Homework 9, DATA 556: Due Tuesday, 11/27/2018

Alexander Van Roijen

November 29, 2018

Please complete the following:

1. Let X be a continuous random variable with PDF f(x) and CDF F(x). For a fixed number x_0 such that $F(x_0) < 1$, define the function

$$g(x) = \begin{cases} \frac{f(x)}{1-F(x_0)} & x \geq x_0 \\ 0, & x < x_0 \end{cases} \tag{1}$$

Prove the g(x) is a pdf

$$g(x) \geq 0 \forall x \in X \tag{2}$$

Trivial, as $1 - F(x_0) > 0 \, \& \, f(x) > 0 \forall x \in X \Rightarrow g(x) = \frac{f(x)}{1-F(x_0)} \geq 0 \text{ or } g(x) = 0 \tag{3}$

$$\int_{-\infty}^{\infty} g(x)dx = 1 \tag{4}$$

$$\int_{x_0}^{\infty} \frac{f(x)}{1-F(x_0)}dx = \frac{1}{1-F(x_0)} \int_{x_0}^{\infty} f(x)dx = \frac{1}{1-F(x_0)} \int_{-\infty}^{\infty} f(x)dx - \int_{-\infty}^{x_0} f(x)dx \tag{5}$$

$$= \frac{1}{1-F(x_0)}(1 - F(x_0)) = 1 \square \tag{6}$$

2. Let X be a random variable with PDF $f(x) = \frac{1}{2}(1+x)$ for $-1 < x < 1$

Note that $0 < x^2 < 1$

(a) Find the pdf of $Y = X^2$

$$f_y(y) = \frac{d}{dy}P(Y \leq y) = \frac{d}{dy}P(X^2 \leq y) = \frac{d}{dy}P(-\sqrt{y} \leq X \leq \sqrt{y}) \tag{7}$$

$$= \frac{d}{dy}\int_{-1}^{\sqrt{y}} f(x)dx - \int_{-1}^{-\sqrt{y}} f(x)dx = \frac{d}{dy}\frac{x}{2}|_{-1}^{\sqrt{y}} + \frac{x^2}{4}|_{-1}^{\sqrt{y}} - (\frac{x}{2}|_{-1}^{-\sqrt{y}} + \frac{x^2}{4}|_{-1}^{-\sqrt{y}}) \tag{8}$$

$$= \frac{d}{dy}\frac{\sqrt{y}}{2} + \frac{1}{2} + \frac{y}{4} - \frac{1}{4} + \frac{\sqrt{y}}{2} - \frac{1}{2} - \frac{y}{4} + \frac{1}{4} = \frac{d}{dy}\sqrt{y} = \frac{1}{2\sqrt{y}} 0 < y < 1, 0 \text{ otherwise} \tag{9}$$

(b) Find the mean $E[Y]$ and the variance $\text{Var}(Y)$ of the RV Y

$$E[Y] = E[g(x)] \text{ with } g(x) = x^2 => \int_{-1}^{1} x^2 \frac{1}{2}(1+x)dx \tag{10}$$

$$= \int_{-1}^{1} \frac{x^2}{2}dx + \int_{-1}^{1} \frac{x^3}{2}dx \tag{11}$$

$$= \frac{x^3}{6}|_{-1}^{1} + \frac{x^4}{8}|_{-1}^{1} = \frac{1}{6} + \frac{1}{6} + 0 = \frac{1}{3} \tag{12}$$

$$E[Y^2] = E[g(x)] \text{ with } g(x) = x^4 => \int_{-1}^{1} x^4 \frac{1}{2}(1+x)dx \tag{13}$$

$$= \int_{-1}^{1} \frac{x^4}{2}dx + \int_{-1}^{1} \frac{x^5}{2}dx \tag{14}$$

$$= \frac{x^5}{10}|_{-1}^{1} + \frac{x^6}{12}|_{-1}^{1} = \frac{1}{10} + \frac{1}{10} + 0 = \frac{1}{5} \tag{15}$$

$$\text{Var}(Y) = E[Y^2] - E[Y]^2 = \frac{1}{5} - \frac{1}{9} = \frac{4}{45} \tag{16}$$

3. Let X be a random variable with moment generating function $M_X(t)$, for $-h < t < h$. Prove that

(a) $P(X \geq a) \leq e^{-at} M_X(t)$ for $0 < t < h$

Let us define a function $g(t) = e^{t(X-a)} - I_{X \geq a}$. We have two cases

$$X \geq a => g(t) \geq 0 \text{ as } X - a \geq 0 => e^{t(X-a)} \geq 1 = I_{X \geq a} \tag{17}$$

$$X < a => g(t) \geq 0 \text{ as } X - a < 0 => e^{t(X-a)} \geq 0 = I_{X \geq a} \tag{18}$$

If we take the expectation of $g(t)$ we get the following inequality

$$E[e^{t(X-a)}] \geq [I_{X \geq a}] = e^{-at} M_X(t) \geq P(X \geq a) \tag{19}$$

The above is true due to the fundamental bridge and the fact we have shown in both cases that $e^{t(X-a)} \geq I_{X \geq a}$. Note that $0 < t < h$ is important as it determines that the sign of our exponential follows the sign of the difference of $X - a$. If this werent the case, our inequality would need to flip as we will see below.

(b) $P(X \leq a) \leq e^{-at} M_X(t)$ for $-h < t < 0$ Let us define a function $g(t) = e^{t(X-a)} - I_{X \leq a}$.

3

We have two cases

$$X \leq a \implies g(t) \geq 0 \text{ as } t * (X - a) \leq 0 \implies e^{t(X-a)} \geq 1 = I_{X \leq a} \qquad (20)$$

Note that the negative value of t made $X - a$ a positive exponentent, thus greater than 1

$$(21)$$

$$\text{The same occurs below in the second case} \qquad (22)$$

$$X > a \implies g(t) \geq 0 \text{ as } t * (X - a) > 0 \implies e^{t(X-a)} \geq 0 = I_{X \leq a} \qquad (23)$$

$$(24)$$

If we take the expectation of $g(t)$ we get the following inequality

$$E[e^{t(X-a)}] \geq [I_{X \leq a}] = e^{-at} M_X(t) \geq P(X \leq a) \qquad (25)$$

The above is true due to the same reasons as part a, except that the negative sign of our exponent will flip the inequality in the exponent allowing for all this to be true in the first place.

4. Let $X \sim N(\mu, \sigma^2)$. Find the values of μ, σ s.t.

$$P(|X| < 2) = P(-2 < X < 2) = P(X < 2) - P(X < -2) = \frac{1}{2} \qquad (26)$$

$$\text{Now let} \frac{X - \mu}{\sigma} \sim N(0, 1) \qquad (27)$$

$$\implies P(\frac{X - \mu}{\sigma} < \frac{2 - \mu}{\sigma}) - P(\frac{X - \mu}{\sigma} < \frac{-2 - \mu}{\sigma}) = \frac{1}{2} \qquad (28)$$

$$\implies \phi(\frac{2 - \mu}{\sigma}) - \phi(\frac{-2 - \mu}{\sigma}) = \phi(\frac{2 - \mu}{\sigma}) - (1 - \phi(\frac{2 - \mu}{\sigma})) = 2\phi(\frac{2 - \mu}{\sigma}) - 1 = \frac{1}{2} \qquad (29)$$

$$\text{We also have} P(|X| > 2) = P(X < -2) + P(X > 2) = \frac{1}{2} \qquad (30)$$

$$\implies P(\frac{X - \mu}{\sigma} < \frac{-2 - \mu}{\sigma}) + P(\frac{X - \mu}{\sigma} > \frac{2 - \mu}{\sigma}) = P(\frac{X - \mu}{\sigma} < \frac{-2 - \mu}{\sigma}) + 1 - P(\frac{X - \mu}{\sigma} < \frac{2 - \mu}{\sigma}) = \frac{1}{2}$$

$$(31)$$

$$= 1 + \phi(\frac{-2 - \mu}{\sigma}) - \phi(\frac{2 - \mu}{\sigma}) = 1 + 1 - \phi(\frac{2 - \mu}{\sigma}) - \phi(\frac{2 - \mu}{\sigma}) = \frac{1}{2} \qquad (32)$$

$$= 2 - 2\phi(\frac{2 - \mu}{\sigma}) = \frac{1}{2} \qquad (33)$$

$$\text{simplifying, we get} \phi(\frac{2 - \mu}{\sigma}) = \frac{3}{4} \qquad (34)$$

$$\implies \frac{2 - \mu}{\sigma} = \phi^{-1}(\frac{3}{4}) \implies 2 - \sigma\phi^{-1}(\frac{3}{4}) = \mu \square \qquad (35)$$

4

It is easy to see that since the mean and variance are not limited in any particular way, there are many solutions to this inequality and do exist. Thus we are done.

5. Find $P(X > \sqrt{Y})$ if X and Y are jointly distributed with

$$f_{X,Y}(x,y) = x + y \text{ for } 0 \leq x \leq 1 \& 0 \leq y \leq 1 \tag{36}$$

$$P(X > \sqrt{Y}) = \int_0^1 \int_{\sqrt{y}}^1 x + y \, dx \, dy = \int_0^1 (\frac{x^2}{2}|_{\sqrt{y}}^1 + xy|_{\sqrt{y}}^1) dy = \int_0^1 (\frac{1}{2} - \frac{y}{2} + y - y^{3/2}) dy$$

$$\tag{37}$$

$$= \int_0^1 (\frac{1}{2} + \frac{y}{2} - y^{3/2}) dy = \frac{1}{2} + \frac{1}{4} - \frac{2}{5} = \frac{7}{20} \tag{38}$$

6. Suppose X and Y are independent N(0,1) random variables.

   (a) Find $P(X^2 < 1)$

   We know that the chi-squared dsitribution is a sum of squared normal distributions. We can consider the above as a sum of 1 squared normal distribution, and thus can equivalently find $P(Z < 1)$ with $Z \sim \chi_1^2$. The one indicates the one degree of freedom for our distribution.

   We get

   ```
   > print(pchisq(1,1))
   [1] 0.6826895
   ```

   $P(X^2 < 1) = .68269$

   (b) Find $P(X^2 + Y^2) < 1$ Similar to the one above, we have a chisquared distribution ,but now up our degrees of freedom to 2 $Z \sim \chi_2^2$.

   We get the following.

   ```
   > print(pchisq(1,2))
   [1] 0.3934693
   ```

   $P(X^2 + Y^2) < 1 = .3934693$

7. Let $X \sim N(\mu, \sigma^2)$, and let $Y \sim N(\tau, \sigma^2)$. Suppose X and Y are independent. Define two random variables U = X+Y and V = X-Y.

(a) Show that U and V are independent Normal random variables.

Intuitively, if we can show this follows a bivariate normal distribution with no covariance between them we have satisfied this question. This property is describe in one of our earlier lectures.

Note, we already know and have shown in previous assignments and lectures that a sum or difference of independent normal distributions are $N(\mu_1 + \mu_2, 2\sigma^2)$ and $N(\mu_1 - \mu_2, 2\sigma^2)$ respectively

$$U, V = X + Y, X - Y \sim \begin{bmatrix} E[U] \\ E[V] \end{bmatrix}, \begin{bmatrix} \text{Var}[U] & Cov(U,V) \\ Cov(V,U) & \text{Var}[V] \end{bmatrix} \tag{39}$$

We do not need to even bother ourselves with determining the variance of the random variables U and V, even though this would be rather easy. We only need to show that Cov(U,V) = 0

$$Cov(U,V) = E(UV) - E(U)E(V) = E[X^2 - Y^2] - E[X+Y]E[X-Y] \tag{40}$$

$$= E[X^2] - E[Y^2] - (E[X] + E[Y])(E[X] - E[Y]) \text{ by linearity of expecation} \tag{41}$$

$$= \sigma^2 + \mu^2 - (\sigma^2 + \tau^2) - (\mu + \tau)(\mu - \tau) = \mu^2 - \tau^2 - \mu^2 + \mu\tau - \mu\tau + \tau^2 = 0 \tag{42}$$

Thus, we have shown U,V are uncorrelated and have zero covariance. Consequently they are independent Normals $\square$

(b) Find the distribution of U.

We have already stated the distribution of U and V above, but we will do a quick proof here to show the result.

$$P(U \leq u) = P(X + Y \leq u) = \int_{-\infty}^{\infty} P(X + y \leq u | Y = y) f_y(y) dy \tag{43}$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{2\pi\sigma^2}} e^{-1 \frac{(u-y-\mu)^2 + (y-\tau)^2}{2\sigma^2}} dy \tag{44}$$

The remainder of this proof is manipulating the completed squares to achieve the following

$$\tag{45}$$

$$P(U \leq u) = \frac{1}{\sqrt{4\pi}\sigma} e^{-1 \frac{(u-(\tau+\mu))^2}{\sigma^2}} \sim N(\mu + \tau, 2\sigma^2) \tag{46}$$

The integral evaluates to 1 as it is a kernel of a normal distrib $\tag{47}$

6

(c) Find the distribution of V.

The proof will be very similar to the one above and has been done in homework 7.

The result is below.

$$P(U \leq u) = P(X - Y \leq u) = \int_{-\infty}^{\infty} P(X - y \leq u | Y = y) f_y(y) dy \qquad (48)$$

$$P(U \leq u) = \frac{1}{\sqrt{4\pi}\sigma} e^{-1 \frac{(u-(\tau-\mu))^2}{\sigma^2}} \sim N(\mu - \tau, 2\sigma^2) \qquad (49)$$

$$\qquad (50)$$

8. A random variable X is defined by the transformation $Z = \log(X)$, where the mean of the random variable Z is $E(Z) = 0$. Is $E(X)$ greater than, less than, or equal to 1?

$$E[Z] = E[\log(X)] = E[g(x)] = \int_{-\infty}^{\infty} \log(x) f(x) dx \qquad (51)$$

$$\text{We are interested in } E[X] = \int_{-\infty}^{\infty} x f(x) dx \qquad (52)$$

Since f(x) stays the same, we really want to analyze the difference between $\log(x)$ and x. So lets try to find the minimum distance of the following

$$y = x - \log(x) \qquad (53)$$

$$\frac{dy}{dx} = 1 - \frac{1}{x} = 0 \implies \frac{1}{x} = 1 \implies x = 1 \qquad (54)$$

$$y(1) = 1 - 0 = 1 \qquad (55)$$

$$\frac{d^2y}{dx^2} = \frac{1}{x^2} \geq 0 \qquad (56)$$

This implies that at our value of x=1 we are at a local minimum, which in this case is the global minimum, and all other values of x result in a large difference. Thus this means

$$E[X] \geq 1 \qquad (57)$$

This should make sense, as if this were to be a discrete RV, which happens to take only probability at X=1, or Z =0, then $E[X] = 1$ and $E[Z] = 0$. If we have either a discrete or continuous RV with values with probability outside of X=1 and Z=0 s.t. E[Z]=0 still holds true, then we know that the distance between $x - \log(x) > 1$ which would make $E[X] > 1$. Thus we are done □ Thu

Have a great Holiday Season!