∩ Not Online] / P[Not Online]
$= \frac{0.038}{1-0.45} = 0.0691$

**iii)** $P[Bank \mid \text{Withdrawal}] = \frac{P[\text{Withdrawal} \cap Bank]}{P[\text{Withdrawal}]} = \frac{0.280}{0.101} = 0.2772$

3.

**i)** Ordering the marks given, stem and leaf diagram is:

| | |
|---|---|
| 1 | 7, 9 |
| 2 | 1, 1, 2, 2, 6, 6, 7, 7, 7, 7 |
| 3 | 0, 0, 1, 2, 3, 3, 3, 4, 6 |
| 4 | 6 |
| 5 | 3 |
| 6 | 0 |

The stems are 10s and leaves are units.

**ii)** Median: $\left(\frac{1}{2}n + \frac{1}{2}\right)^{th}$ value =12.5$^{th}$ value =(27+30)/2 =28.5.

Mode: 27. (27 appears the maximum number of times-four times)



**iii)** Interquartile Range (IQR) = $Q_3 - Q_1$

Now $Q_1$ = $\frac{n+2}{4}$ th value counting from below = 6.5$^{th}$ value

= (22+26)/2 = 24

$Q_1$ = $\frac{n+2}{4}$ th value counting from above =33

Hence, $IQR = Q_3 - Q_1$ =33-24 =9

*[Alternatively,*

$Q_1$ = $\frac{n+1}{4}$ *th value counting from below = 6.25th value =23 and*

$Q_3$= $\frac{n+1}{4}$ *th value counting from above= 33*

*Hence IQR = 33-23=10]*

4.

The binomial distribution $(n, p)$ has probability function.

$$P(X = x) = \frac{n!}{(n-x)!\,x!}\, p^x (1-p)^{(n-x)} \ ; \quad x=0,1,2,\ldots;\ 0<p<1.$$

and $P(X = x - 1) = \frac{n!}{(n-x+1)!(x-1)!}\, p^{(x-1)} (1-p)^{(n-x+1)}$ ; $x=1,2,3,\ldots$ ;$0<p<1$

Now, $\frac{P(X=x)}{P(X=x-1)} = \frac{n-x+1}{x} \frac{p}{1-p}$

Therefore, $P(X = x) = \frac{n-x+1}{x} \frac{p}{1-p} P(X = x - 1)$ ; $\quad x=1,2,3,\ldots$

5.

**i)** Let Success: = Getting one passenger to go to *B* from *A*.

$p$ = Probability [Getting one passenger to go to *B* from *A*.] = 0.3;

$q = 1 - p = 0.7$

$k$ = the number of successes (number of passengers going to Town *B*) = 4

$X+k$ = the number of trials=15

$X$ = the number of failures = 11

$X$ follows *NB* ( $k$=4, $p$ = 0.3)

$$[X \sim NB(k,p) \rightarrow P(X = x) = \binom{x+k-1}{x} p^k q^x; = x = 0,1,\ldots \quad ]$$

Hence, $P(X = 11) = \binom{11+4-1}{11} 0.3^4 0.7^{11}$

$$= \frac{14!}{11!3!} 0.3^4 0.7^{11} = 0.0583.$$

**ii)** The average number of persons to be asked in order to get 4 passengers

$= E$ (the number of trials)

$= E(X + k) = E(X) + k = \frac{kq}{p} + k = \frac{k}{p}$

$= \frac{4\times0.7}{0.3} + 4 = \frac{4}{0.3}$

$= 13.33 \sim 14$ persons.

6.

**i)** Mode: For fixed $\vartheta > 0$, the density function $f(x)$ is an increasing function of $x$.

Thus, $f(x)$ has maximum at the right end point of the interval $[0, \vartheta]$.

Hence the mode of this distribution is $\vartheta$.

Median:

$$\tfrac{1}{2} = \int_0^q f(x)\,dx$$

$$= \int_0^q \frac{3x^2}{\theta^3}\,dx = [x^3/\theta^3] \text{ from } 0 \text{ to } q$$

Thus, $\tfrac{1}{2} = q^3/\theta^3$ implies $q = \dfrac{\theta}{2^{\frac{1}{3}}}$

**ii)** Let $A$ be the ratio of the mode of this distribution to the median

$A$ = mode/median = $\theta \times \dfrac{2^{\frac{1}{3}}}{\theta} = 2^{\frac{1}{3}} = 1.2599$

$$P(X < A) = \int_0^A f(x)\,dx$$

$$= \int_0^A 3x^2/\theta^3\,dx$$

$$= [x^3/\theta^3] \text{ from } 0 \text{ to } A$$

$$= A^3/\theta^3$$

$$= \begin{cases} \dfrac{2}{\theta^3} & if \quad \theta > 2^{1/3} = 1{\cdot}2599 \\ 1 & otherwise \end{cases}$$

7.

**i)** Median = $\left(\frac{1}{2}n + \frac{1}{2}\right)^{th}$ *value*

$Q_1 = \left(\frac{1}{4}n + \frac{1}{2}\right)^{th}$ value and $Q_3 = \left(\frac{3}{4}n + \frac{1}{2}\right)^{th}$ value.

For city *A*:

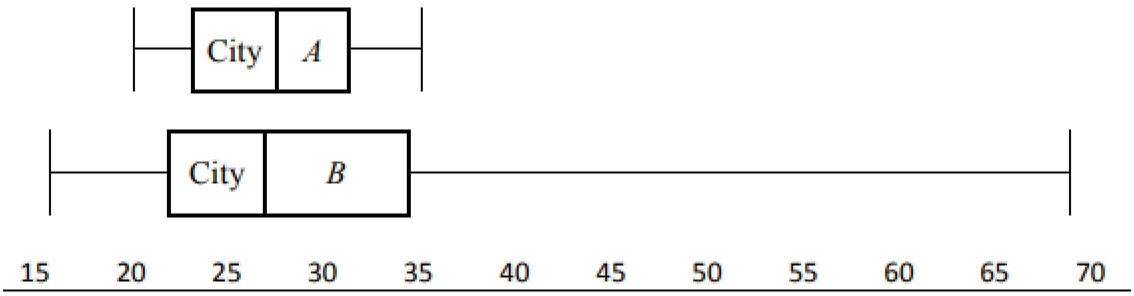
Median = 27.50

$Q_1 = 23.50$ ; $Q_3 = 31.50$

For city *B*:

Median = 27.00

$Q_1 = 22.50$ ; $Q_3 = 34.50$




ii) Looking at the box plots, we see that the median of both distributions are close to 27 ° Celsius. This suggests that the monthly maximum temperatures for City A and B may have the averages close to each other.

However, the overall spread of the figures for city B appears to be greater than the corresponding spread for city A which can be confirmed by measuring by IQR. This suggests that the variability in the monthly maximum temperatures for City B is greater than the corresponding variability for city A (although conclusions drawn from such small sample sizes should be treated with caution).

The value of 68 for city B could be an outlier.

City A distribution seems symmetric and City B distribution is clearly positively skewed.

For comparison:
- for city A: the modes (26.00) ≈ median (27.50) = mean (27.50)
- for city B: the mode (25.00) < median (27.00) < mean (30.75)

8.

The $\chi_9^2$ distribution has mean 9 and variance 18.

Setting the expressions for the mean and variance of the lognormal distribution to the above values and solving the two equations:

$$e^{\mu + \frac{1}{2}\sigma^2} = 9\,; \qquad e^{2\mu + \sigma^2}\left(e^{\sigma^2} - 1\right) = 18$$

Squaring the first equation and substituting into the second, we get:

$$81\left(e^{\sigma^2} - 1\right) = 18$$

Solving this, we get $\sigma^2 = \log\frac{99}{81} = 0.2007$

Now substituting $\sigma^2$in any one equation, we get $\mu = 2.0969$.

If $X \sim$ log Normal then log $X \sim$ Normal, so: $P\ (X > 9) = P\ (\log X > \log 9)$

$$= \mathrm{P}\ \left(Z > \frac{\log 9 - 2.0969}{\sqrt{0.2007}}\right) = \mathrm{P}\ (Z > \ 0.224)$$

Using interpolation from tabulated values for 0.22 and 0.23 in page 160

$= 1 - P\ (Z < 0.224) = 1 - 0.5886 = 0.4114$ **[5 Marks]**

9.

Let $X$ be the time for the first exotic cake and $Y$ be the time for the second. Then:

$X \sim N(120, 225)$ and $Y \sim N(120, 225)$ [in terms of minutes]

We require: $P(|X - Y| < 25)$

The distribution of $X$ - $Y$, is:

$$(X - Y) \sim N(120 - 120, 225 + 225) \sim N(0, 450)$$

$$P(|X - Y| < 25) = P(-25 < X - Y < 25)$$

$$= P\left(\frac{-25}{\sqrt{450}} < Z < \frac{25}{\sqrt{450}}\right)$$

$$= P(Z < 1.1785) - P(Z < -1.1785)$$

$$= P(Z < 1.1785) - (1 - P(Z < 1.1785))$$

$$= 2P(Z < 1.1785) - 1$$

$$= (2)(0.8807) - 1 = 0.7614$$

10.

**(i)** Let S be the salary (constant) of each of the 99 employees.

Hence, average salary of 99 employees = (S+S+...+S)/99 = S
The variance of the salary of 99 employees was 0 (as they are getting the constant salary)

**[1]**

**(ii)** S+1000 is the salary of the 100th employee.

So, the average salary of 100 employees = [(S+S+...+S) +(S+1000)] /100
= (100S+1000)/100 = S+10

With the addition of the 100th employee, the average salary of 100 employees has increased by Rs 10.

Now, the variance of the salary of 100 employees

$$\frac{1}{100}[\{(-10)^2+(-10)^2+\ldots+(-10)^2\}+990^2]$$

$$= \frac{1}{100}[9900+980100] = \frac{1}{100}(990000) = 9900$$

The standard deviation = $\sqrt{9900}$ = Rs 99.50, which is positive

11.

**(i)** Let $T$: Accident with tyre burst

$C$: Accident due to collision with the road divider and

$D$: Death casualty in a car accident

Given that: $P(T) = 0.6$, $P(C) = 0.4$, $P(D|T) = 0.3$ and $P(D|C) = 0.5$.

Given accidental death casualty, the probability of tyre burst is:

$$P(T|D) = \frac{P(T)P(D|T)}{P(T)\,P(D|T) + P(C)\,P(D|C)}$$

$$= \frac{0.6\,(0.3)}{(0.6\,(0.3) + 0.4\,(0.5))} = \frac{18}{38} = 0.4737.$$

**(ii)** To find the most probable cause of death casualty due to accidents, we need to compare given accidental death casualty, the probability of tyre burst and probability of collision with the road divisor; i.e $P(T|D)$ and $P(C|D)$

Now, $$P(C|D) = \frac{P(C)P(D|C)}{P(T)\,P(D|T) + P(C)\,P(D|C)}$$

$$= \frac{0.4\,(0.5)}{(0.6\,(0.3) + 0.4\,(0.5))} = \frac{20}{38} = 0.5263.$$

[Alternately, $P(C/D) = 1 - P(T/D) = \frac{20}{38}$]

Since, $P(T|D) < P(C|D)$, collision with the road divider is the most probable cause of accidental death casualty.

12.

**(i)** P(obtaining correct password in the third try)

= P(obtaining incorrect password in the first two attempts and obtaining correct password in the third attempt)

= (1- 1/100) (1- 1/99) (1/98)

=1/100. **[2]**

**(ii)** File can be accessed if the password is correct in the first attempt or second attempt or third attempt.

That is: P (Password is correct in the first attempt) + P(password is incorrect in the first attempt and correct in the second attempt) + P(password is incorrect in the first two attempts and correct in the third attempt)

= 1/100 + (1- 1/100)(1/99) + (1 - 1/100) (1 - 1/99) (1/98)

=3/100. **[2]**



**(iii)** P (Correct password is found on the 10th try)

= [P (Incorrect password in the first 9 attempts)][P(correct password on the 10th attempt)]

[1]

= $[(1 - 1/100)^9][1/100]$ =0.009135. **[2]**

13.

Let $X_i$ (i = 1,2,3) denote the number of hospitalisations in the month of October, November and December respectively.

From the information provided, $X_1 \sim Poi\ (2)$, $X_2 \sim Poi\ (3)$ and $X_3 \sim Poi\ (1)$

Let the total hospitalisation over this period be denoted by X where $X = X_1 + X_2 + X_3$

Since all $X_i$'s are independent

$X \sim Poi\ (2 + 3 + 1)$

Thus, $P [X < 5] = \sum (6^x e^{-6})/x!$ (summation over x = 0 to 4)

= 0.0248 + 0.0149 + 0.0446 + 0.0892+ 0.1339

= 0.2851

14.

**i)** **a)** Let $X_i$ represent each motor claim amount for i = 1 to 10

Moment generating function for exponential distribution, $M_x(t) = (1 - t/\lambda)^{-1}$

Hence, for $Y = \sum X_i$ (i = 1 to 10)

$M_Y(t) = (M_X(t))$ ^ 10

$= (1 - t/\lambda)^{-10}$

which is the moment generating function of gamma distribution with $\alpha = 10$ and $\lambda = 1.25$

**b)** MGF of 2.5Y is $E[e^{(2.5t)Y}]$

$= M_Y[2.5t]$

$= (1 - 2t)^{-10}$

$= (1 - t/0.5)^{-10}$

which is the moment generating function of gamma (10, 0.5)

i.e. $\chi^2_{20}$ distribution

ii) B

15. Since only one claim is eligible for each of the ailments, claims from Heart, Cancer and Liver related ailments can be modelled as three Bernoulli Variables (indicator variables). It is given in the question that the three can be assumed to be independent.
H = Claims from Heart related ailments H ~ Bernoulli (0.01)
C = Claims from Cancer related ailments C ~ Bernoulli (0.02)
L = Claims from Liver related ailments L ~ Bernoulli (0.005)

Let X be the claim amount to be paid out in the next year on a single policy $X = 20 \times H + 25 \times C + 15 \times L$ We have to find E(X) and s.d.(X)

$E(X) = 20 \times E(H) + 25 \times E(C) + 15 \times E(L)$
$= 20 \times (0.01) + 25 \times (0.02) + 15 \times (0.005)$ ....... *$E(A) = p$ for $A \sim$ Bernoulli(p)*
$= 0.775$ lakhs
$=$ INR 77,500

$Var(X) = 20^2 \times Var(H) + 25^2 \times Var(C) + 15^2 \times Var(L)$
...... *Since H, C and L are independent, no co-variance terms*

$= 400 \times (0.01)(1\text{-}0.01) + 625 \times (0.02)(1\text{-}0.02) + 225 \times (0.005)(1\text{-}0.005)$
............ *$Var(A) = p(1\text{-}p)$ for $A \sim$ Bernouli(p)*
$= 17.32938$ lakhs

$SD(X) = (17.32938)^{1/2}$
$=$ INR 4.1628 lakhs

**ii)**
Exactly one claim has occurred. We don't know whether it is related to H, C or L.

P(exactly 1 claim)
$= P(H) \times (1\text{-}P(C)) \times (1\text{-}P(L)) + (1\text{-}P(H)) \times P(C) \times (1\text{-}P(L)) + (1\text{-}P(H)) \times (1\text{-}P(C)) \times P(L)$
$= (0.01)(1\text{-}0.02)(1\text{-}0.005) + (1\text{-}0.01)(0.02)(1\text{-}0.005) + (1\text{-}0.01)(1\text{-}0.02)(0.005)$
$= 0.009751 + 0.019701 + 0.004851$
$= 0.034303$

$P(H \mid 1 \text{ claim has occurred}) = 0.009751 / 0.034303 = 0.284261$
$P(C \mid 1 \text{ claim has occurred}) = 0.019701 / 0.034303 = 0.574323$
$P(L \mid 1 \text{ claim has occurred}) = 0.004851 / 0.034303 = 0.141416$

These should total up to 1.

So, we have to find $E(X \mid 1 \text{ claim has occurred})$

$E(X \mid 1 \text{ claim has occurred})$
$= 20 \times P(H \mid 1 \text{ claim has occurred}) + 25 \times P(C \mid 1 \text{ claim has occurred}) + 15 \times P(L \mid 1 \text{ claim has occurred})$
$= 20 \times 0.284261 + 25 \times 0.574323 + 15 \times 0.141416$
$=$ INR 22.16453 lakhs

............ *Kindly note that even after taking account the condition that one claim has occurred, H,C and L continue to be Bernoulli variables and hence their mean will be equal to p ...... although the value of p has changed now.*

iii) There are three independent risks covered under this policy with relatively very small probability of incidence of a claim in the next year. The probability of no claim during the next one year = (1-0.01) (1-0.02) (1-0.005) = 0.965349. Since in almost 96% of the cases, there will be no claim, the expected pay-out at the inception of the policy is quite low (lower than1 lakh). However, after one claim has occurred, we have actually experienced something which has a possibility of 3.4% to occur. After its occurrence we are finding out the expected amount since we don't know whether it relates to H, C or L (otherwise there was no need of expectation, we could directly infer it to be 20 lakhs, 25 lakhs or 15 lakhs). Since something which was only 3.4% probable has actually occurred, there is a significant increase in the expected claim pay-out from (i) to (ii).

INSTITUTE OF ACTUARIAL & QUANTITATIVE STUDIES

P&S 1 – UNIT 1 & 2

ASSIGNMENT SOLUTIONS