XX50215 Statistics for Data Science

Problems 3 - Solutions

1. Given X has the geometric pmf $f_X(x) = \frac{1}{3}\left(\frac{2}{3}\right)^x$, x = 0, 1, 2, … Determine the probability distribution of Y = X/(X+1). Note that here both X and Y are discrete random variables. To specify the probability distribution of Y, specify its pmf.

P(Y = y) = P($\frac{X}{X+1}$ = y) = P(X = $\frac{y}{1-y}$) = $\frac{1}{3}\left(\frac{2}{3}\right)^{y/(1-y)}$, where y = 0, 1/2, 2/3, 3/4,….,x/x+1,…

2. Show that the following function is a cdf and find $F_x^{-1}(y)$.

$$F_x(x) = \begin{cases} 0 & if\ x < 0 \\ 1 - e^{-x} & if\ x \geq 0 \end{cases}$$

i. $\lim_{x\to 0} F(x) = 1 - e^{-0} = 0,\ \lim_{x\to\infty} F(x) = 1 - e^{-\infty} = 1$

ii. 1 - e^{-x} is increasing in x.

iii. 1 - e^{-x} is continuous.

iv. $F_x^{-1}(y) = -\log(1-y)$

3. Consider a sequence of independent coin flips where the probability of each being heads is p.

    a. Define a random variable X as the length of the run of either heads or tails started by the first trial. (e.g. X = 4 if the sequence is HHHHT or TTTTH.)

    b. Find the distribution of X and EX.

$$P(X = k) = (1-p)^k p + p^k(1-p), k = 1,2, ...$$

$$\begin{aligned} \mathrm{E}X &= \sum_{k=1}^{\infty} k[(1-p)^k p + p^k(1-p)] = (1-p)p\left[\sum_{k=1}^{\infty} k(1-p)^{k-1} + \sum_{k=1}^{\infty} kp^{k-1}\right] \\ &= (1-p)p\left[\frac{1}{p^2} + \frac{1}{(1-p)^2}\right] = \frac{1-2p+2p^2}{p(1-p)}. \end{aligned}$$

4. If a couple decides to continue having children until a daughter is born. What is the expected number of children of the couple?

If X =number of children until the first daughter then $P(X = k) = (1-p)^{k-1}p$ where p = probability of a daughter. X is a geometric random varianble.dd

$$EX = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = p - \sum_{k=1}^{\infty} \frac{d}{dp}(1-p)^k = -p\frac{d}{dp}\left[\sum_{k=0}^{\infty}(1-p)^k - 1\right] = -p\frac{d}{dp}\left[\frac{1}{1p} - 1\right]$$
$$= \frac{1}{p}$$

Therefore, if p = 1/2, the expected number of children is two.

5. Find the moment generating function (mgf) corresponding to f(x) = 1/c, 0 < x < c.

$$E(e^{tX}) = \int_0^c e^{tx}\frac{1}{c}dx = \frac{1}{ct}e^{tx}|_0^c = \frac{1}{ct}e^{tc} - \frac{1}{ct}1 = \frac{1}{ct}(e^{tc} - 1)$$

6. Does a distribution exist for $M_X(t) = t/(1-t)$, $|t| < 1$?

If yes, find it. If no, prove it.

No, evaluates to 0 at t=0. Needs to evaluate to 1 to be an mgf.

7. A median of a distribution is a value m such that P(X <= m) >= 1/2 and P(X >= m) >= 1/2.

If X is continuous, m satisfies

$$\int_{-\infty}^m f(x)dx = \int_m^{\infty} f(x)dx = 1/2.$$

Find the median of the distribution f(x) = $3x^2$, 0 <x<1.

$$\int_0^m 3x^2dx = m^3 \overset{set}{\longrightarrow} \frac{1}{2} \Rightarrow m = (\frac{1}{2})^{1/3} = 0.794$$

8. Suppose the pdf $f_X(x)$ of a random variable X is an even function.

Show that:

a. X and –X are identically distributed.

b. $M_X(t)$ is symmetric about 0.

Note: fx(x) is an even function of $f_X(x) = f_X(-x)$ for every x.

a. $Y = -X$ and $g^{-1}(y) = -y$. Thus $f_Y(y) = f_X(g^{-1}(y))|\frac{d}{dy}g^{-1}(y)| = f_X(-y)|-1| = f_X(y)$ for every y.

b. To show that $M_X(t)$ is symmetric about 0 we must show that $M_X(0+\epsilon) = M_X(0-\epsilon)$ for all $\epsilon > 0$.

$$\begin{aligned}
M_X(0+\epsilon) &= \int_{-\infty}^{\infty} e^{(0+\epsilon)x} f_X(x)dx = \int_{-\infty}^{0} e^{\epsilon x} f_X(x)dx + \int_{0}^{\infty} e^{\epsilon x} f_X(x)dx \\
&= \int_{0}^{\infty} e^{\epsilon(-x)} f_X(-x)dx + \int_{-\infty}^{0} e^{\epsilon(-x)} f_X(-x)dx = \int_{-\infty}^{\infty} e^{-\epsilon x} f_X(x)dx \\
&= \int_{-\infty}^{\infty} e^{(0-\epsilon)x} f_X(x)dx = M_X(0-\epsilon).
\end{aligned}$$