MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.436J/15.085J — Fall 2018

Recitation 7

Exercise 1. Let X_1 and X_2 be independent random variables, uniform over the interval $(0, 1)$. Find the PDF of X_1X_2.

Solution:

1. **The Jacobian approach:** We wish to derive the PDF of $Y_1 = g(X_1, X_2) = X_1X_2$. Thus, we define $Y_2 = X_2$ and use find the Jacobian. From the relation $x_1 = y_1/x_2$ we see that $h(y_1, y_2) = y_1/y_2$. The partial derivative $\partial h/\partial y_1$ is $1/y_2$. We obtain

$$f_{Y_1}(y_1) = \int f_X(y_1/y_2, y_2)\frac{1}{y_2}\, dy_2 = \int f_X(y_1/x_2, x_2)\frac{1}{x_2}\, dx_2.$$

   Recall that $X_1, X_2 \stackrel{d}{=} U(0, 1)$, and independent. Then, their common PDF is $f_{X_i}(x_i) = 1$, for $x_i \in [0, 1]$. Note that $f_{Y_1}(y_1) = 0$ for $y \notin (0, 1)$. Furthermore, $f_{X_1}(y_1/x_2)$ is positive (and equal to 1) only in the range $x_2 \geq y_1$. Also $f_{X_2}(x_2)$ is positive, and equal to 1, iff $x_2 \in (0, 1)$. In particular,

$$f_X(y_1/x_2, x_2) = f_{X_1}(y_1/x_2)f_{X_2}(x_2) = 1, \qquad \text{for } x_2 \geq y_1.$$

   We then obtain

$$f_{Y_1}(y_1) = \int_{y_1}^{1} \frac{1}{x_2}\, dx_2 = -\log y_1, \qquad y_1 \in (0, 1).$$

2. **The direct approach:** The direct approach to this problem would first involve the calculation of $F_{Y_1}(y_1) = \mathbb{P}(X_1X_2 \leq y_1)$. It is actually easier to calculate

$$\begin{aligned} 1 - F_{Y_1}(y_1) = \mathbb{P}(X_1X_2 \geq y_1) &= \int_{y_1}^{1}\int_{y_1/x_1}^{1} dx_2\, dx_1 \\ &= \int_{y_1} \Big(1 - \frac{y_1}{x_1}\Big)\, dx_1 \\ &= (x_1 - y_1 \log x_1)\Big|_{y_1}^{1} = (1 - y_1) + y_1 \log y_1. \end{aligned}$$

   Thus, $F_{Y_1}(y_1) = y_1 - y_1 \log y_1$. Differentiating, we find that $f_{Y_1}(y_1) = -\log y_1$.

3. **The easiest approach:** An even easier solution for this particular problem (along the lines of the stick example in Lecture 9) is to realize that conditioned on $X_1 = x_1$, the random variable $Y_1 = X_1X_2$ is uniform on $[0, x_1]$, and using the total probability theorem,

$$f_{Y_1}(y_1) = \int_{y_1}^{1} f_{X_1}(x_1)f_{Y_1|X_1}(y_1 \mid x_1)\, dx_1 = \int_{y_1}^{1} \frac{1}{x_1}\, dx_1 = -\log y_1.$$

Exercise 2. Let $\{X_n\}$ be a sequence of i.i.d. random variables, $with X_1 \sim \exp(\lambda)$, and let $N \sim Geom(\beta)$ be an independent geometric random variable. Show that $T = X_1 + \cdots + X_N \sim \exp(\lambda\beta)$.

Solution: It is enough to show that its mgf is

$$E\left[e^{sT}\right] = \frac{\beta\lambda}{\beta\lambda - s}$$

Taking conditional expectation, we have

$$E\left[e^{sT}\right] = E\left[e^{s\sum_{n=1}^{N} X_n}\right] = E\left[E\left[e^{s\sum_{n=1}^{i} X_n}\middle| N = i\right]\right].$$

For a fixed i, we know that

$$E\left[e^{s\sum_{n=1}^{i} X_n}\middle| N = i\right] = \left(\frac{\lambda}{\lambda - s}\right)^i$$

Combining this with what we had before, we obtain

$$E\left[e^{sT}\right] = E\left[\left(\frac{\lambda}{\lambda - s}\right)^N\right] = \sum_{n=1}^{+\infty}\left(\frac{\lambda}{\lambda - s}\right)^n \beta(1-\beta)^{n-1} = \frac{\beta}{1-\beta}\sum_{n=1}^{+\infty}\left[\frac{\lambda(1-\beta)}{\lambda - s}\right]^n,$$

and thus

$$\frac{\beta}{1-\beta}\left[\frac{1}{1-\frac{\lambda(1-\beta)}{\lambda-s}} - 1\right] = \frac{\beta}{1-\beta}\left(\frac{\lambda - s}{\lambda - s - \lambda + \lambda\beta} - 1\right) = \frac{\beta}{1-\beta}\left(\frac{\lambda - s + s - \lambda\beta}{\lambda\beta - s}\right) = \frac{\lambda\beta}{\lambda\beta - s}.$$

Exercise 3. (Discrete-continuous Bayes rule) As part of a clinical trial, a patient undergoes either medical treatment A or medical treatment B. The treatment is chosen randomly, and each treatment has equal probability of being chosen. After the treatment, some health index X is observed for the patient. If treatment A is selected, the PDF of X is

$$f_{X|A}(x) = \begin{cases} 1 & \text{if } 0 < x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

If treatment B is selected, the PDF of X is

$$f_{X|B}(x) = \begin{cases} 3 & \text{if } 0 < x \leq 1/3, \\ 0 & \text{otherwise.} \end{cases}$$

If we are told that the value of X was less than $1/4$, what is the conditional probability that treatment A was the one selected?

Solution: We have

$$\begin{aligned}
\mathbb{P}(A \mid X < 1/4) &= \frac{\mathbb{P}(A)\mathbb{P}(X \leq 1/4 \mid A)}{\mathbb{P}(A)\mathbb{P}(X \leq 1/4 \mid A) + \mathbb{P}(B)\mathbb{P}(X \leq 1/4 \mid B)} \\
&= \frac{\mathbb{P}(A) \int_0^{1/4} f_{X|A}(x)\, dx}{\mathbb{P}(A) \int_0^{1/4} f_{X|A}(x)\, dx + \mathbb{P}(B) \int_0^{1/4} f_{X|B}(x)\, dx} \\
&= \frac{0.5 \int_0^{1/4} 1\, dx}{0.5 \int_0^{1/4} 1\, dx + 0.5 \int_0^{1/4} 3\, dx} \\
&= \frac{1}{4}.
\end{aligned}$$

Exercise 4. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. Bernoulli random variables (coin tosses), such that $\mathbb{P}(X_1 = H) = p \in (0,1)$. Let

$$L_n = \max\{m \geq 0 : X_n = H,\ X_{n+1} = H,\ \ldots,\ X_{n+m-1} = H,\ X_{n+m} = T\}$$

be the length of the run of heads starting from the n-th coin toss. Prove that

$$\limsup_{n\to\infty} \frac{L_n}{\log(n)} = \frac{1}{\log(1/p)} \qquad \text{a.s.}. \tag{1}$$

Solution: First, note that L_n has the same geometric distribution for all n, i.e., we have

$$\mathbb{P}(L_n = k) = (1-p)p^k, \qquad \forall\, k \geq 0,$$

for all n.

For any $\epsilon > 0$, we have

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{P}\left(\frac{L_n}{\log(n)} > \frac{1+\epsilon}{\log(1/p)}\right) &\leq \sum_{n=1}^{\infty} p^{(1+\epsilon)\frac{\log(n)}{\log(1/p)}} \\
&= \sum_{n=1}^{\infty} e^{-(1+\epsilon)\log(n)} \\
&= \sum_{n=1}^{\infty} n^{-(1+\epsilon)} \\
&< \infty.
\end{aligned}$$

Thus, Borel-Cantelli implies that

$$\mathbb{P}\left(\limsup_n \left\{\frac{L_n}{\log(n)} > \frac{1+\epsilon}{\log(1/p)}\right\}\right) = 0.$$

Since $L_n > (1+\epsilon)\frac{\log(n)}{\log(1/p)}$ only happens finitely many times, we also have

$$\mathbb{P}\left(\limsup_n \frac{L_n}{\log(n)} > \frac{1+\epsilon}{\log(1/p)}\right) = 0.$$

Since this holds for all $\epsilon > 0$, we must have

$$\mathbb{P}\left(\limsup_n \frac{L_n}{\log(n)} \leq \frac{1}{\log(1/p)}\right) = 1.$$

On the other hand, consider the sequence of events

$$A_n = \{X_{r_n} = H,\ \ldots,\ X_{r_n+d_n-1} = H\},$$

where $r_n = n^n$ and $d_n = \lfloor \log(n)/\log(1/p) \rfloor$. We have

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} p^{d_n} = \sum_{n=1}^{\infty} e^{d_n \log(p)} \geq \sum_{n=1}^{\infty} \frac{1}{n} = \infty.$$

Furthermore, note that the events A_n are independent. Thus, Borel-Cantelli implies that

$$\mathbb{P}(A_n\ i.o.) = 1.$$

This means that there are runs of at least $\lfloor \log(n)/\log(1/p) \rfloor$ heads infinitely often, and thus

$$\mathbb{P}\left(\limsup_n \frac{L_n}{\log(n)} \geq \frac{1}{\log(1/p)}\right) = 1.$$