STATISTICS 641 - ASSIGNMENT 2

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

Name Jack Rodoni

Email Address jrodoni @tamu.edu

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.

Jack Rodoni
STAT 641
Assignment 2
9/20/21

• Assignment covers H.O. 3; Chp 2,3,4 in the Devore's book.

1.) Assume that the R.V. Y has a PMF w/ parameter p, $0<p<1$:

$$f(y) = \begin{cases} p(1-p)^y & \text{for } y = 0,1,2,3,\dots \\ 0 & \text{o.w.} \end{cases}$$

(a.) Find the cdf $F(y)$ for Y: $\boxed{\text{Hint: } \sum_{k=0}^{m} ab^k = a\frac{1-b^{m+1}}{1-b}}$

• $F(y) = \sum_{y=0}^{m} p(1-p)^y = p\left(\frac{1-(1-p)^{y+1}}{1-(1-p)}\right) = 1-(1-p)^{y+1}$

$$\boxed{F(y) = \begin{cases} 1-(1-p)^{y+1} & \text{for } y = 0,1,2,\dots \\ 0 & \text{o.w.} \end{cases}}$$

NOTE: $F(y) = 1-(1-p)^{y+1}$ is a valid cdf as:

① $\forall y \in [0,1,2,\dots]$ $F(y) \geq 0$.

Proof:

• $F'(y) = -\ln(1-p)\, p^{y+1}$

since $0<p<1 \Rightarrow -\ln(1-p) > 0 \ \forall p \in (0,1)$

similarly $p^{y+1} > 0 \ \forall p \in (0,1)$. $\therefore$ Thus $F(y)$ is monotonically increasing.

• Since $F(y)$ is monotonically increasing, its lowest value on the interval $y \in [0,1,\dots]$ occurs when $y=0$.

$F(0) = 1-(1-p)$.

$\lim_{p\to 0} F(0) = \lim_{p \to 0} 1-(1-p) = 0$.

Thus $F(y) \geq 0$. QED.

②. $\forall y \in [0,1,2,\dots]$, $F(y) \leq 1$.

Proof: From ① we know y is monotonically increasing. Thus $F(y)$ will be at its largest value as $y \to \infty$.

• $\lim_{y\to\infty} F(y) = \lim_{y\to\infty} 1-(1-p)^{y+1} = 1 - \lim_{y\to\infty}(1-p)^{y+1}$

• since $(1-p)<1$; $\lim_{y\to\infty} F(y) = 1$.

Thus $F(y) \leq 1$ QED

③ $F'(y) \geq 0$ $\forall y \in [0,1,2,\dots]$.

From ①, we know $F'(y) > 0$ $\therefore$ Thus $F(y)$ is nondecreasing.

1.) (Contd)

(b) Find the 80^{th} percentile of $F(y)$ if $p = 0.4$. That is, evaluate $Q(0.8)$ for $p = 0.4$.

① Find $Q(y)$.

$$y = 1 - (1-p)^{Q(y)+1} \quad * \; p = 0.4 \Rightarrow y = 1 - (0.6)^{Q(y)+1}$$

$$\Leftrightarrow 0.6^{Q(y)+1} = 1 - y \Leftrightarrow (Q(y)+1)\ln(0.6) = \ln(1-y)$$

$$\boxed{Q(y) = \frac{\ln(1-y)}{\ln(0.6)} - 1}$$

$$Q(0.8) = (\ln(0.2)/\ln(0.6)) - 1 \approx 2.15$$

$$\boxed{Q(0.8) = 3}$$

2.) Let Y have a 3-parameter Weibull dist, that is, Y has pdf & cdf $\alpha > 0, \gamma$

$$f(y) = \begin{cases} \gamma/\alpha^\gamma (y-\theta)^{\gamma-1} e^{-(y-\theta/\alpha)^\gamma} & \text{for } y \geq \theta \\ 0 & \text{for } y < \theta \end{cases}$$

$$F(y) = \begin{cases} 1 - e^{-(y-\theta/\alpha)^\gamma} & \text{for } y \geq \theta \\ 0 & \text{for } y < \theta \end{cases}$$

a.) Verify the pair (θ, α) are location-scale parameters for this family of dist.

Let $w = Y - \theta/\alpha \Rightarrow$ pdf w is $f_w(\omega) = \alpha f(\alpha w + \theta)$

$$f_w(\omega) = \alpha\left(\frac{\gamma}{\alpha^\gamma}((\alpha w + \theta) - \theta)^{\gamma-1} e^{-\left(\frac{(\alpha w + \theta) - \theta}{\alpha}\right)^\gamma}\right.$$

$$= \alpha^{1-\gamma} \gamma \alpha^{\gamma-1} w^{\gamma-1} e^{-w^\gamma}$$

$$\boxed{f_w(w) = \gamma w^{\gamma-1} e^{-(w)^\gamma}}$$

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

$$Q(x) = F^{-1}(x)$$

$$\cdot\; x = 1 - e^{-\left(\frac{Q(x)-\theta}{\alpha}\right)^\gamma} \Leftrightarrow e^{-\left(\frac{Q(x)-\theta}{\alpha}\right)^\gamma} = 1 - x$$

$$\cdot\; -\left(\frac{Q(x)-\theta}{\alpha}\right)^\gamma = \ln(1-x) \Leftrightarrow -\left(\frac{Q(x)-\theta}{\alpha}\right) = (\ln(1-x))^{1/\gamma}$$

$$\boxed{Q(x) = \alpha(-\ln(1-x))^{1/\gamma} + \theta}$$

2.) (Cont'd)

(c) What is the probability that a random selected value from a Weibull dist w/ $\theta = 5$, $\gamma = 4$; $\alpha = 20$ has a value greater than 23.

$$P(Y > 23) = 1 - F(23) = 1 - \left(1 - e^{-((23-5)/20)^4}\right) = \boxed{0.51884}$$

(d) $Q(0.25) = 20(-\ln(1-0.25))^{1/4} + 5 = \boxed{19.647}$

3.) An alternative form of the 2 parameter weibull distribution is given as follows w/ parameters $\beta > 0, \gamma > 0$

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

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

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

[NOTE: θ in a family of pdfs for the R.V. Y $\{f_y(y;\theta) : \theta \in \Theta\}$ is said to be a scale parameter if the dist $W = Y/\theta$ doesn't depend on θ. That is if the pdf of W $f_w(w) = \theta f_y(\theta w)$ does not depend on θ.]

$$f_w(w) = \beta f_y(\beta w) = \beta\left(\frac{\gamma}{\beta}(\beta w)^{\gamma-1} e^{-(\beta w)^\gamma/\beta}\right)$$

$$= \underbrace{(\beta w)^{\gamma-1} e^{-(\beta^{\gamma-1} w^\gamma)}}$$

doesn't reduce any further $\Rightarrow f_w(w)$ is dependent on β

$\Rightarrow \beta$ is not a scale parameter.

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

$$f_w(w) = \beta^{1/\gamma} f(\beta^{1/\gamma} w) = \beta^{1/\gamma}\left(\frac{\gamma}{\beta}(\beta^{1/\gamma} w)^{\gamma-1} e^{-(\beta^{1/\gamma} w)^\gamma/\beta}\right)$$

$$= \beta^{1/\gamma}\left(\frac{\gamma}{\beta}\left(\beta^{\frac{\gamma-1}{\gamma}} w^{\gamma-1}\right)\right) e^{-w^\gamma} = \beta^{1/\gamma}\left(\frac{\gamma}{\beta}\left(\beta^{1-1/\gamma} w^{\gamma-1}\right) e^{-w^\gamma}\right)$$

$$= \beta^{1/\gamma}\left(\gamma\left(\beta^{-1/\gamma} w^{\gamma-1}\right) e^{-w^\gamma}\right)$$

$$\boxed{f_w(w) = \gamma w^{\gamma-1} e^{-w^\gamma}}$$

4.) An experiment measures the number of particle emissions from a radioactive substance. The number of emissions has a poisson distribution w/ rate $\lambda = 0.15$ particles

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

$$P(y \geq 2) = 1 - P(y=1) - P(y=0) = 1 - \left(\frac{e^{-0.15}\, 0.15^1}{1!}\right) - \left(\frac{e^{-0.15}\, 0.15^0}{0!}\right)$$

$$\boxed{P(y \geq 2) = 0.089}$$

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

$$\lambda_{52} = 0.15(52) = 7.8$$

$$P(y \geq 2) = 1 - p(y=1) - p(y=0) = 1 - \left(\frac{e^{-7.8}(7.8)^1}{1!}\right) - \left(\frac{e^{-7.8}\, 7.8^0}{0!}\right)$$

$$\boxed{P(y \geq 2) = 0.99676}$$

5.) Let $Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, Z_7, Z_8$ be independent $N(0,1)$ R.Vs. Identify the distributions of the random variables A, B, C, D by providing the name of the distribution and the appropriate degrees of freedom

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

$\boxed{\text{t-dist w/ 3 df}}$

(b) $B = Z_5 / Z_6$

$\boxed{\text{Cauchy}}$

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

$\boxed{\text{Chi-squared w/ 3df}}$

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

$\boxed{\text{Fisher w/ dF (4,3)}}$

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

$\boxed{\text{Fisher w/ dF (1,3)}}$

6.) Let $U = .26$ be a realization from a uniform $(0,1)$ distribution. Express a single realization from each of the following distributions using just the fact that $U = 0.26$.

(a) Weibull ($\gamma = 4$, $\alpha = 1.5$)

$$F(y) = 1 - e^{-(y/\alpha)^\gamma} \Rightarrow y = 1 - e^{-(Q(y)/\cancel{\alpha})^\gamma} \Rightarrow e^{-(Q(y)/\alpha)^\gamma} = 1 - y$$

$$\Rightarrow Q(y) = \alpha(-\ln(1-y))^{1/\gamma}, \quad Q(0.26) = 1.5(-\ln(0.74))^{1/4} = \boxed{1.11145}$$

6.) (Cont'd)

(b) N = negbinom(r=8, p=0.7)

Rcode:

```
sum = 0
q = 0.26
p_given = 0.7
r_given = 8
i = r_given

while (sum < q) {
    sum = sum + choose(i-1, r_given-1)*(p_given^r_given)*((1-p_given)^(i-r_given))
    if (sum < q) {
        i = i+1
    } else {
        i = i
    }
}
```

$\boxed{Q(0.26) = 10}$

(c) B = Bin(20, 0.4)

rcode: $\boxed{\text{qbinom}(0.26, 20, 0.4) = 7}$

(d) P = Poisson ($\lambda = 3$)

rcode: $\boxed{\text{qpois}(0.26, 3) = 2}$

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

rcode: $\boxed{\text{qunif}(0.26, \text{min} = 0.3, \text{max} = 2.5) = 0.872}$

7.) (a) $\boxed{N \sim \text{Hypergeometric}}$ We are sampling from a population, presumably, w/o replacement and each outcome is binary; either the person has cancer or they do not. Though, b/c the ratio of the sample size to population size is so small, this could also be realistically modeled w/ a Binomial distribution.

(b.) $\boxed{D \sim \text{Exponential}}$ From Course Notes: Exponential (B) R.V.

- In a poisson process w/ λ (in our case $\lambda = 30$) being the average number of occurrences (in our case micro-cracks) in a unit of [space] (in our case 10ft length of column), let T (D) be the time (distance) btwn the occurence of two events.

7.) (cont'd)

(c) $\boxed{S \sim \text{Hypergeom}}$ We are sampling from a population, presumably, w/o replacement and each outcome is binary; either the student has chronic fatigue, or they do not. However, b/c the ratio of sample size to population size is so small this situation could also be realistically modeled by a Binomial distribution.

(d) $\boxed{C \sim \text{Poisson}}$ C is a discrete R.V. and we are measuring the # of occurrences of an 'event' (cracks in cooling pipes) during a specified period of space (1 pipe length).

(e) $\boxed{p \sim \text{Beta}}$ We are modeling a probability and Beta distributions are useful when trying to model probabilities.

(f) $\boxed{R \sim \text{Normal}}$ R is a continuous R.V. Presumably the distribution of R would be symmetric given the information we have as there is no reason the max daily ozone reading would be skewed one way or another. Similarly, there is nothing in the information given to indicate the tails of the distribution should be heavier than normal.

(g) $\boxed{C \sim \text{Poisson}}$ same argument as (d) above.

(h) $\boxed{G \sim \text{Binomial}}$ Discrete R.V., the population size (the # of genes in a mouse) is effectively infinite.

(i) $\boxed{Q \sim \text{Exponential}}$ similar argument as (b) above.

(j) $\boxed{S \sim \text{Neg Binomial}}$ We are counting the # of attempts until r successes.

7.) (cont'd)

(k) $D \sim t$ D should by symmetric, but w/ heavier tails than the normal dist would have, as we are told 20% of the bearings have diameters which deviate more than 3 SD from the mean.

(l) $L \sim$ Poisson Similar argument to (c) above.

(m) $T \sim$ Gamma The length of time btwn each plane arriving is exponential. Thus $T = \sum_{i=1}^{15} E_i$ where $E_1 \ldots E_{15}$ are i.i.d R.V. Thus $T \sim$ Gamma.

(n) $T \sim$ Hypergeom Similar to c & a, but population size is unknown. If we assume population size to be very large, then the distribution of T could also be modeled by the binomial distribution.

(o) $X \sim$ HyperGeom Similar to a & c. Binomial not realistic model b/c ratio of sample size to population size is relatively large.