Statistical and Mathematical Methods DS-GA 1002
, 2019

Recitation Session 5 Solutions

Problem

1. *Minimum of Random Variables* Given a collection of independent exponentially distributed random variables $\{X_1; X_2; ...; X_n\}$ with parameter β. Let $Y = min\{X_1; X_2; ...; X_n\}$. Find the distribution of Y .

   SOLUTION

   In general we have
   $$F_Y(y) = 1 - [1 - F_X(y)]^n$$

2. *Independent Random Variables.* If the joint density function of X and Y is
   $$f(x, y) = 6e^{-2x}e^{-3y}, \quad 0 < x < \infty,\ 0 < y < \infty$$
   and is equal to 0 outside this region, are the random variables independent? What if the joint density function is
   $$f(x, y) = 24xy, \quad 0 < x < 1,\ 0 < y < 1,\ 0 < x + y < 1$$

   SOLUTION

   (1) Yes, notice $f(x, y) = 6e^{-2x}e^{-3y} = 2e^{-2x} * 3e^{-3y}$
   (2) No, since $f(y) = \int_0^{1-y} 24xydx = 12y(1-y)^2$

3. *Conditional Independence.* There is a disease which affects 2% of the population. A medical test is available which gives a false positive rate of 5% and a mis-detection rate of 3%. Compute the false discovery rate and the false omission rate of using this test.

   SOLUTION

   Define the random variables H and T as in the lecture notes ($H = 0$ if the patient is healthy, $H = 1$ if affected, $T = 0$ if the test is negative, $T = 1$ if positive). $\mathrm{P}(H = 1) = 0.02$. Mis-detection rate is $\mathrm{P}(T = 0|H = 1) = 0.03$. False positive rate is $\mathrm{P}(T = 1|H = 0) = 0.05$. To compute the false discovery rate, by Bayes rule,
   $$\mathrm{P}(H = 0|T = 1) = \frac{\mathrm{P}(H = 0)\mathrm{P}(T = 1|H = 0)}{\mathrm{P}(H = 0)\mathrm{P}(T = 1|H = 0) + \mathrm{P}(H = 1)\mathrm{P}(T = 1|H = 1)}$$
   $$= \frac{(1 - 0.02)0.05}{(1 - 0.02)0.05 + 0.02(1 - 0.03)} \approx 0.716.$$

To compute the false omission rate, by Bayes rule,

$$\mathrm{P}(H=1|T=0) = \frac{\mathrm{P}(H=1)\mathrm{P}(T=0|H=1)}{\mathrm{P}(H=1)\mathrm{P}(T=0|H=1)+\mathrm{P}(H=0)\mathrm{P}(T=0|H=0)}$$
$$= \frac{0.02\cdot 0.03}{0.02\cdot 0.03+(1-0.02)(1-0.05)} \approx 0.000644.$$

4. *Generating Random Variables* Generate random numbers: suppose you are able to generate uniform random variables over $\{0,1\}$. How to generate a sequence of independent uniform random variables over $\{0,1,\cdots,12\}$? Propose your algorithm and briefly explain your answer (You don't have to justify it rigorously).

   Solution

   The idea is to encode $\{0,1,\ \cdots,\ 12\}$ with $\{0,1\}$, i.e., binary encoding. For example, we can let

   $$0 = \text{“0000”},\ 1 = \text{“0001”},\ \cdots 11 = \text{“1011”},\ 12 = \text{“1100”}.$$

   Then we can get a random number between $\{0,1,\ \cdot,\ 12\}$ by first generating a number of uniformly distributed random variables between $\{0,\ \cdots,\ 15\}$ and then only accept those between 0 and 12. It means you waste some randomness by rejecting the number between 13 and 15.

Extra Problem

1. *Transformations* Let (X,Y) be uniformly distributed on the unit disk $\{(x,y): x^2+y^2\le 1\}$, i.e.,

   $$f(x,y) = \begin{cases} \frac{1}{\pi}, & \text{if } x^2+y^2\le 1, \\ 0, & \text{otherwise.} \end{cases}$$

   Let $R=\sqrt{X^2+Y^2}$. Find the cdf and pdf of R.

   Solution

   Note that $0\le R\le 1$ since $(X,\ Y)$ is uniformly distributed on the unit circle. By definition of cdf, we have

   $$\mathbb{P}(R\le t) = \mathbb{P}(X^2+Y^2\le t^2) = \frac{1}{\pi}\int_{\{x^2+y^2\le t^2\}} \mathrm{d}x\mathrm{d}y = \pi t^2$$

   since it is basically the area of a circle with radius t. Therefore, the cdf of R is

   $$F_R(t) = \frac{1}{\pi}\cdot \pi t^2 = t^2,\ 0\le t\le 1$$

   and the pdf is

   $$f_R(t) = \begin{cases} 2t, & \text{if } 0\le t\le 1, \\ 0, & \text{otherwise.} \end{cases}$$

2. *Conditional Independence II* Monty Hall Problem: You are given three doors, behind one door is a car; behind the other two are goats. You pick one door and announce it. Once you've done that, Monty opens one of the other two doors that he knows has a goat behind it. Then he gives you the option to switch or not. Would you rather switch?

SOLUTION

Denote $p_{NS}()$ and $p_S()$ two strategies and I the information given by Monty. Then

$$p_{NS}(\text{win} \mid I) = p_{NS}(\text{win}) = 1/3$$

and

$$p_S(\text{lose} \mid I) = p_S(\text{lose}) = 1/3$$

so you should switch

3. *Generating Random Variables* We investigate how to generate discrete random variables given the pmf. For simplicity, we assume a random variable X has its pmf as:

$$\Pr(X = k) = p_k > 0, \qquad k = 0, 1, 2, \cdots,$$

and

$$\sum_{k=0}^{\infty} p_k = 1.$$

a. Write down the cdf of X by using the partial sum $S_n = \sum_{k=0}^{n} p_k$.

b. Consider the generalized inverse of the cdf.

$$F_X^{-1}(x) := \inf\{t : F_X(t) \geq x\}, \quad 0 < x < 1,$$

which is the smallest value t such that $F_X(t)$ is greater than or equal to x. Specify the following set:

$$A_n := \{x : F_X^{-1}(x) = n\}, \quad n \in \{0, 1, 2, \cdots\}.$$

Try to use the expression of the cdf obtained from (a). (You may use Geo(1/2) as a concrete example to help you think. Sketching the cdf is very helpful).

c. Suppose U is a Unif[0,1] random variable. What is $\Pr(U \in A_n)$?

d. Can you provide a random number generator for Poisson(λ)? You are encouraged to implement the algorithm but not required

SOLUTION

a.

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ S_{\lfloor x \rfloor}, & \text{if } x \geq 0 \end{cases}$$

where $\lfloor x \rfloor$ is the largest integer no greater than x.

b. So long as $S_{n-1} < x \leq S_n$, then we have

$$F_X(n) = S_n \geq x.$$

On the other hand,

$$F_X(n-1) = S_{n-1} < x.$$

Therefore, we have

$$F_X^{-1}(x) = F_X(n)$$

and

$$A_n = (S_{n-1},\ S_n],\ n \geq 1.$$

If $n = 0, A_0 = (0,\ S_0] = (0, p_0]$.

c. If U is a Unif$[0,1]$ random variable, the probability of U falling in A_n is equal to the size of A_n, which is $S_n - S_{n-1} = p_n$. Therefore,

$$\mathbb{P}(U \in A_n) = p_n,\ n \geq 0.$$

d. We can just let

$$p_k = \frac{\lambda^k e^{-\lambda}}{k!},\ S_n = \sum_{k=0}^{n} p_k,\ n \in \{0, 1, 2, 3,\ \cdots\}.$$

Sample a uniform random variable U and we take

$$X = \begin{cases} 0, & \text{if } 0 < U \leq p_0, \\ n, & \text{if } S_{n-1} < U \leq S_n,\ n \geq 1. \end{cases}$$

Then $X \sim \text{Poisson}(\lambda)$ according to our discussion above.

Lab 4

1). $X_1, X_2, \ldots, X_n$ iid.

$X_i \sim \exp(\beta) \Rightarrow$ pdf: $\lambda e^{-\lambda x}$
CDF: $1 - e^{-\lambda x}$

$Y = \min(X_1, X_2, \ldots, X_n)$.

Pdf: a relative likelihood that the value of the random var would equal that sample
$\Rightarrow$ the prob of the random var falling within a particular range of value
$\rightarrow$ as opposed to taking on any one value.

$P(Y \le y) = P(\min(X_1, X_2, \ldots, X_n) \le y)$.

$= 1 - P(Y \geq y)$

$= 1 - P(\min(X_1, \ldots X_n) \geq y)$

$= 1 - [P(X_1 \geq y) \cdot P(X_2 \geq y) \cdot \ldots \cdot P(X_n \geq y)]$

$= 1 - [P(X_1 \geq y) + P(X_2 \geq y) \ldots + P(X_n \geq y)$

$= 1 - P(X_1 \geq y)\, P(X_2 \geq y) \ldots$

$= 1 - [e^{-\beta y} e^{-\beta y} \ldots e^{-\beta y}]$

$= 1 - e^{-\beta(y + y + \ldots + y)}$

$= 1 - e^{-\beta n y}$ $\quad y > 0$

$P(X_1 \geq y)$
$= 1 - \underbrace{P(X_1 \le y)}$
$= 1 - [1 - e^{-\beta y}]$
$= e^{-\beta y}$

$\Rightarrow Y \sim \exp(\beta n)$

2) joint density func

a) $f_{X,Y}(x, y) = \begin{cases} 6e^{-2x} e^{-3y} & 0 < x < \infty \\ & 0 < y < \infty \\ 0 & \text{o.t.w.} \end{cases}$

X, Y independent ?

? X, Y are independent $\Leftrightarrow$. $F_{X,Y}(x,y) = F_X(x) F_Y(y)$.

$f_{X,Y}(x, y) = f_X(x) f_Y(y)$?

$$f_{X,Y}(x,y) = 6e^{-2x} e^{-3y}$$

$$= \underbrace{(2e^{-2x})}_{\text{PDF} \sim \exp(2)} \underbrace{(3e^{-3y})}_{\sim \exp(3)}$$

$$= f_X(x) \quad f_Y(y) \quad \Rightarrow \text{ independ.}$$

b) $f_{X,Y}(x,y) = 24xy \quad 0 < x < 1, \; 0 < y < 1, \; 0 < x+y < 1$

can't factorize $\Rightarrow$ not independent.

$I(x,y) = \begin{cases} 1 & 0 < x < 1, \; 0 < y < 1, \; 0 < x+y < 1. \\ 0 & \text{o.t.w.} \end{cases}$

3) Disease affects 2% of the pop.

FP : 5%; FN = miss detect° rate = 3%.

False discovery = ? $P(H=0 \mid T=1) = ?$

False omission rate ? $P(H=1 \mid T=0) = ?$

$H=1$: affected ; 0 : not affected.

$T=0$: neg , 1 : pos.

$P(H=1) = 0.02.$

$P(T=0 \mid H=1) = 0.03.$

$P(T=1 \mid H=0) = 0.05.$

$$P(H=0 \mid T=1) = \frac{P(H=0,\ T=1)}{P(T=1)}$$

$$= \frac{P(H=0)\ P(T=1 \mid H=0)}{P(T=1)}.$$

$$= \frac{(1-0.02)(0.05)}{P(H=0)\ P(T=1 \mid H=0) + P(H=1)\ P(T=1 \mid H=1)}$$

$$= \frac{0.98 \times 0.05}{(1-0.02)\ 0.05 + 0.02 \times (1-0.03)} = 0.716.$$

~~P(H=1| T=0)~~

$$P(H=1 \mid T=0) = \frac{P(H=1)\,P(T=0 \mid H=1)}{P(T=0)}$$

$$= \frac{0.02 \cdot 0.03}{P(H=1)\,P(T=0 \mid H=1) + P(H=0)\,P(T=0 \mid H=0)}.$$

$$= \frac{0.02 \cdot 0.03}{0.02 \times 0.03 + (1-0.02)(1-0.05)} \approx 0.000644$$

4)

$$f(x) = \begin{cases} 0 & \text{if } 0 \le x \le 1/13 \\ 1 & \text{-- } \cancel{1}/13 < x \le 2/13 \\ \vdots & \\ 12 & \text{if } \frac{12}{13} < x \le 1. \end{cases}$$

Uniform ; 13 values

$P(0) = \dots = P(12) = 1/13$

---

5) $x^2+y^2=1$ $\quad f(x,y) = \begin{cases} 1/\pi & x^2+y^2 \le 1 \\ 0 & \text{otw.} \end{cases}$ $\Rightarrow F_{(X,Y)} =$

$R = \sqrt{X^2+Y^2}$. $\quad F_R(r) = P(R \le r) = P(\sqrt{X^2+Y^2} \le r) = P(X^2+Y^2 \le r^2)$

$$= \frac{\pi r^2}{\pi} = r^2$$

End

$f_R(r) = 2r.$