MTL 390 (Statistical Methods)
Minor Examination Assignment 1 Report

Name: Bhumika Chopra Entry Number: 2018MT10748

1. Descriptive Statistics

(a) The following table represents the final marks (out of 100) scored by 60 university students in the course- 'The Science of Well Being'.

60	77	65	55	45	56	39	92	70	40
28	25	34	33	34	34	34	29	24	38
42	22	27	78	42	26	75	86	28	50
59	32	87	43	95	49	96	55	25	90
71	65	29	33	22	64	35	39	77	71
88	87	99	56	23	84	92	42	83	59

(a) Construct the frequency table using appropriate intervals and then plot the frequency histogram of the data.

(b) Give a descriptive statistics report with relevant information.

(c) The professor is a nice person and does not want any of her students to earn a bad grade, so she decides to give a bonus of 3 marks to the students having the lowest 5% of the test scores. Find the highest score that a student can obtain and still receive the bonus points?

Solution

(a) The frequency table is -

Interval	Frequency
1-10	0
11-20	0
21-30	12
31-40	12
41-50	7
51-60	7
61-70	4
71-80	6
81-90	7
91-100	5

To plot the frequency histogram we need continuous intervals. We will modify the above table.

Interval	Frequency
0.5-10.5	0
10.5-20.5	0
20.5-30.5	12
30.5-40.5	12
40.5-50.5	7
50.5-60.5	7
60.5-70.5	4
70.5-80.5	6
80.5-90.5	7
90.5-100.5	5

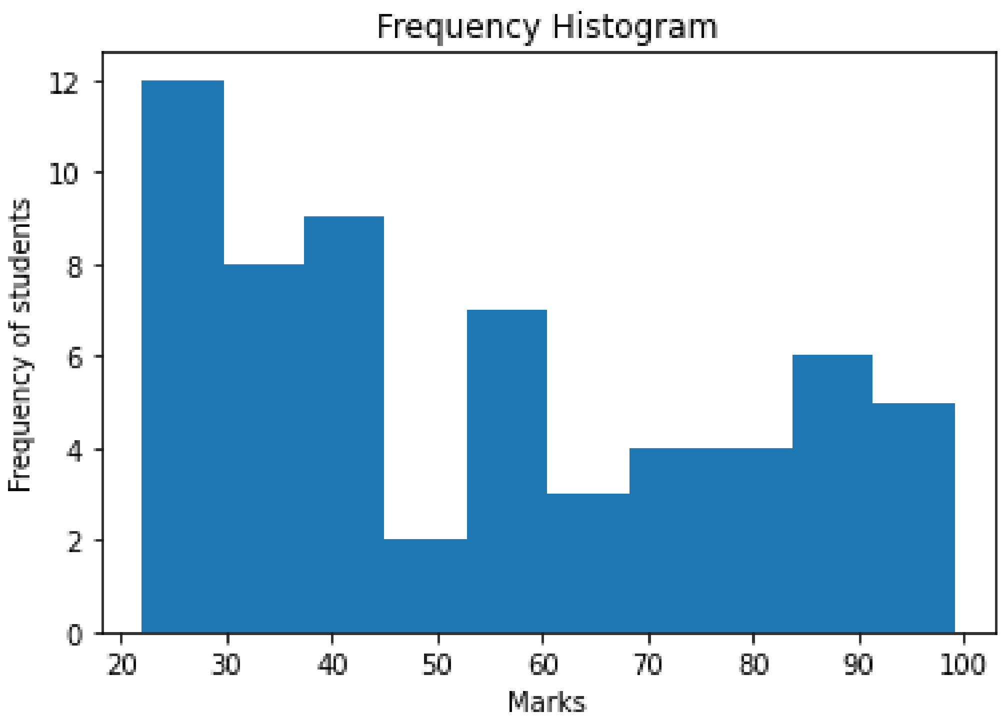

Figure 1: Frequency Histogram

(b) The important statistics of the data have been summarized in the below table.

Statistic	Value
Maximum	99
Minimum	22
Range	77
Mean	53.97
Mode	34
Median	49.5
Variance	571.355
Standard Deviation	23.903
Coefficient of Variation	0.443
Skewness	0.359
Kurtosis	-1.240
Excess Kurtosis	-4.240

(c) The professor has decided to give bonus marks to all the students having the lowest 5% of test scores. After sorting the test scores in ascending order we find that the highest marks a student can obtain and still receive the bonus is 29 points. After the bonus the new frequency histogram of the distribution of marks is given below.

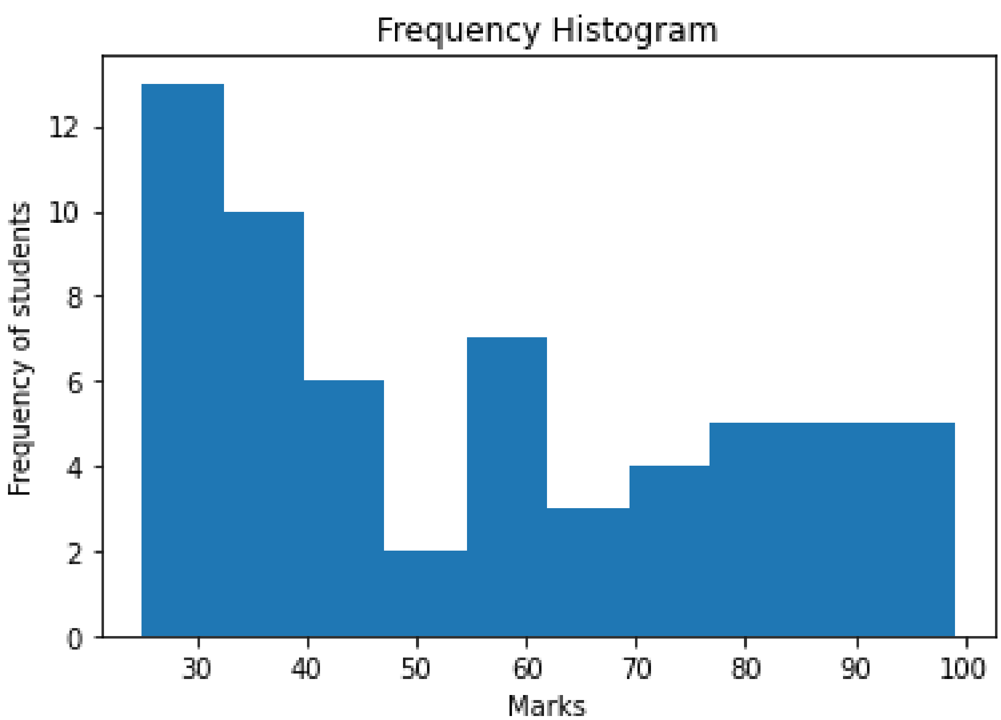

Figure 2: Frequency Histogram

(b) Let X be a random variable such that $X \sim Gamma(\alpha, k)$ and let α is known. Define

$Y = \gamma log(X)$. Show that if $\gamma > 0$ is know, then the distribution of Y is an exponential distribution.

Solution

The density function of X is given by

$$f(x) = \frac{1}{\Gamma(\alpha)k^{\alpha}} x^{\alpha-1} e^{-x/k}(x)$$

Logarithm is a is a continuous Borel measurable function, therefore the density function of Y is,

$$t(y) = \frac{1}{\Gamma(\alpha)\gamma} e^{\alpha(y-\gamma log(k))/\gamma} exp[-e^{(y-\gamma log(k))/\gamma}]$$

The above distribution belongs to a location-scale family with location parameter $\eta = \gamma log(k)$ and scale parameter γ.

When γ is know, we can write the density function as,

$$t(y) = \frac{1}{\Gamma(\alpha)\gamma} e^{\alpha y/\gamma} exp[-\frac{e^y \gamma}{k} - \alpha log(k)]$$

This belongs to the exponential family.

2. Descriptive Statistics

(a) The following 3 tables represents the temperature (measured in °C) in Mumbai, Delhi, and Bangalore in the month of April.

Mumbai

29	30	28	31	31	30	33	33	34	29
30	31	32	31	35	36	32	34	30	32
33	31	30	33	33	33	32	35	32	33

Delhi

26	28	35	22	29	25	30	32	31	30
29	30	30	32	33	34	27	29	30	31
34	33	30	33	29	30	32	31	31	30

Bangalore

27	28	29	27	29	25	28	27	28	29
29	30	28	29	30	27	30	29	28	30
29	31	32	29	28	29	30	29	30	29

(a) Give a descriptive statistics report of the temperature in the 3 cities.

(b) Make box plots for the 3 cities.

Solution

(a) The descriptive statistics report for the temperature in the 3 cities in the month of April is -

Statistics	Mumbai	Delhi	Bangalore
Maximum	36	35	32
Minimum	28	22	25
Range	8	12	7
Mean	31.88	30.2	28.77
Mode	33	30	29
Median	32	30	29
Variance	3.706	7.614	1.909
Standard Deviation	1.925	2.759	1.382
Coefficient of Variation	0.0594	0.0594	0.0594
Skewness	0.104	0.1035	0.1035
Kurtosis	-0.507	-0.5069	-0.507
Q1 Quartile	30.25	29	28
Q3 Quartile	33	32	29.75
Interquartile range	2.75	3	1.75

From the table we observe that-

- Mumbai has the highest temperatures, therefore people who are more accustomed to high temperatures should tend to live in Mumbai.
- Delhi offers a mix of low and high temperatures with the widest range and largest standard deviation. This shows that the temperature tends to fluctuate a lot in Delhi.
- Bangalore offers the most consistent temperature with lowest mean temperature. It has the least range and smallest standard deviation. People who prefer consistent and cool temperature conditions tend to live in Bangalore.

(b) The box plots are -

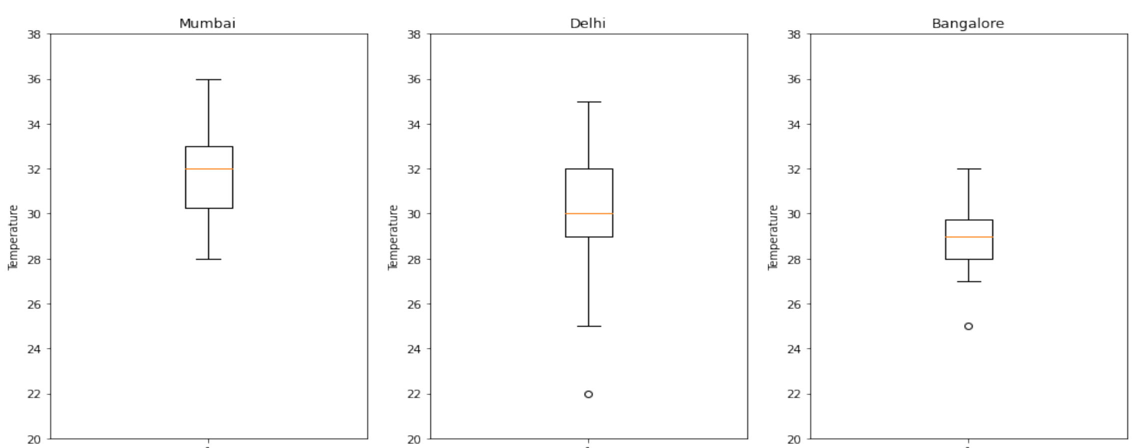

Figure 3: Box Plot

We are able to clearly visualize our findings using the box plots.

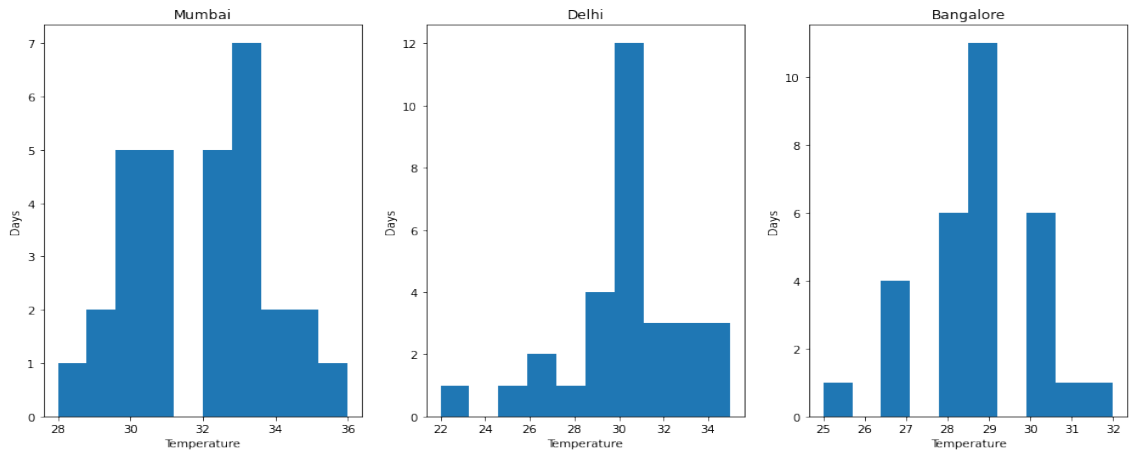

Figure 4: Frequency Histogram

(b) Let $X = (X_1, X_2, ...X_n)$ be a random sample from $N(0,1)$. Show that $A_i = \frac{X_i^2}{\sum_{j=1}^{n} X_j^2}$ and $B = \sum_{k=1}^{n} X_k^2$ are independent $\forall i = 1, 2, 3...n$.

Solution We know that the square of a normal distribution follows chi^2 distribution. Since all X_i are independent implies all X_i^2 are independent $\forall i = 1, 2, ...n$. Let $Y = (X_1^2, ..., X_n^2)$. The joint density distribution is,

$$f(\mathbf{y}) = \frac{ce^{-(y_1+y_2+....+y_n)/2}}{\sqrt{y_1 y_2 ... y_n}}$$

We observe that $A_i B = X_i^2$ and $\sum_{i=1}^{n} A_i = 1$.
The density function of $Z = (B, A_1, A_2, ..., A_n)$ is

$$f(\mathbf{z}) = \frac{ce^{-b/2} a_n b^{n-1}}{\sqrt{u^n v_1 ... v_n}} = cb^{n/2-1} e^{-b/2} \sqrt{\frac{1 - a_1 a_2 ... a_n}{a_1 a_2 ... a_n}}\, b > 0, a_j > 0$$

Therefore, we get the random variables B and $A_1/B, ..., A_{n-1}/B$ are independent. Since $A_{n-1} = 1 - (A_1 + A_2 + ... + A_n)$, we can conclude that B and A_n/B are independent.

3. Sampling Distributions

(a) Akira conducted a survey to analyze the study time and play time balance for her university students. She found out that on average after every 2 hours a student spent studying rigorously, they would take a short break for some time T, which follows an exponential distribution with a mean of 20 minutes.

Consider a random sample of 50 students, let X_i denote the time spent playing by the i^{th} student after having studied for 2 hours.

(a) Find the probability that the average time a student spent playing was more than 15 minutes.

(b) Now, consider that only 1 student was randomly picked and find the probability that the time they spent playing is more than 15 minutes.

(c) Why are the probabilities in the above 2 parts different?

Solution

(a) Let $X = (X_1, X_2, ...X_{50})$ be the random sample of 50 students that was drawn. Given that $X_1 \sim Exp(1/20)$. We have to find $P(\bar{X} > 15)$ where $\bar{X}$ is the sample mean. S
Since the sample size is large (¿30) we can use the Central Limit Theorem to get a distribution of $\bar{X}$.

$$\bar{X} = \frac{1}{50}\sum_{i=1}^{50} X_i$$

$$E(\bar{X}) = E(X_i) = 20$$

$$Var(\bar{X}) = \frac{Var(X_i)}{50} = 400/50 = 8$$

Let,

$$Z = (\frac{\bar{X} - 20}{\sqrt{8}}) \sim N(0, 1)$$

By using the Central Limit Theorem, we get

$$P(Z > \frac{15 - 20}{2.828}) = P(Z > -1.768)$$

Using z-table, we get,

$$P(Z > -1.768) = 0.430$$

Therefore, the probability that the average time a student spent playing was more than 15 minutes is 0.430.

(b) Let X is the random variable corresponding to the picked student. $X \sim Exp(1/20)$, we have

$$P(X > 15) = 1 - P(X \leq 15) = 1 - (1 - e^{-\frac{15}{20}}) = e^{-\frac{3}{4}} = 0.4723$$

(c) In part (a) we have drawn a random sample of reasonable size (so as to apply CLT) while in the (b) we have just taken 1 student. Naturally, the answer in (a) is a more accurate.

(b) Let X_1, X_2, X_3 are 3 independent normally distributed random variables. Given that $X_1 \sim N(0, 9)$, $X_2 \sim N(0, 16)$, and $X_5 \sim N(0, 25)$. The find $P(\frac{X_1+X_2}{|X_3|} \leq 3)$

Solution

We observe that $X_1 + X_2 \sim N(0, 25)$ and after normalization $S = \frac{X_1+X_2}{5} \sim N(0, 1)$.
Also, $\frac{X_3}{5} \sim N(0, 1)$ and $Q = ((\frac{X_3}{5})^2) \sim \chi^2(1)$.
We know, $T = \frac{S}{\sqrt{Q/1}} \sim t(1)$ follows the Students t-distribution with 1 degree of freedom.
Therefore,

$$P(\frac{X_1 + X_2}{|X_3|} \leq 3) = P(\frac{5S}{5\sqrt{Q}} \leq 3)$$
$$= P(T \leq 3) = 0.8976$$

The value is calculated using students t-distribution table.

4. Sampling Distributions

   (a) The marks of the final year end semester examination were posted and it was found out that they formed a normal distribution with mean 65 and standard deviation 5.4.

      (a) Find the probability that the marks scored by a randomly selected student lie within 5% of the mean.
      (b) Find the probability that the mean score of 70 randomly selected exam papers is between 60 and 70.
      (c) A group of students were surprised by the how high that average was and decided to conduct a survey to gauge the fairness of the result. Aarti took a sample of 10 students, Ayush took a sample of 15 students and Bhavya took a sample of 50 students. Find the sample mean and variance these students must have observed and plot their distributions.

   Solution

      (a) Let X be the random sample (size = 1) drawn from the data. We need to find $P(61.75 < X < 68.25)$.
      We know $X \sim N(65, 5.4)$, after normalizing X,

$$Z = (\frac{X - 65}{5.4}) \sim N(0, 1)$$

      We need to find,

$$P(\frac{61.75 - 65}{5.4} < Z < \frac{68.25 - 65}{5.4}) = P(\frac{-3.25}{5.4} < Z < \frac{3.25}{5.4})$$
$$P(-0.601 < Z < 0.601) = 2(P(Z < 0.601) - 0.5)$$

      Since the normal distribution is symmetric about its mean. Using z-table we get,

$$P(Z < 0.601) = 0.726$$

$$2(P(Z < 0.601) - 0.5) = 0.452$$

.

Therefore, the probability that thee marks scored by a randomly selected student lie within 5% of the mean is 0.452.

(b) Let $T = (X_1, X_2, ...X_{70})$ be the sample selected. Since the sample size if large (¿30) we can use the central limit theory to estimate the mean of T. We need to find, $P(60 < T < 70)$

$$\bar{T} = \frac{1}{70}\sum_{i=1}^{70} X_i$$

$$E(\bar{T}) = E(X_i) = 65$$

$$Var(\bar{T}) = \frac{Var(X_i)}{70} => S(T) = 5.4/8.367 = 0.645$$

Let,

$$Z = (\frac{\bar{T} - 65}{0.645}) \sim N(0, 1)$$

By using the Central Limit Theorem, we get

$$P(\frac{60 - 65}{0.645} < Z < \frac{70 - 65}{0.645}) = P(-7.75 < Z < 7.75) = 2(P(Z < 7.75) - 0.5) \approx 1$$

Therefore, the probability that the sample mean lies between $[60, 70]$ is nearly 1.

(c) The sample mean for all 3 of them will be same but the standard deviation $= \frac{s}{\sqrt{n}}$ where s is the standard deviation of the actual distribution and n is the sample size.

For Aarti, $s_1 = 5.4/\sqrt{10} = 1.708$, for Ayush $s_2 = 5.4/\sqrt{15} = 1.394$ and for Bhavya $s_3 = 5.4/\sqrt{50} = 0.764$.

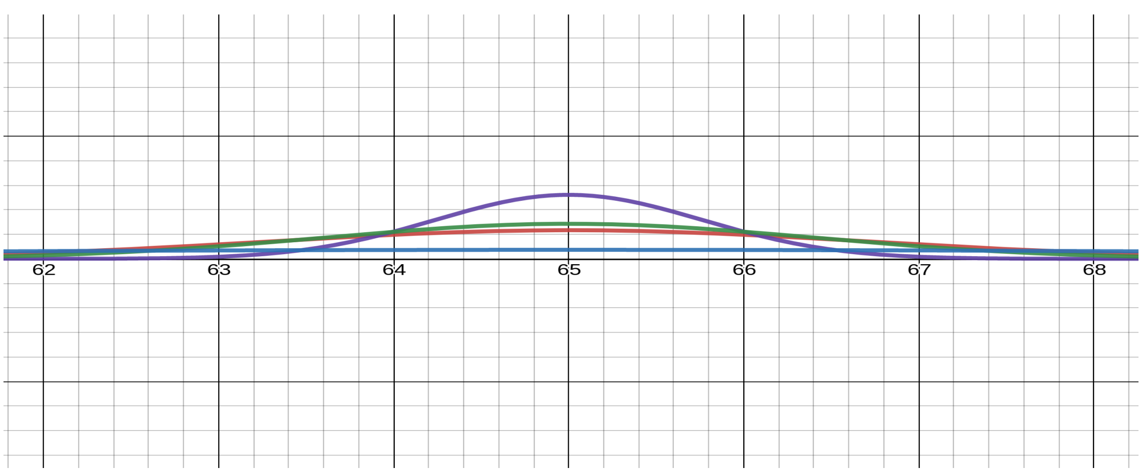

Figure 5: All the distributions

Purple - obtained by Bhavya - $N(65, 0.764)$
Green - obtained by Ayush - $N(65, 0.1.394)$
Red - obtained by Aarti - $N(65, 1.708)$
Blue - the original distribution - $N(65, 5.4)$

(b) Let $X_1, X_2, ...X_5$ be independent normally distributed random variables such that $E(X_i) = 0 \forall i = 1, 2, ...5$. Given that $E(X_1^2) = E(X_2^2) = E(X_3^2) = 16$ and $E(X_4^2 = E(X_5^2) = 64$. Find the value of a such that $P[(X_1^2 + X_2^2 + X_3^2) \leq a(X_4^2 + X_5^2)] = 0.9$.

Solution

We get that $X_i \sim N(0, 16)$ $i = 1, 2, 3$ and $X_i \sim N(0, 64)$ $i = 4, 5$. We can normalize the random variables. $\frac{X_i}{4} \sim N(0, 1)$ $i = 1, 2, 3$ and $\frac{X_i}{8} \sim N(0, 1)$ $i = 4, 5$.

We know that square of normal distributions is a χ_2 distribution with parameter 1. Therefore, $(\frac{X_i}{4})^2 \sim \chi^2(1)$ $i = 1, 2, 3$ and $(\frac{X_i}{8})^2 \sim \chi^2(1)$ $i = 4, 5$.

χ_2 distribution has reproductive property, let $T = \sum_{i=1}^{3}(\frac{X_i}{4})^2 \sim \chi_2(3)$ and $Q = \sum_{i=4}^{5}(\frac{X_i}{8})^2 \sim \chi_2(2)$.

Finally, we note that the ratio of 2 chi^2 distributed random variables divided by their respective parameters (degrees of freedom) follows the F-distribution. So,

$$F = \frac{T/3}{Q/2} \sim F(3, 2)$$

Using this we have to compute,

$$P[(X_1^2 + X_2^2 + X_3^2) \leq a(X_4^2 + X_5^2)] = 0.9$$
$$P[16T \leq 64aQ] = 0.9$$
$$P[48(T/3) \leq 128a(Q/2)] = 0.9$$
$$P[F(3, 2) \leq 8a/3] = 0.9$$

Now from the F table we can compute the value of $8a/3$. We get $P[F(3, 2) \leq 9.16] = 0.9$. And we have $8a/3 = 9.16 => a = 3.435$.

5. Point and Interval Estimations

(a) Prof. Anisha taught the course 'Introduction to Statistics' and had always received good reviews from her students. This year she decided to spice it up by giving out a challenge to her students. Each day at the end of her class she would bring a black box full of balls. Each student would be blindfolded and allowed to draw 1 ball from the box, they could do whatever they wanted with the ball and then had to put it back in. The task was to find an estimate of the number of balls in the box. She would continue to do the experiment until any student gave a close enough answer. That student would also receive a special prize. Naturally all the students were excited but also very confused. Let the total number of balls, N is a large number.

(a) Give a method by which you will be able to estimate the number of balls using an unbiased estimator.

(b) Is the above estimator consistent?

(c) Can you find a better estimator?

Solution

(a) Each day you will go and pick 1 ball from the box. Make a mark on this ball to distinguish it from the other unmarked ones and if the ball is already marked then put it back in. Continue this process until you have marked n balls and after this stop marking the balls. Now go back each day and draw a ball until you get a marked one. You can assume that the probability of getting a marked ball will be $p = n/N$ (since N is a large number this estimation is valid). Let, the number of days before you get a marked ball is X_1. Then you can show that $T = nX_1$ will be an unbiased estimator for N.

To prove this, note that $X_1 \sim Geometric(p)$ (we have constructed the experiment in this manner). The pmf of this distribution is given by -

$$f(x) = (1-p)^{k-1}p$$

for k = 1,2,3....

$$E(X_1) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = p[\sum_{k=1}^{\infty}(1-p)^{k-1} + \sum_{k=2}^{\infty}(1-p)^{k-1}\sum_{k=3}^{\infty}(1-p)^{k-1} + ...]$$

$$= p[\frac{1}{p} + \frac{(1-p)}{p} + \frac{(1-p)^2}{p} + ...]$$

$$= 1 + (1-p) + (1-p)^2 + ... = \frac{1}{p} = \frac{N}{n}$$

If we take the estimator $T = nX_1$ we get

$$E(T) = nE(X_1) = N$$

. Therefore, T is an unbiased estimator of N.

(b) The variance of T is equal to,

$$Var(T) = Var(nX_1) = n^2 Var(X_1)$$

$$Var(X_1) = E(X_1^2) - (E(X_1)^2 = \frac{(1-p)}{p^2}$$

We get,

$$Var(T) = Var(nX_1) = \frac{n^2(1-p)}{p^2}$$

Therefore, T is not a consistent estimator.

(c) We will use the same method as above, but instead of using only X_1, we will repeat the experiment k number of times which will give $X_1, X_2, ...X_k$ and $X_i \sim Geometric(p)$ for $i = 1, 2, 3....$ and $p = \frac{n}{N}$ Now, our estimator will be

$$S = n\frac{\sum_{i=1}^{k} X_i}{k}$$

We get,

$$E(S) = n\frac{\sum_{i=1}^{k} E(X_i)}{k} = n\frac{kE(X_1}{k} = nE(X_1) = N$$

also,

$$Var(S) = \frac{n^2}{k^2}\sum_{i=1}^{k} Var(X_i) = \frac{n^2}{k^2}\frac{(1-p)}{p^2}$$

$$\lim_{k\to\infty} Var(S) \to 0$$

Therefore, S is an unbiased and consistent estimator.
Both T and S are unbiased estimators, to compare them we will use relative efficiency. Since $Var(S) < Var(T)$ we get that S is the relatively efficient estimator.

(b) Suppose that $X_1, X_2, ...X_n$ are i.i.d random variables and $X_1 \sim Poisson(\lambda)$. Show that $S = \sum_{i=1}^{n} X_i$ is a sufficient statistic for λ.
Solution Let $\mathbf{X} = (X_1, X_2, ...X_n)$. We observe that,

$$P(\mathbf{X} = \mathbf{x}/S(\mathbf{X}) = s) = \frac{P(\mathbf{X} = \mathbf{x}, S(\mathbf{X}) = s)}{P(S(\mathbf{X}) = s)}$$

$$P(\mathbf{X} = \mathbf{x}, S(\mathbf{X}) = s) = \begin{cases} 0, & S(\mathbf{x}) \neq s \\ P(\mathbf{X} = \mathbf{x}), & S(\mathbf{x}) = s \end{cases} \tag{1}$$

$$P(X_i = x_i) = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

Since X_i, $i = 1, 2, 3...$ are i.i.d.

$$P(\mathbf{X} = \mathbf{x}) = \frac{e^{-n\lambda}\lambda^{S(x)}}{\prod_{i=1}^{n} x_i!}$$

We know that the Poisson distribution has the reproductive property. Therefore,

$$P(S(\mathbf{X} = s)) = \frac{e^{-n\lambda}(n\lambda)^s}{s!}$$

$$\frac{P(\mathbf{X}=\mathbf{x})}{P(S(\mathbf{X})=s)} = \frac{\frac{e^{-n\lambda}\lambda^s}{\prod_{i=1}^n x_i!}}{\frac{e^{-n\lambda}(n\lambda)^s}{s!}} = \frac{s!}{\prod_{i=1}^n x_i! n^s}$$

Since this ratio does not depend on λ, $S = \sum_{i=1}^n$ is a sufficient statistic for λ.

(c) Let $X = (X_1, X_2, ...X_{10})$ is the random sample drawn from some distribution with mean $\bar{X} = 14.5$ and standard deviation $S = 2.5$. Find the 95% confidence interval of the mean of the distribution.

Solution Since the sample mean and sample standard deviation are given, we will us students t-distribution to obtain the critical values.

We have to find a 95% confidence interval for the sample mean, so $1 - \alpha = 0.95 =>$ $\alpha = 0.05$

Sample size, $n = 10$, degrees of freedom on Student's t-distribution is $n - 1 = 9$. Using the lookup table we get

$$t_{n-1,\alpha/2} = 2.262$$

,

$$sample\, mean = 14.5$$

and

$$\frac{s}{\sqrt{n}} = \frac{2.5}{\sqrt{10}} = 0.7905$$

. Therefore the 95% confidence interval =

$$= (\bar{X} - \frac{s}{\sqrt{n}} * t_{n-1,\alpha/2},\ \bar{X} + \frac{s}{\sqrt{n}} * t_{n-1,\alpha/2})$$

$$= (12.712, 16.288)$$

6. Point and Interval Estimations

(a) Suppose that $X = (X_1, X_2, ...X_n)$ is a random sample drawn from $e^{\frac{1}{\lambda}}$ with $f(x;\lambda) = \lambda e^{-\lambda x}$ for $x > 0$.

  (a) Find the Crammer Rao lower bound on the variance of unbiased estimators of $\psi(\lambda) = \frac{1}{\lambda}$.

  (b) Determine if the sample mean is a UMVUE of $\psi(\lambda) = \frac{1}{\lambda}$.

Solution

  (a) We first compute the Fisher information of λ.

$$I(\lambda) = -E[\frac{\delta^2}{\delta\lambda^2} lnf(x;\lambda)]$$

We have,

$$f(X;\lambda) = \lambda e^{-\lambda X}$$

$$lnf(X;\lambda) = ln(\lambda) - \lambda X$$
$$\frac{\delta}{\delta\lambda} lnf(X;\lambda) = \frac{1}{\lambda} - X$$
$$\frac{\delta^2}{\delta\lambda^2} lnf(x;\lambda) = \frac{-1}{\lambda^2}$$

Also,

$$\psi(\lambda) = \frac{1}{\lambda}$$
$$\psi^{'}(\lambda) = \frac{-1}{\lambda^2}$$
$$(\psi^{'}(\lambda))^2 = \frac{1}{\lambda^4}$$

Now the Crammer Rao Lower Bound (CRLB) is defined as,

$$CRLB = \frac{(\psi^{'}(\lambda))^2}{nI(\lambda)} = \frac{\frac{1}{\lambda^4}}{\frac{n}{\lambda^2}} = \frac{1}{n\lambda^2}$$

(b) We first verify that sample mean is an unbiased estimator.

$$E(\bar{X}) = \frac{\sum_{i=1}^{n} E(X_i)}{n}$$

Since all X_i are iid random variables,

$$E(\bar{X}) = \frac{n * E(X_i)}{n} = E(X_i) = \frac{1}{\lambda} = \psi(\lambda)$$

Now we compute the variance and check if it is equal to the CRLB.

$$Var(\bar{X}) = Var(\frac{\sum_{i=1}^{n} X_i}{n})$$
$$Var(\bar{X}) = \frac{Var(\sum_{i=1}^{n} X_i)}{n^2} = \frac{nVar(X_i)}{n^2}$$
$$= \frac{Var(X_i)}{n} = \frac{1}{n\lambda^2} = CRLB$$

Therefore, the sample mean is a UMVUE of $\psi(\lambda) = \frac{1}{\lambda}$

(b) Consider a distribution with the probability density function, $f(x) = \frac{1}{\beta} e^{\frac{-|x-\theta|}{\alpha}}$ for $x, \theta \in R$ and $\alpha > 0$.

(a) Find value of β such that $f(x)$ is a probability density function.

(b) Suppose $X = (X_1, X_2, ...X_n)$ is a random sample drawn from the distribution. Find estimates for the parameters θ and α.

Solution

(a) Since $f(x)$ is a probability density function, we know

$$\int_{-\infty}^{\infty} \frac{1}{\beta} e^{\frac{-|x-\theta|}{\alpha}} dx = 1$$

$$\int_{\theta}^{\infty} \frac{1}{\beta} e^{\frac{-|x-\theta|}{\alpha}} dx + \int_{-\infty}^{\theta} \frac{1}{\beta} e^{\frac{-|x-\theta|}{\alpha}} dx = 1$$

$$\frac{1}{\beta} = 2\alpha$$

$$\beta = 2\alpha$$

(b) We will do this using 2 methods.
Method of moments

$$E(X_1) = \int_{-\infty}^{\infty} \frac{1}{2\alpha} x e^{\frac{-|x-\theta|}{\alpha}} dx = \frac{1}{2} \int_{-\infty}^{\infty} (\alpha t + \theta) e^{-|s|} dt = \theta$$

$$E(X_1^2) = \int_{-\infty}^{\infty} \frac{1}{2\alpha} x^2 e^{\frac{-|x-\theta|}{\alpha}} dx = \frac{1}{2} \int_{-\infty}^{\infty} (\alpha t + \theta)^2 e^{-|s|} dt = 2\alpha^2 + \theta^2$$

We can equate,

$$Sample\ mean(\bar{X}) = \theta$$

$$Sample\ variance(S^2) = E(X_1^2 - (E(X_1)^2) = 2\alpha^2$$

$$\alpha = S/\sqrt{(}2)$$

Maximum Likelihood Estimate We compute the likelihood function of the data,

$$L(\theta, \alpha) = \prod_{i=1}^{n} \frac{1}{2\alpha} e^{\frac{-|X_i-\theta|}{\alpha}} = 2^{-n} \alpha^{-n} e^{\sum_{i=1}^{n} frac-|X_i-\theta|\alpha}$$

The log-likelihood function is

$$ln(L(\theta, \alpha)) = -nln2 - nln\alpha - \frac{1}{\alpha} \sum_{i=1}^{n} |X_i - \theta|$$

We need to maximize the log-likelihood function. Observe that $sum_{i=1}^{n} |X_i - \theta|$ is minimum when X_i is the median.

$$\sum_{i=1}^{n} |X_i - \theta| = |X_1 - \theta| + |X_2 - \theta| + ... + |X_n - \theta|$$

$$|X_n - X_1| <= |X_n - \theta| + |X_1 - \theta|$$

$$|X_{n-1} - X_2| <= |X_{n-1} - \theta| + |X_2 - \theta|$$
$$|X_{n-2} - X_3| <= |X_{n-2} - \theta| + |X_3 - \theta|$$
$$....$$

and so on.
All inequalities become equalities only when θ is the median of all X_i.
Hence, we get that

$$\theta = M$$

where M is the median of X.
After putting $\theta = M$ in the log-likelihood function we can differentiate with respect to α.

$$\frac{\delta log(L(\theta, \alpha))}{\delta \alpha} = \frac{-n}{\alpha} + \frac{1}{\alpha^2} \sum_{i=1}^{n} |X_i - M|$$

Equating this to 0, we get

$$\alpha = \frac{1}{n} \sum_{i=1}^{n} |X_i - M|$$

Hence, we have obtained the estimates for parameters θ and α.