STATISTICS 641 - ASSIGNMENT 2

DUE DATE: Noon (CDT), MONDAY, SEPTEMBER 20, 2021

Name ____

Email Address ____

Please TYPE your name and email address. Often we have difficulty in reading the handwritten names and email addresses. Make this cover sheet the first page of your Solutions.

STATISTICS 641 - ASSIGNMENT #2 - Due Noon, Monday - 9/20/21

- Read Handout 3
- Supplemental reading: Chapter 2, 3, 4 in the Devore's book
- Hand in the following Problems:

(1.) (10 points) Assume that the random variable Y has pmf with parameter $p,\ 0 < p < 1$:

$$f(y) = \begin{cases} p(1-p)^y & \text{for } \ y = 0, 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

(a.) Find the cdf, $F(y)$ for Y

Hint: $\sum_{k=0}^{m} ab^k = a\frac{1-b^{m+1}}{1-b}$

(b.) Find the 80th percentile of $F(y)$ if $p = .4$. That is, evaluate $Q(.8)$ for $p = .4$

(2.) (20 points) Let Y have a 3-parameter Weibull distribution, that is, Y has pdf and cdf in the following form with $\alpha > 0,\ \gamma > 0,\ \theta > 0$:

$$f(y) = \begin{cases} \frac{\gamma}{\alpha^\gamma}(y-\theta)^{\gamma-1} e^{-\left(\frac{y-\theta}{\alpha}\right)^\gamma} & \text{for} \quad y \geq \theta \\ 0 & \text{for} \quad y < \theta \end{cases} \qquad F(y) = \begin{cases} 1 - e^{-\left(\frac{y-\theta}{\alpha}\right)^\gamma} & \text{for} \quad y \geq \theta \\ 0 & \text{for} \quad y < \theta \end{cases}$$

(a.) Verify that the pair $(\theta,\ \alpha)$ are location-scale parameters for this family of distributions.

(b.) Derive the quantile function for the three parameter Weibull family of distributions.

(c.) What is the probability that a random selected value from a Weibull distribution with $\theta = 5,\ \gamma = 4$ and $\alpha = 20$ has value greater than 23?

(d.) Compute the 25th percentile from a Weibull distribution with with $\theta = 5,\ \gamma = 4$ and $\alpha = 20$.

(3.) (10 points) An alternative form of the 2-parameter Weibull distribution is given as follows with parameters $\beta > 0,\ \gamma > 0$

$$f(y) = \begin{cases} \frac{\gamma}{\beta} y^{\gamma-1} e^{-y^\gamma/\beta} & \text{for} \quad y \geq 0 \\ 0 & \text{for} \quad y < 0 \end{cases} \qquad F(y) = \begin{cases} 1 - e^{-y^\gamma/\beta} & \text{for} \quad y \geq 0 \\ 0 & \text{for} \quad y < 0 \end{cases}$$

(a.) Show that β is not a scale parameter for this family of distributions.

(b.) Show that $\alpha = \beta^{1/\gamma}$ is a scale parameter for this family of distributions.

(4.) (10 points) An experiment measures the number of particle emissions from a radioactive substance. The number of emissions has a Poisson distribution with rate $\lambda = .15$ particles per week.

(a.) What is the probability of at least 2 emission occurring in a randomly selected week?

(b.) What is the probability of at least 2 emission occurring in a randomly selected year?

(5.) (10 points) Let $Z_1,\ Z_2,\ Z_3,\ Z_4,\ Z_5,\ Z_6,\ Z_7,\ Z_8$ be independent N(0,1) r.v.'s. Identify the distributions of the random variables, A, B, C, D by providing the name of the distribution and the appropriate degrees of freedom, if needed.

(a.) $A = Z_7/\sqrt{[Z_1^2 + Z_2^2 + Z_3^2]/3}$.

(b.) $B = Z_5/Z_6$.

(c.) $C = Z_1^2 + Z_2^2 + Z_3^2$

(d.) $D = 3(Z_4^2 + Z_5^2 + Z_6^2 + Z_7^2)/4[Z_1^2 + Z_2^2 + Z_3^2]$

(e.) $E = 3Z_1^2/[Z_2^2 + Z_3^2 + Z_4^2)]$.

(6.) (10 points) Let $U = .26$ be a realization from a Uniform on (0,1) distribution.

Express a single realization from each of the following distributions using just the fact $U = .26$.

(a.) W = Weibull(γ=4, α=1.5)

(b.) N = NegBin(r=8, p=.7)

(c.) B = Bin(20, .4)

(d.) P = Poisson(λ=3)

(e.) U = Uniform on (0.3, 2.5)

(7.) (30 points) For each of the following situations described below, select the distribution which best models the given situation (you may need to use some distributions multiple times). Provide a very short justification for your answer.

Hypergeometric	Equally Likely	Poisson	Binomial	Geometric	Negative Binomial	Normal
Uniform	Gamma	Exponential	Chi-square	Lognormal	Gamma	Exponential
Chi-square	Lognormal	Cauchy	Weibull	F	t	Beta

(a.) In an epidemiological study of the incidence of skin cancer for those individuals who use artificial tanning procedures, a researcher randomly selected 100 of the 10,000 customers of a large tanning facility. The 100 customers were examined and the number N who had developed skin cancer was recorded. The distribution of N is is ___?

(b.) A civil engineer has determined that micro-cracks in bridge support columns occur according to a Poisson process with an average rate of 30 micro-cracks in a 10 feet length of column. She wants to assess the possible distance, D, between micro-cracks in a column. The distribution of D is is ___?

(c.) In an epidemiological study of the incidence of chronic fatigue within those individuals who are fully employed and pursing a college degree, a researcher decided to interview 500 of the 100,000 students enrolled in a distance learning program at a well known university. The number S of students who had been diagnosed with chronic fatigue was recorded and used in the analysis. The distribution of S is _____?

(d.) A metallurgist designed a study to estimate the distribution of cracks in the cooling pipes at nuclear power plants. She randomly selected 200 sections of pipes within the cooling systems at various nuclear power plants. An x-ray is taken of each pipe and the number of cracks, C, of length greater than 20mm is recorded. A possible probability model for C, the number of cracks of length greater than 20mm in a randomly selected pipe, is _____?

(e.) The automobile industry has thousands of small suppliers of parts for its assembly process. In a randomly selected delivery of parts to the assembly plant, let p be the proportion of parts in the shipment that are not within specification. From previous studies, it is known that around 80% of the suppliers produce parts with values of p in the 0 to 0.03 range but the remaining suppliers have values of p between 0.03 and 0.25. A possible probability model for p is _____?

(f.) For each day during a six month period in Stamford, Connecticut, the maximum daily ozone reading R was recorded. The distribution of R is _____?

(g.) A mechanical engineer for a natural gas distributor is investigating the occurrence of leaks in gas pipelines. From 30 years of data she finds that the number of major cracks in any 100 feet of pipe appears to be independent of the number of major cracks in any adjacent 100 feet of pipe. From the 30 year data set, she determines that the average number of major cracks per 100 feet of pipe is relatively constant. Let **C** be the number of major cracks in a randomly selected 100 feet section of pipe. The distribution of the C is _____?

(h.) A biostatistician wishes to investigate the distribution of defective genes occurring in the kidney of mice exposed to a toxin. She uses 100 mice in a lab study and determines for each mouse the number of defective genes in their kidney. A possible probability model for G, the number of defective genes in the kidney of a mouse exposed to the toxin, is _____?

(i.) A geologist is studying the frequency of occurrence of minor earthquakes in Texas. From hundreds of years of data she finds that the number of earthquakes in any 24 hour period is independent of the number of earthquakes in any other 24 period of time. Over the past 100 years, the daily earthquake rate has been relatively constant. Let **Q** be the number of days in which there were no earthquakes in a randomly selected year. The distribution of the Q is _____?

(j.) In an epidemiological study of the incidence of skin cancer for those individuals who use artificial tanning procedures, a researcher needs at least 100 subjects in the study for it to have validity. The researcher interviews subjects to determine if they fit the criteria to be included in the study. Let S be the number of subjects interviewed until 100 subjects are determined to be acceptable for the study. The distribution of S is _____?

(k.) A quality control engineer wants to document major problems in a process which produces ball bearings. He measures the difference D between the nominal diameter of a 5 cm ball bearing and the true bearing diameter. He finds that the bearings are equally likely to have a diameter larger than or smaller than 5 cm. Furthermore, approximately 20% of the bearings have diameters which deviate more than 3 standard deviations from 5 cm. The distribution of D is _____?

(l.) The wings on an airplane are subject to stresses which cause cracks in the surface of the wing. After 1000 hours of flight the wing is inspected with an x-ray machine and the number of cracks N are recorded. The distribution of N is _____?

(m.) Suppose small aircraft arrive at a certain airport according to a Poisson process with rate 8 aircraft per hour. For the next 100 days, the length of time, T, until the 15th aircraft arrives each day is recorded. The distribution of T is _____?

(n.) An entomologist is studying the number of ticks on cattle in a feed lot. From 20 years of tick inspection data, she finds that the number of ticks on a randomly selected cow appears to be independent of the number of ticks on any other randomly selected cow in the feed lot. The density of ticks per cow appears to be fairly constant over the past 20 years. Let **T** be the number of cows in a random selection of 100 cows in the feed lot having fewer than 5 ticks on their torso. The distribution of T is _____?

(o.) An administrator is studying the quality of high-school curriculums in Michigan. She randomly selects 50 high schools out of the 357 high schools in Michigan for the study. A careful examination of their curriculum is performed. Let X be the number of high schools in which the curriculum was found to be unsatisfactory. The distribution of X is ____?

Stat 641 Fall 2021

Solutions for Assignment 2

(1.) (10 points) In the following expressions let m be a non-negative integer. Using the expression $\sum_{k=0}^{m} ab^k = a\frac{1-b^{m+1}}{1-b}$, we have with a = p, b = 1- p, [y] = greatest integer $\leq$ y

(a.) For $y < 0$, F(y) = 0; for $y \geq 0$,

$$F(y) = P[Y \leq y] = \sum_{k=0}^{[y]} p(1-p)^k = 1-(1-p)^{[y]+1} = \begin{cases} 0 & \text{if } y < 0 \\ p & \text{for } 0 \leq y < 1 \\ 1-(1-p)^2 & \text{for } 1 \leq y < 2 \\ 1-(1-p)^3 & \text{for } 2 \leq y < 3 \\ 1-(1-p)^4 & \text{for } 3 \leq y < 4 \\ \vdots & \end{cases}$$

(b.)

Using the definition of $Q(u)$, $Q(u) = inf(y : F(y) \geq u) \Rightarrow$

$$Q(u) = \text{ smallest nonnegative integer } y_u \text{ such that } 1-(1-p)^{y_u+1} \geq u \text{ with } Q(0) = 0$$

$$Q(u) = \text{ smallest nonnegative integer } y_u \text{ such that } y_u \geq \frac{log(1-u)}{log(1-p)} - 1 \text{ with } Q(0) = 0$$

That is,

$$Q(u) = \begin{cases} 0 & \text{if } u \leq 0 \\ 0 & \text{for } 0 < u \leq p \\ 1 & \text{for } p < u \leq 1-(1-p)^2 \\ 2 & \text{for } 1-(1-p)^2 < u \leq 1-(1-p)^3 \\ 3 & \text{for } 1-(1-p)^3 < u \leq 1-(1-p)^4 \\ \vdots & \end{cases}$$

Using the above expression with p = .4, Q(.8) = 3

(2.) (20 Points) (a.) Let $W = \frac{Y-\theta}{\alpha}$ then the pdf of W is

$$f_W(w) = \alpha f(\theta+\alpha w) = \alpha \frac{\gamma}{\alpha^\gamma}((\theta+\alpha w)-\theta)^{\gamma-1} e^{-\left(\frac{(\theta+\alpha w)-\theta}{\alpha}\right)^\gamma} \text{ for } \theta+\alpha w \geq \theta \;\Rightarrow$$

$$f_W(w) = \begin{cases} \gamma w^{\gamma-1} e^{-w^\gamma} & \text{for } w \geq 0 \\ 0 & \text{for } w < 0 \end{cases}$$

Because the expression for $f_W(w)$ does not contain $(\theta,\ \alpha)$, we can conclude that $(\theta,\ \alpha)$ are location-scale parameters for the family of distributions.

(b.) To find the quantile function, set

$$u = F(y_u) = 1 - e^{-\left(\frac{y_u-\theta}{\alpha}\right)^\gamma}$$

and solve for y_u. In this case,

$$y_u = \theta + \alpha(-log(1-u))^{1/\gamma} \quad \Rightarrow \quad Q(u) = \theta + \alpha(-log(1-u))^{1/\gamma}$$

(c.) With $\theta = 5,\ \gamma = 4,\ \alpha = 20,\quad P(Y > 23) = 1 - P(Y \leq 23) = 1 - F(23) = e^{-\left(\frac{23-5}{20}\right)^4} = .5189$

(d.) With $\theta = 5,\ \gamma = 4,\ \alpha = 20,\quad Q(.25) = 5 + 20(-log(1-.25))^{1/4} = 19.647$

(3.) (10 points) (a.) Let $W = Y/\beta$. The pdf of W is

$$f_W(w) = \beta f(\beta w) = \beta\frac{\gamma}{\beta}(\beta w)^{\gamma-1} e^{-(\beta w)^\gamma/\beta} = \gamma\beta^{\gamma-1} w^{\gamma-1} e^{-\beta^{\gamma-1} w^\gamma} \quad \text{for} \quad w > 0$$

Because the expression for the pdf of W contains β, β cannot be a scale parameter for the given family of distributions.

(b.) With $W = Y/\alpha$, the pdf of W is given by

$$f_W(w) = \alpha f(\alpha w) = \alpha\frac{\gamma}{\alpha}(\alpha w/\alpha)^{\gamma-1} e^{-(\alpha w/\alpha)^\gamma} = \gamma w^{\gamma-1} e^{-w^\gamma} \quad \text{for} \quad w > 0$$

The expression for f_W is free of α, therefore α is a scale parameter.

(4.) (10 points)

(a.) Let X be the number of emissions in a week. X has a Poisson distribution with $\lambda = 0.15$.

$$P(X \geq 2) = 1 - P(X=0) - P(X=1) = 1 - \frac{e^{-0.15}(0.15)^0}{0!} - \frac{e^{-0.15}(0.15)^1}{1!} = 1 - .8607 - .1291 = 0.0102$$

Using the R-function dpois, $P(X \geq 2) = 1 - dpois(0, .15) - dpois(1, .15) = 1 - .8607 - .1291 = 0.0102$

(b) Let Y be the number of emissions in a year. Y has a Poisson distribution with $\lambda = 0.15 \times 52 = 7.8$.

$$P(Y \geq 2) = 1 - P(Y=0) - P(Y=1) = 1 - \frac{e^{-7.8}(7.8)^0}{0!} - \frac{e^{-7.8}(7.8)^1}{1!} = 1 - .0004097 - .003200 = .9964$$

Using the R-function dpois, $P(X \geq 2) = 1 - dpois(0, 7.8) - dpois(1, 7.8) = 1 - .0004097 - .003200 = .9964$ or

Using the R-function ppois, $P(X \geq 2) = 1 - P(X \leq 1) = 1 - ppois(1, 7.8) = 1 - .0036 = .9964$

(5.) (10 points)

(a.) A has a t-distribution with df = 3 (A is the ratio of a N(0,1) r.v. and the square root of a Chi-square r.v. divided by its df. with the numerator and denominator r.v's having independent distributions)

(b.) B has a Cauchy distribution with location = 0 and scale =1 (B is the ratio of two independent N(0,1) r.v.'s)

(c.) C has a chi-squared distribution with df = 3 (C is the sum of independent squared N(0,1) r.v.s)

(d.) D has an F-distribution with $df_1 = 4,\ df_2 = 3$ (An F-distribution is the ratio of two independent Chi-square r.v.'s divided by their df's.)

(e.) E has an F-distribution with $df_1 = 1,\ df_2 = 3$ (An F-distribution is the ratio of two independent Chi-square r.v.'s divided by their df's.)

(6.) (10 points) Let U = .26 be a realization from a Uniform on (0,1) distribution.

(a.) W = Weibull(γ=4,α=1.5): $Q(u) = 1.5[-log(1-u)]^{1/4} \ \Rightarrow$

$W = Q(.26) = 1.5[-log(1-.26)]^{1/4} = 1.111$

(b.) $N = NegBin(r = 8, p = 0.7)$. Recall that the R functions for Negative Binomial are modeling the number of failures. Using the R function **pnbinom(x,8,.7)** with

x=c(0,1,2,3), we obtain the cdf, F(x), for X equal to the number failures before the 8th success:

$$F(x) = \begin{cases} 0.05764801 & x = 0 \\ 0.19600323 & x = 1 \\ 0.38278279 & x = 2 \\ 0.56956234 & x = 3 \end{cases}$$

Thus, with U=.26, we obtain $X = 2$ because $F(1) = .196 < .26 < .383 = F(2)$.

Therefore, N, the number of trials before the 8th success, $N = X + 8 = 2 + 8 = 10$.

(c.) B = Bin(20,.4): Using the R function **pbinom(x,20,.4)** with

x=c(5,6,7), we obtain

$$F(x) = \begin{cases} .1256 & x = 5 \\ .2500 & x = 6 \\ .4159 & x = 7 \end{cases}$$

Thus, with U=.26, we obtain $B = 7$ because $F(6) = .25 < .26 < .4159 = F(7)$

(d.) P = Poisson(λ=3): Using the R function **ppois(x,3)** with x=c(0,1,2), we obtain

$$F(x) = \begin{cases} .04978707 & x = 0 \\ .19914827 & x = 1 \\ .42319008 & x = 2 \end{cases}$$

Thus, with U=.26, we obtain $P = 2$ because $F(1) = .1991 < .26 < .4232 = F(2)$.

(e.) Y = Uniform on (0.3,2.5). Then the pdf is $f(y) = 1/(2.5 - .3)$ for $.3 < y < 2.5$; 0 otherwise. Therefore, the cdf is given by

$$F(y) = 0 \text{ for } y \leq .3;\ F(y) = 1 \text{ for } y \geq 2.5;\ \text{For } .3 < y < 2.5,\ F(y) = \int_{.3}^{y} \frac{1}{2.5 - .3} dy = \frac{1}{2.5 - .3}(y - .3)$$

Let $u = F(y_u) = \frac{1}{2.5-.3}(y_u - .3)$, then solve for y_u yields $Q(u) = y_u = .3 + (2.5 - .3)u$.

Therefore, with $U = .26$, $Y = .3 + (2.5 - .3)(.26) = 0.872$

(7.) (30 points)

(a.) Hypergeometric - Sampling from finite population in which there are two types of units

(b.) Exponential - Distance between events in a Poisson process

(c.) Binomial (assuming 100,000 is very large) or Hypergeometric - Sampling from finite population in which there are two types of units

(d.) Poisson - counting the number of cracks larger than 20mm in a randomly selected pipe

(e.) Beta- Values of p are within (0,1) and have a skewed distribution

(f.) Weibull - Modeling extremes, maximum daily ozone level

(g.) Poisson - number of events occurring in a fixed length of pipe

(h.) Poisson - number of events occurring in a large number of genes

(i.) Binomial - number of occurrences of independent events in a fixed number of trials

(j.) Negative Binomial - number of trials until success, 100 subjects are accepted into the study

(k.) Cauchy - symmetric distribution with a large number of extreme values

(l.) Poisson - recording number of events in space - cracks on wing

(m.) Gamma - Time until 15th event in a Poisson process

(n.) Binomial - Counting number of trials (cows) in which a success occurs (5 or fewer ticks)

(o.) Hypergeometric - Sampling from finite population in which there are two types of units