MTL 390 (Statistical Methods)
Minor Examination Assignment 1 Report

Name: Arpit Saxena Entry Number: 2018MT10742

1. Descriptive Statistics

x_1	x_2
0.199072	1.227659
-0.668432	-1.050136
1.071085	-0.897862
0.507552	-1.056094
0.434084	-0.063376
-2.059958	-0.159318
1.329366	-0.388192
1.734604	0.834843
-1.982315	0.360589
-2.721066	-2.263463

2. Descriptive Statistics

3. Sampling Distributions

At an organization, there are 10 systems and each one has a different processing power. Suppose there are 10 jobs whose loads (as defined by some metric) are randomly picked from the interval $(0, 100)$. One job is assigned to one system, and the assignment is done by arranging the systems and jobs in an increasing order (of processing power and load respectively) and assign each job to the system at the same index in the other list.

Find the probability distribution of the load at each system. Hence or otherwise, find the expected load of the job ending up on the third system (arranged in increasing order).

Answer

Let $X_1, \ldots, X_{10}$ be samples from the distribution $U(0, 1)$. Then $100X_i \sim U(0, 100)\,\forall i = 1, \ldots, 10$ would represent the job loads. The ordering amongst X_i's and $100X_i$'s is the same, so we work with X_i's to make our life easier.

Suppose X_i's arranged in ascending order are $X_{(1)}, X_{(2)}, \ldots, X_{(10)}$.

Let $k \in \{1, \ldots, 10\}$ and $n = 10$. Also, let $\epsilon > 0$ be small. Then:

$$\begin{aligned} P(X_{(k)} \in [x, x+\epsilon]) = P(&\text{one of } X_1, \ldots, X_{10} \text{ lies in } [x, x+\epsilon] \\ &\text{and exactly } k-1 \text{ are less than } x) \end{aligned}$$

(Now using the fact that X_i's are independent)

$$= \binom{n}{1} P(\text{one lies in } [x, x+\epsilon]) P(\text{exactly } k-1 \text{ are less than } x)$$

$$
\begin{aligned}
&= n\epsilon\binom{n-1}{k-1} P(X_1 \le x)^{k-1} P(X_1 > x)^{n-k} && \text{(Since pdf is } f(x) = 1\text{)} \\
&= n\epsilon \frac{(n-1)!}{(k-1)!\,(n-k)!} x^{k-1}(1-x)^{n-k} \\
\implies f_{X_{(k)}}(x) &= \frac{x^{k-1}(1-x)^{n-k}}{\frac{(k-1)!\,(n-k)!}{n!}} \\
&= \frac{x^{k-1}(1-x)^{n-k}}{\frac{\Gamma(k)\Gamma(n-k+1)}{\Gamma(n+1)}} \\
&= \frac{x^{k-1}(1-x)^{n-k}}{B(k, n-k+1)} && \text{(where } B \text{ is the Beta function)}
\end{aligned}
$$

We note that

$$
X \sim B(\alpha, \beta) \implies f_X(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}
$$

where $B(\cdot,\cdot)$ is the Beta-distribution.

Comparing the equations of Beta distribution's pdf and the pdf we obtained for $X_{(k)}$, we conclude that

$$
X_{(k)} \sim B(k, n-k+1)\, \forall k \in \{1, \dots, 10\}
$$

Now, we find the expected value of $X_{(3)}$ as:

$$
\begin{aligned}
E(X_{(k)}) &= \int_0^1 x \times \frac{x^{k-1}(1-x)^{n-k}}{B(k, n-k+1)} dx \\
\implies E(X_{(3)}) &= \int_0^1 x \times \frac{x^2(1-x)^8}{B(3,8)} dx && (\because n = 10) \\
&= \frac{1}{B(3,8)} \int_0^1 x^3(1-x)^8 dx \\
&= \frac{1}{B(3,8)} \int_0^1 (1-x)^3 x^8 dx \\
&= \frac{1}{B(3,8)} \int_0^1 x^8(1 - 3x + 3x^2 - x^3) dx \\
&= \frac{\Gamma(11)}{\Gamma(3)\Gamma(8)} \int_0^1 (x^8 - 3x^9 + 3x^{10} - x^{11}) dx \\
&= \frac{10!}{2!\,7!} \left[\frac{1}{9} - \frac{3}{10} + \frac{3}{11} - \frac{1}{12}\right] \\
&= 360 \left[\frac{1}{36} - \frac{3}{110}\right] \\
&= 0.182
\end{aligned}
$$

Therefore, the expected load of the job ending up at the 3rd system is $100E(X_{(k)}) = 18.2$

4. Sampling Distributions

5. Point and Interval Estimations

   For a distribution with k unknown parameters, method of moments uses k moments to form a system of equations and solves it to find estimates for the parameters. This throws away information contained in higher order moments. To remedy that, the **Generalized Method of Moments (GMM)** takes $q(> k)$ moments and minimizes the sum of squares of difference between sample moments and moments calculated from the distribution.

   Consider the following samples taken from a Poisson distribution with unknown λ. Find an estimate for the parameter using both method of moments as well as generalized method of moments (with 3 moments)

30	21	24	18
28	25	24	25
26	19	19	21
22	34	22	15
22	25	16	22

Answer

We first calculate expressions of three moments $E[X], E[X^2]$ and $E[X^3]$ for $X \sim P(\lambda)$.

Using the MGF of the Poisson distribution, we find moments around 0:

$$\begin{aligned}
&M_X(t) = exp(\lambda(e^t - 1)) \\
\implies &M_X'(t) = \lambda e^t exp(\lambda e^t - \lambda) \\
\implies &M_X''(t) = (\lambda e^t)^2 exp(\lambda e^t - \lambda) + \lambda e^t exp(\lambda e^t - \lambda) \\
\implies &M_X'''(t) = (\lambda e^t)^3 exp(\lambda e^t - \lambda) + 2(\lambda e^t)^2 exp(\lambda e^t - \lambda) \\
&\qquad + (\lambda e^t)^2 exp(\lambda e^t - \lambda) + \lambda e^t exp(\lambda e^t - \lambda) \\
&\qquad = (\lambda e^t)^3 exp(\lambda e^t - \lambda) + 3(\lambda e^t)^2 exp(\lambda e^t - \lambda) + \lambda e^t exp(\lambda e^t - \lambda)
\end{aligned}$$

Using these, we calculate the moments as:

$$\begin{aligned}
E[X] &= M_X'(0) = \lambda \\
E[X^2] &= M_X''(0) = \lambda^2 + \lambda \\
E[X^3] &= M_X'''(0) = \lambda^3 + 3\lambda^2 + \lambda
\end{aligned}$$

Next we calculate the sample moments. Let the samples be written as $x_1, ..., x_{20}$. Then:

$$m_1 = \sum_{i=1}^{20} x_i = 22.9$$
$$m_2 = \sum_{i=1}^{20} {x_i}^2 = 544.4$$
$$m_3 = \sum_{i=1}^{20} {x_i}^3 = 13424.5$$

Using the method of moments, we get:

$$\widehat{\lambda}_1 = m_1 = 22.9$$

For the generalized method of moments, we note that we can take any weighting of the sample moments. In fact we can also take a positive definite matrix and define the cost function that way. Suppose we somehow decided to keep the weights for m_1, m_2, m_3 to be $100, 10, 1$ respectively. Then, we have the function:

$$\begin{aligned} Q(\lambda) &= 100(m_1 - E[X])^2 + 10(m_2 - E[X^2])^2 + 1(m_3 - E[X^3])^2 \\ &= 100(22.9 - \lambda)^2 + 10(544.4 - \lambda^2 - \lambda)^2 + (13424.5 - \lambda^3 - 3\lambda^2 - \lambda)^2 \end{aligned}$$

Then the estimator is given by:

$$\widehat{\lambda}_2 = \operatorname*{argmin}_{\lambda} Q(\lambda)$$

We note that $Q(\lambda)$ is a polynomial in lambda with degree 6, so it's not practical to calculate the minimum by hand. Using computer tools, we find:

$$\widehat{\lambda}_2 = 22.79$$

6. Point and Interval Estimations

   Find an unbiased estimator for the parameter λ of an exponential distribution $\mathrm{Exp}(\lambda)$ as a multiple of the sample mean. Given that the estimator found is the UMVUE, show that it does not achieve equality in the Cramer-Rao inequality, consequently showing that achieving the Cramer Rao Lower Bound is not a necessary condition for being UMVUE.

   Customers arrive at a checkout counter with an average time of 10 minutes as observed from 30 customers. What is the time in which their order should be processed so that 70% of the customers don't find a line at the checkout counter.

Answer

Let $X_1, \ldots, X_n \sim \text{Exp}(\lambda)$ be n samples from an exponential distribution with parameter λ. Now, let

$$
\begin{aligned}
T(X) &= \frac{c}{\overline{X}} \\
&= \frac{cn}{\sum_{i=1}^{n} X_i}
\end{aligned}
$$

Using the fact that sum of exponentially distributed variables follows the Gamma distribution, we define $Z = \sum_{i=1}^{n} X_i$ and note that $Z \sim \text{Gamma}(n, \lambda)$. Then $T(X) = \frac{cn}{Z}$

$$
\begin{aligned}
E(T(X)) &= E\left(\frac{cn}{Z}\right) \\
&= cnE\left(\frac{1}{Z}\right) \\
&= cn \int_0^\infty \frac{1}{z} \frac{\lambda^n z^{n-1} e^{-\lambda z}}{\Gamma(n)} dz && (\because Z \sim Gamma(n, \lambda)) \\
&= \frac{cn\lambda}{n-1} \underbrace{\int_0^\infty \frac{\lambda^{n-1} z^{n-2} e^{-\lambda z}}{\Gamma(n-1)} dz}_{=1} && (\because \Gamma(n) = (n-1)\Gamma(n-1)) \\
&= \frac{cn\lambda}{n-1}
\end{aligned}
$$

To make this estimator unbiased, we have

$$
\begin{aligned}
E(T(X)) &= \lambda \\
\implies \frac{cn\lambda}{n-1} &= \lambda \\
\implies c &= \frac{n-1}{n} \\
\therefore T(X) &= \frac{n-1}{\sum_{i=1}^{n} X_i} && \text{Since } T(X) = \frac{cn}{\sum_{i=1}^{n} X_i}
\end{aligned}
$$

Thus, we have found an unbiased estimator for T(X).

Now, for Cramer-Rao inequality:

$$
E\left(T(X)^2\right) = E\left(\left(\frac{n-1}{\sum_{i=1}^{n} X_i}\right)^2\right)
$$

$$
\begin{aligned}
&= (n-1)^2 E\left(\frac{1}{Z^2}\right) \\
&= (n-1)^2 \int_0^\infty \frac{1}{z^2}\frac{\lambda^n z^{n-1} e^{-\lambda z}}{\Gamma(n)} dz && (\because Z \sim \text{Gamma}(n, \lambda)) \\
&= \frac{(n-1)^2\lambda^2}{(n-1)(n-2)} \underbrace{\int_0^\infty \frac{\lambda^{n-2} z^{n-3} e^{-\lambda z}}{\Gamma(n-2)} dz}_{=1} && (\because \Gamma(n) = (n-1)\Gamma(n-1)) \\
&= \frac{(n-1)\lambda^2}{n-2}
\end{aligned}
$$

Using this, we find the variance:

$$
\begin{aligned}
Var(T(X)) &= E(T(X)^2) - (E(T(X)))^2 \\
&= \frac{(n-1)\lambda^2}{n-2} - \lambda^2 \\
&= \lambda^2\left(\frac{n-1}{n-2} - 1\right) \\
\implies Var(T(X)) &= \frac{\lambda^2}{n-2} \qquad (1)
\end{aligned}
$$

Now, we find the information for X_1:

$$
\begin{aligned}
I_1(\lambda) &= E\left[\left(\frac{\partial}{\partial\lambda}\log f(X_1;\lambda)\right)^2\right] \\
&= E\left[\left(\frac{\partial}{\partial\lambda}\log\left\{\lambda e^{-\lambda X}\right\}\right)^2\right] \\
&= E\left[\left(\frac{\partial}{\partial\lambda}\left\{\log\lambda - \lambda X\right\}\right)^2\right] \\
&= E\left[\left(\frac{1}{\lambda} - X\right)^2\right] \\
&= Var(X) && (\because E(X) = \tfrac{1}{\lambda}) \\
&= \frac{1}{\lambda^2}
\end{aligned}
$$

From that we get total information of the sample as:

$$
I(\lambda) = \frac{n}{\lambda^2} \qquad (2)
$$

$$
\text{LHS of Cramer Rao inequality} = Var(T(X)) = \frac{\lambda^2}{n-2} \qquad \text{From (1)}
$$

$$\text{RHS of Cramer Rao inequality} = \frac{1}{I(\lambda)} = \frac{\lambda^2}{n} \qquad \text{Since } E(T(X)) = \lambda \text{ and from (2)}$$

Thus, LHS of Cramer Rao inequality $\neq$ RHS of Cramer Rao inequality

Now, assuming that the inter-arrival time of customers at the checkout counter is independent, we can model the inter-arrival time as an exponential distribution.

We find an estimate for the parameter λ using the estimator derived previously as:

$$\widehat{\lambda} = \frac{n-1}{\sum_{i=1}^{n} X_i} = \frac{n-1}{n\overline{X}} = \frac{29}{30} \times 6 \text{ hour}^{-1} = 5.8 \text{ hour}^{-1}$$

Let t be the time to process one customer's order at the counter, then we want $P(X > t) >= 0.7$ where $X \sim Exp(\widehat{\lambda}) = Exp(5.8)$.

$$\begin{aligned}
& P(X > t) \geq 0.7 \\
\implies & e^{-\widehat{\lambda}t} \geq 0.7 \\
\implies & \widehat{\lambda}t \leq -\log 0.7 \\
\implies & t \leq -\frac{\log 0.7}{\widehat{\lambda}} \\
\implies & t \leq 0.062 \text{ hours} = 3.72 \text{ minutes}
\end{aligned}$$

Thus, each order should be processed in not more than 3.72 minutes for 70% of the customers to not have to wait in line at the counter.