Statistical and Mathematical Methods DS-GA 1002 September 16, 2019

Recitation Session 2

1. *Discrete Random Variables.* Suppose that 15 percent of the families in a certain community have no children, 20 percent have 1 child, 35 percent have 2 children, and 30 percent have 3. Suppose further that in each family each child is equally likely (independently) to be a boy or a girl. If a family is chosen at random from this community, then B, the number of boys, and G, the number of girls, Please write down the the joint probability mass functions and their corresponding marginal probability mass functions in this family.

Solution

i	j = 0	1	2	3	$f_B(i)$
0	.15	.10	.0875	.0375	.3750
1	.10	.175	.1125	0	.3875
2	.0875	.1125	0	0	.2000
3	.0375	0	0	0	.0375
$f_G(j)$	.3750	.3875	.2000	.0375	

2. *Continues Random Variables.* The joint density of X and Y is given by

$$f(x,y) = \begin{cases} e^{-(x+y)} & 0 < x < \infty,\ 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Find the density function of the random variable X/Y.

Solution

We start by computing the distribution function of X/Y. For $a > 0$,

$$\begin{aligned} F_{X/Y}(a) &= p\left\{\frac{X}{Y} \le a\right\} \\ &= \iint_{x/y \le a} e^{-(x+y)} dx dy \\ &= \int_0^\infty \int_0^{ay} e^{-(x+y)} dx dy \\ &= \int_o^\infty (1 - e^{-ay}) e^{-y} dy \\ &= \left\{ -e^{-y} + \frac{e^{-}(a+1)y}{a+1} \right\} |_0^\infty \\ &= 1 - \frac{1}{a+1} \end{aligned}$$

Differentiation shows that the density function of X/Y is given by $f_{X/Y}(a) = 1/(a+1)^2,\ 0 < a < \infty$.

3. *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

In the first instance, the joint density function factors, and thus the random variables, are independent (with one being exponential with rate 2 and the other exponential with rate 3). In the second instance, because the region in which the joint density is nonzero cannot be expressed in the form $x \in A$, $y \in B$, the joint density does not factor, so the random variables are not independent. This can be seen clearly by letting

$$I(x, y) = \begin{cases} 1 & 0 < x < 1,\ 0 < y < 1,\ 0 < x + y < 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

and writing $f(x, y) = 24xyI(x, y)$, which clearly does not factor into a part depending only on x and another depending only on y.

4. *Discrete uniform - binomial.* Suppose there are $N + 1$ boxes labeled by $b = 0, 1, 2, ..., N$. Box b contains b black and $N - b$ white balls. A box is picked uniformly at random, and then n balls are drawn at random with replacement from whatever box is picked (the same box for each of the n draws). Let S_n denote the total number of black balls that appear among the n balls drawn. Find the distribution of S_n.

Solution

Let P be the proportion of black balls in the box picked. Possible values for P are $\{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}, 1\}$. Once a box is picked and the value of P is known, then the distribution of S_n is binomial(n, p):

$$\mathrm{P}(S_n = k)|P = p) = \binom{n}{k} p^k (1 - p)^{n-k} \tag{2}$$

All $N + 1$ possible values for p are equally likely because the priors are uniform. Therefore one can average over those values (or use the definition of conditional probability) to get the unconditional density of of S_n:

$$\mathrm{P}(S_n = k) = \sum_p \binom{n}{k} p^k (1 - p)^{n-k} \frac{1}{N + 1} \tag{3}$$

$$= \binom{n}{k} \frac{1}{(N + 1)N^n} \sum_{b=0}^{N} b^k (N - b)^{n-k} \tag{4}$$

5. *Change of variables.* Let X be uniformly distributed over $[0,1]$, and $Y = -\lambda^{-1}\log(X)$ where λ is positive. Find the distribution of Y.

Solution

Main idea for such problems is that probabilities measure sizes of sets, hence if the set is unchanged then probabilities are also unchanged. In other words $\mathrm{P}(X \in E) = \mathrm{P}(f(X) \in f(E))$ note the slight abuse of notation here, what $X \in E$ means is $\{\omega \in \Omega | X(\omega) \in E\}$. Again with a slight abuse of notation we can write $\mathrm{P}(X \in dx) = \mathrm{P}(Y \in dy)$ which gives us: $f_Y(y) = f_X(x)/|dy/dx|$. Absolute values are there because P measures a quantity that is always positive.

$$\frac{dy}{dx} = \frac{1}{\lambda x} \tag{5}$$
$$f_Y(y) = \frac{1}{1/\lambda x} = \lambda x \tag{6}$$
$$= \lambda \exp(-\lambda y) \tag{7}$$
$$\tag{8}$$

It turns out that Y is an exponential distribution with parameter λ. This is a standard way to simulate exponential distribution using uniform distribution.

6. *Inverse transform: discrete RV (Pittman, Sec. 4.5, Ex. 6).* Figure 2 shows the cdf of a binomial (2, 0.5) random variable.

   a. How would you obtain this cdf as a function $g(U)$ from the uniform $(0,1)$ variable U?

   b. How would you simulate it using fair coin flips?

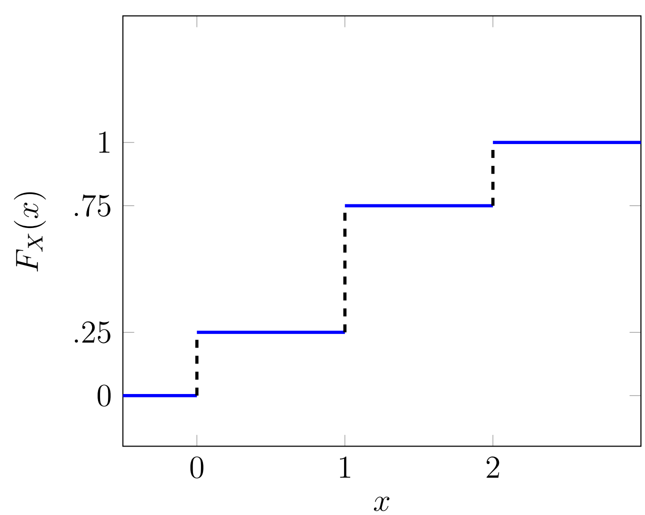

Figure 1: cdf of a binomial (2, 0.5) random variable.

Solution

a. Figure 2 below shows a function g from $(0,1)$ to $\{0,1,2\}$. The rule for getting from the uniform $(0,1)$ variable U to the binomial (2,0.5) variable $g(U)$ is

$$g(u) = \begin{cases} 0 & \text{if } 0 \leq u \leq 0.25 \\ 1 & \text{if } 0.25 < u \leq 0.75 \\ 2 & \text{if } 0.75 < u \leq 1 \end{cases}$$

This is because by construction the intervals on which g takes the values 0,1,2 have lengths $0.25, 0.5$, and 0.25 respectively, as required by the binomial $(2, 0.5)$ distribution.

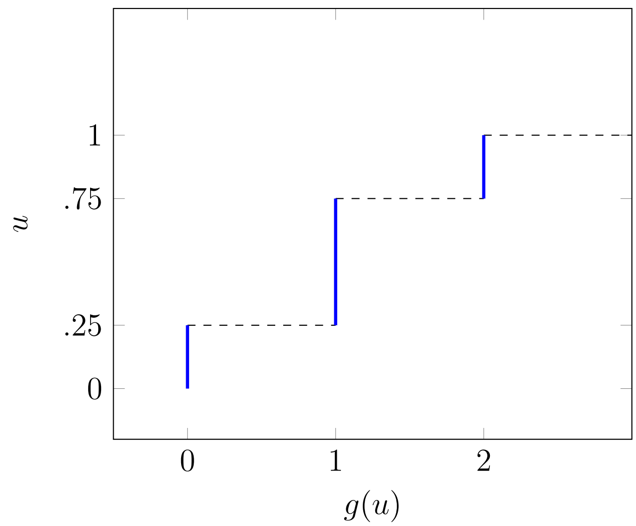

Figure 2: cdf of a binomial (2, 0.5) random variable.

b. To simulate it using fair coin flips, use

$$g(u) = \begin{cases} 0 & \text{if HH} \\ 1 & \text{if HT or TH} \\ 2 & \text{if TT} \end{cases}$$