MATH 8050 Homework 1

Parampreet Singh, C19377466

Due February 7, 2024 at 11:59pm

Instructions: For the problems that follow, please show all of your work. For problems that you solve using software, please provide the code you used either with the problem or in an appendix at the end of your homework document.

Problem 1

Suppose that we studied 2 random variables, X and Y. We know that X has expectation μ_X and standard deviation σ_X. Similarly, Y has expectation μ_Y and standard deviation σ_Y.

Now suppose that we need to convert the units of our data, so we define two transformations of our original random variables, $Z = a - Y$ and $U = bX + d$, where a, b, and d are constants.

(a) Express the expectation and variance of Z and U in terms of those for X and Y.
(b) Given the above information, can we find the expectation and variance of U^2? If so, provide the expectation and variance of U^2.

Problem 2

R has several built-in data sets. We will study one of the variables contained in the `USArrests` data set. To load the data into your R environment, run the following in R:

```
data("USArrests")
```

We will look at the `UrbanPop` variable, which contains the percent of people in each state that live in urban areas.

(a) Plot the `UrbanPop` variable.
(b) Find the mean and standard deviation of `UrbanPop`. Use these to fit a Normal density to `UrbanPop`, and provide a plot of the densities.
(c) Construct and plot a cumulative distribution function (CDF) for `UrbanPop` based on the Normal densities you found in part (b).
(d) Calculate the probability that the percentage of people in a state living in urban areas is less than or equal to 60.
(e) Calculate the probability that the the percentage of people in a state living in urban areas is between 50 and 80.
(f) What is the 75_{th} percentile for the `UrbanPop` variable?

Problem 3

Suppose we have a class of 20 students taking an exam. Assume that each student has a 70% chance of passing the exam, and that the students' test scores are independent of one another.

(a) How many of the students do we expect to pass the exam?
(b) What is the variance associated with the number of students who pass the exam?
(c) What is the probability that only 3 students pass the exam?
(d) What is the probability that at least 15 students pass the exam?

Problem 4

Suppose we have a random variable X with density function:

$$f(x|\lambda) = \lambda e^{-\lambda x},\ x > 0$$

.

(a) Derive the cumulative distribution function (CDF) for X.

Now suppose we have n independent and identically distributed (iid) random variables $X_1, \ldots, X_n$ with the following density function:

$$f(x_i|\lambda) = \lambda e^{-\lambda x_i},\ x_i > 0$$

(b) What is the joint likelihood function of $X_1, \ldots, X_n$?

(c) What is the maximum likelihood estimator for the parameter λ?

MATH 8050 HOMEWORK 1

Ans 1. given

$X, Y \rightarrow$ Random variables

$E(X) = \mu_X$ $\quad$ $std(X) = \sigma_X$

$E(Y) = \mu_Y$ $\quad$ $std(Y) = \sigma_Y$

$Z = a - Y$ $\quad$ $U = bX + d$

where,

a, b & d are constants

(a) $E(Z) = E(a - Y)$

$= E(a) - E(Y)$

$= a - E(Y)$ $\quad$ ($\because$ a is constant)

$\boxed{E(Z) = a - \mu_Y}$ $\quad$ ($\because E(Y) = \mu_Y$)

$E(U) = E(bX + d)$

$= E(bX) + E(d)$

$= b\,E(X) + d$ $\quad$ ($\because$ b, d are constants)

$\boxed{E(U) = b\mu_X + d}$ $\quad$ ($\because E(X) = \mu_X$)

Variance $\rightarrow$

$Var(Z) = Var(a - Y)$

$= Var(-Y)$ $\quad$ ($\because$ a is constt)

$= (-1)^2\, Var(Y)$

$\boxed{Var(Z) = \sigma_Y^2}$ $\quad$ ($\because Var(Y) = (std\,Y)^2 = \sigma_Y^2$)

$$Var(U) = Var(bX + d)$$
$$= b^2 Var(X) \quad (\because b, d \text{ are const.})$$
$$\boxed{Var(U) = b^2 \sigma_x^2}$$

(b) To find the expectation of ~~a r.v. X~~ ~~as a~~ a function of r.v X, we use

$$E(g(X)) = \int g(x) f_X(X) \, dx$$

where,

$g(X)$ = function of r.v X

$f_X(x)$ = probability density function over all values of X

In our case,

$$g(X) = U^2 = (bX + d)^2$$

for $f_X(x)$, we need to know the distribution of X, which is unknown. Therefore the best representation that could given is

$$\boxed{E(U^2) = \int (bX + d)^2 f_X(X) \, dx}$$

For $Var(U^2)$ we use

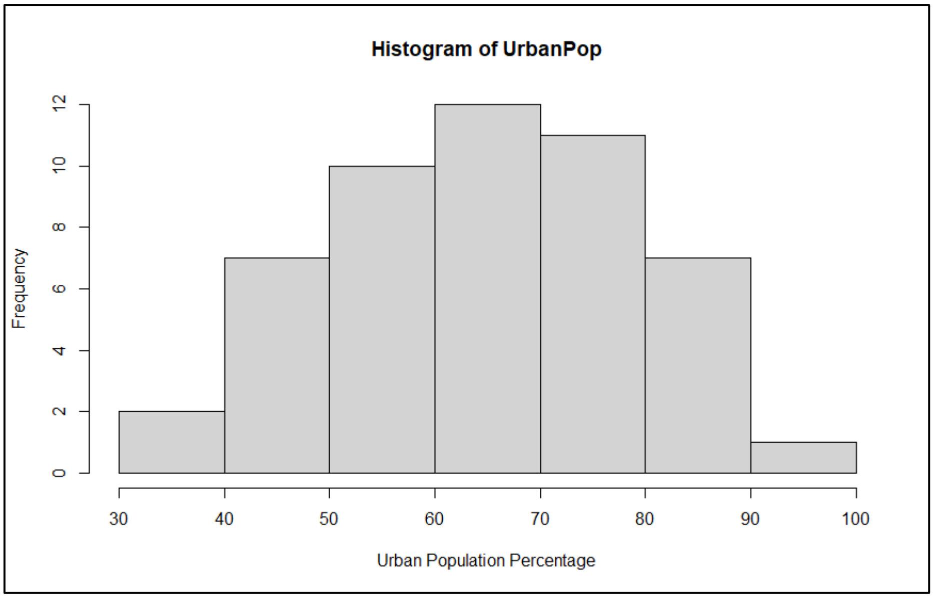
$$Var(U^2) = E(U^4) - [E(U^2)]^2$$

Finding the fourth moment of U depends upon the distribution of r.v X which is unknown.

Hence, with the given information, above provided expressions are the best representation that could be given for Expectation & variance of U^2.

Ans 2 –

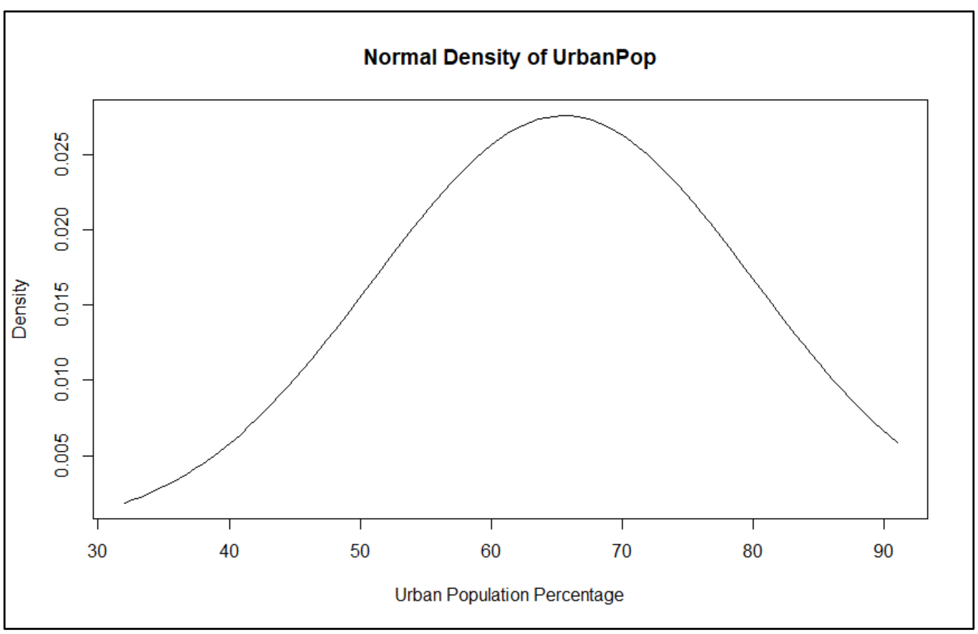
(A) - Plot the UrbanPop variable



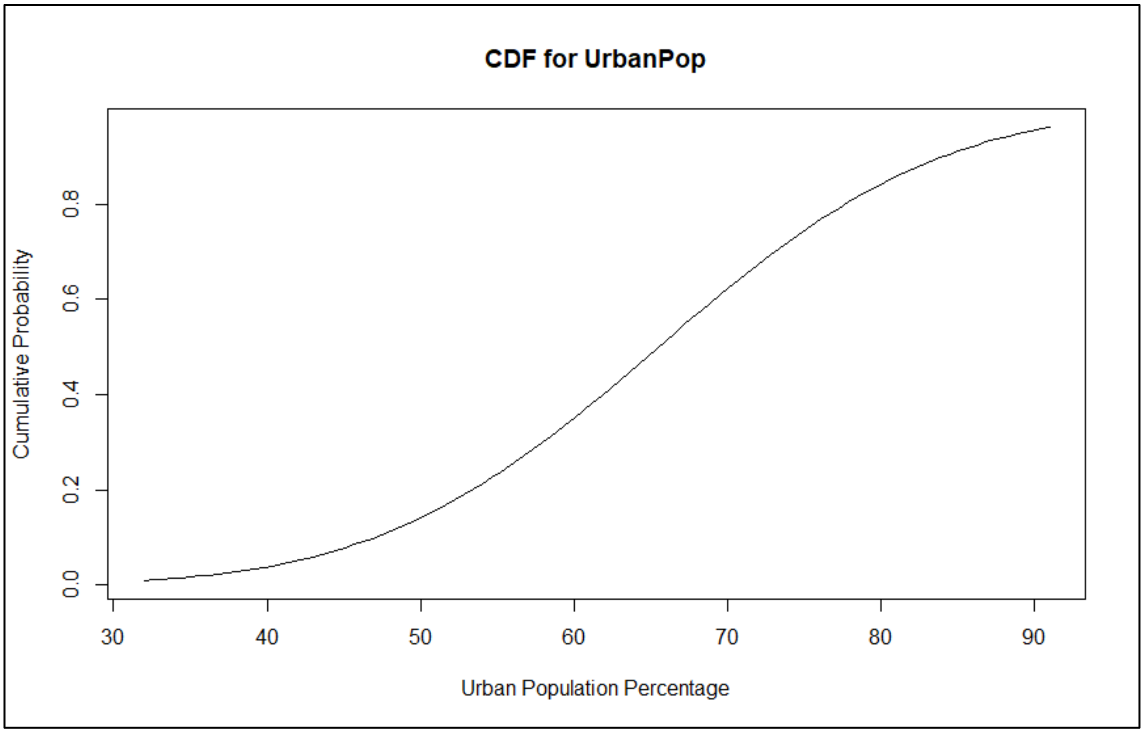
(B) - Find the mean and standard deviation of UrbanPop and fit a Normal density

Mean of UrbanPop: **65.54**

Standard Deviation of UrbanPop: **14.475**



(C) - Construct and plot a cumulative distribution function (CDF) for UrbanPop



(D)- Probability less than or equal to 60: **0.35095**

(E) - Probability between 50 and 80: **0.69959**

(F) - 75th percentile of UrbanPop: **75.3031**

```r
# Load the USArrests dataset
data("USArrests")

# Part (a): Plot the UrbanPop variable.
hist(USArrests$UrbanPop, main="Histogram of UrbanPop", xlab="Urban Population Percentage")

# Part (b): Find the mean and standard deviation of UrbanPop and fit a Normal density.
mean_urban_pop <- mean(USArrests$UrbanPop)
sd_urban_pop <- sd(USArrests$UrbanPop)

# Density Plot for UrbanPop
x <- seq(min(USArrests$UrbanPop), max(USArrests$UrbanPop), length=100)
y <- dnorm(x, mean=mean_urban_pop, sd=sd_urban_pop)
plot(x, y, type="l", main="Normal Density of UrbanPop", xlab="Urban Population
Percentage", ylab="Density")

# Part (c): Construct and plot a cumulative distribution function (CDF) for UrbanPop.
y_cdf <- pnorm(x, mean=mean_urban_pop, sd=sd_urban_pop)
plot(x, y_cdf, type="l", main="CDF for UrbanPop", xlab="Urban Population Percentage",
ylab="Cumulative Probability")

# Part (d): Calculate the probability that the percentage of people in a state living in
urban areas is less than or equal to 60.
prob_less_than_60 <- pnorm(60, mean=mean_urban_pop, sd=sd_urban_pop)

# Part (e): Calculate the probability that the percentage of people in a state living in
urban areas is between 50 and 80.
prob_between_50_80 <- pnorm(80, mean=mean_urban_pop, sd=sd_urban_pop) - pnorm(50,
mean=mean_urban_pop, sd=sd_urban_pop)

# Part (f): What is the 75th percentile for the UrbanPop variable?
percentile_75th <- qnorm(0.75, mean=mean_urban_pop, sd=sd_urban_pop)

# Print the results
print(paste("Mean of UrbanPop:", mean_urban_pop))
print(paste("Standard Deviation of UrbanPop:", sd_urban_pop))
print(paste("Probability less than or equal to 60:", prob_less_than_60))
print(paste("Probability between 50 and 80:", prob_between_50_80))
print(paste("75th percentile of UrbanPop:", percentile_75th))
```

Ans 3- Given,

Binomial Distribution

No. of trials $n = 20$

~~Pro~~ Probability of passing $p = 0.7$

Probability of not passing $1-p = 0.3$

(a) Expected no. of students passing

$$E(X) = np = 20 \times 0.7$$
$$= 14$$

(b) Variance associated

$$Var(X) = np(1-p)$$
$$= 20 \times 0.7 \times 0.3$$
$$= 4.2$$

(c) Probability that only 3 students pass

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$P(X = 3) = \binom{20}{3} (0.7)^3 (0.3)^{17}$$

$$P(X = 3) \simeq 5.05 \times 10^{-7}$$

(d) Probability atleast 15 students pass the exam

$$P(X \geq 15) = \sum_{k=15}^{20} \binom{20}{k} (0.7)^k (0.3)^{20-k}$$

$$P(X \geq 15) \simeq 0.416$$

Ans 4. Given,

$$f(x|\lambda) = \lambda e^{-\lambda x}, \quad x > 0$$

(a) CDF $F_X(x|\lambda) = \int_0^x f(t|\lambda)\,dt$

$$= \int_0^x \lambda e^{-\lambda t}\,dt$$

$$= \lambda \int_0^x e^{-\lambda t}\,dt$$

$$= -\frac{\lambda}{\lambda}\left[e^{-\lambda t}\right]_0^x$$

$$F(x|\lambda) = 1 - e^{-\lambda x}$$

(b) For independent and identically distributed random variables $X_1, \ldots X_n$ with given pdf, the joint likelihood function $L(\lambda)$ is the product of individual pdfs :

$$L(\lambda) = \prod_{i=1}^{n} f(x_i \mid \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$$

(C) To find MLE for λ, we take natural log of the joint likelihood function

$$\ell(\lambda) = \ln(L(\lambda))$$

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}$$

$$= (\lambda e^{-\lambda x_1}).(\lambda e^{-\lambda x_2}) \ldots (\lambda e^{-\lambda x_n})$$

$$= \lambda^{n} e^{-\lambda \sum_{i=1}^{n} x_i}$$

From properties of exponents

let,
$$Y = \sum_{i=1}^{n} x_i$$

$$L(\lambda) = \lambda^{n} e^{-\lambda Y}$$

$$\ell(\lambda) = \ln(\lambda^{n} e^{-\lambda Y})$$
$$\ell(\lambda) = n \ln(\lambda) - \lambda(Y)$$

To find MLE, differentiate $\ell(\lambda)$ w.r.t λ

$$\frac{d\, l(\lambda)}{d\lambda} = \frac{n}{\lambda} - y$$

Now, equate $\frac{d\, l(\lambda)}{d\lambda} = 0$

$$\frac{n}{\lambda} - y = 0$$

$$\frac{n}{\lambda} = y$$

$$\lambda = \frac{n}{y}$$

$$\lambda = \frac{n}{\sum_{i=1}^{n} x_i} \qquad \left(\because y = \sum_{i=1}^{n} x_i\right)$$

$$\lambda = \frac{1}{\left(\frac{\sum_{i=1}^{n} x_i}{n}\right)}$$

$$\boxed{\hat{\lambda} = \frac{1}{\bar{x}}}$$ is the MLE