Problem Set 3
Data Analysis and Interpretation (EE 223)

Instructor: Prof. Prasanna Chaporkar
EE Department, IIT Bombay

1. The joint density function of X and Y is given by

$$f(x,y) = \begin{cases} ce^{-x}e^{-2y} & 0 < x < \infty,\ 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Compute (a) c, (b) $P\{X > 1, Y < 1\}$, (c) $P\{X < Y\}$, and (d) $P\{X < a\}$.

(a) By definition of pdf

$$\begin{aligned} \int_0^\infty \int_0^\infty ce^{-x}e^{-2y} dx\ dy &= 1 \\ c\int_0^\infty e^{-x}dx \int_0^\infty e^{-2y}dy &= 1 \\ c \cdot 1 \cdot \frac{1}{2} &= 1 \\ c &= 2 \end{aligned}$$

(b)

$$\begin{aligned} P\{X > 1, Y < 1\} &= \int_0^1 \int_1^\infty 2e^{-x}e^{-2y} dx\ dy = -\int_0^1 2e^{-2y}(e^{-x}\big|_1^\infty)dy \\ &= e^{-1}\int_0^1 2e^{-2y}dy = e^{-1}(1-e^{-2}) \end{aligned}$$

(c)

$$\begin{aligned} P\{X < Y\} &= \iint\limits_{(x,y):x<y} 2e^{-x}e^{-2y} dx\ dy = \int_0^\infty \int_0^y 2e^{-x}e^{-2y} dx\ dy \\ &= \int_0^\infty 2e^{-2y}(1-e^{-y})dy = \int_0^\infty 2e^{-2y}dy - \int_0^\infty 2e^{-3y}dy \\ &= 1 - \frac{2}{3} = \frac{1}{3} \end{aligned}$$

(d)

$$P\{X < a\} = \int_0^a \int_0^\infty 2e^{-x}e^{-2y}dy\ dx = \int_0^a e^{-x}dx = 1 - e^{-a}$$

2. 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.

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

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

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

3. Sonia and Narendra decide to meet at a certain location. If each of them independently arrives at a time uniformly distributed between 5 PM and 6 PM, find the probability that the first to arrive has to wait longer than 10 minutes.

   If we let X and Y denote, respectively, the time past 5 PM that Sonia and Narendra arrive, then X and Y are independent random variables, each of which is uniformly distributed over $(0, 60)$. The desired probability, $P\{X + 10 < Y\} + P\{Y + 10 < X\}$, which, by symmetry, equals $2P\{X + 10 < Y\}$, is obtained as follows:

$$\begin{aligned} 2P\{X+10<Y\} &= 2\iint\limits_{x+10<y} f(x,y)dx\ dy = 2\iint\limits_{x+10<y} f(x)f(y)dx\ dy \\ &= 2\int_{10}^{60}\int_{0}^{y-10}\left(\frac{1}{60}\right)^2 dx\ dy = \frac{2}{(60)^2}\int_{10}^{60}(y-10)\ dy \\ &= \frac{25}{36} \end{aligned}$$

4. 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

$$f(x,y) = 24xy \quad 0 < x < 1,\ 0 < y < 1,\ 0 < x + y < 1$$

   and is equal to 0 otherwise?

   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

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

   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. Alternatively, this can be done by finding $f(x)$ and $f(y)$ and evaluating whether $f(x, y) = f(x)f(y)$ or not.

5. Nitish and Lalu shoot at a target. The distance of each shot from the center of the target is uniformly distributed on $(0, 1)$, independently of the other shot. What is the PDF of the distance of the losing shot from the center?

   Let X and Y be the distances from the center of the first and second shots, respectively. Let also Z be the distance of the losing shot:

$$Z = \max\{X, Y\}.$$

   We know that X and Y are uniformly distributed over $[0, 1]$, so that for all $z \in [0, 1]$, we have

$$P\{X \le z\} = P\{Y \le z\} = z.$$

   Thus, using the independence of X and Y, we have for all $z \in [0, 1]$,

$$\begin{aligned} F_Z(z) &= P\{\max\{X,Y\} \le z\} = P\{X \le z, Y \le z\} \\ &= P\{X \le z\}P\{Y \le z\} = z^2. \end{aligned}$$

   Differentiating, we obtain

$$f_Z(z) = \begin{cases} 2z & 0 \le z \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

6. Let X, Y, Z be independent and uniformly distributed over $(0, 1)$. Compute $P\{X \geq YZ\}$.

   Since $f(x, y, z) = f(x)f(y)f(z) = 1 \quad 0 \leq x, y, z \leq 1$, we have

$$
\begin{aligned}
P\{X \geq YZ\} &= \iiint_{x \geq yz} f(x, y, z) dx\ dy\ dz = \int_0^1 \int_0^1 \int_{yz}^1 dx\ dy\ dz \\
&= \int_0^1 \int_0^1 (1 - yz) dy\ dz = \int_0^1 \left(1 - \frac{z}{2}\right) dz \\
&= \frac{3}{4}.
\end{aligned}
$$

7. Sum of two independent random variables

   (a) If X and Y are independent random variables, both uniformly distributed on $(0, 1)$, calculate the probability density of $X + Y$.

   We use the result that the pdf f_{X+Y} is the convolution of the pdfs f_X and f_Y.

$$
f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a - y) f_Y(y) dy = \int_0^1 f_X(a - y) dy
$$

   For $0 \leq a \leq 1$, this yields

$$
f_{X+Y}(a) = \int_0^a dy = a
$$

   For $1 < a < 2$, this yields

$$
f_{X+Y}(a) = \int_{a-1}^1 dy = 2 - a.
$$

   Because of the shape of its density function, the random variable $X+Y$ is said to have a triangular distribution

   (b) If X and Y are independent Gamma random variables with respective parameters $(\alpha_1,\ \beta)$ and $(\alpha_2,\ \beta)$, then prove that $X + Y$ is also a Gamma random variable with parameters $(\alpha_1 + \alpha_2,\ \beta)$. Recall that a Gamma random variable has a density of the form

$$
f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha - 1} e^{-\beta x} \qquad x, \alpha, \beta > 0.
$$

   We use the result that the pdf f_{X+Y} is the convolution of the pdfs f_X and f_Y.

$$
\begin{aligned}
f_{X+Y}(a) &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \int_0^a \beta^{\alpha_1} (a - y)^{\alpha_1 - 1} e^{-\beta(a-y)} \beta^{\alpha_2} y^{\alpha_2 - 1} e^{-\beta y} dy \\
&= K e^{-\beta a} \int_0^a (a - y)^{\alpha_1 - 1} y^{\alpha_2 - 1} dy \\
&= K e^{-\beta a} a^{\alpha_1 + \alpha_2 - 1} \int_0^1 (1 - z)^{\alpha_1 - 1} z^{\alpha_2 - 1} dz \quad \text{by letting } x = \frac{y}{a} \\
&= C e^{-\beta a} a^{\alpha_1 + \alpha_2 - 1}
\end{aligned}
$$

   where C is a constant that does not depend on a. But, as the preceding is a density function and thus must integrate to 1, the value of C is determined, and we have

$$
f(x; \alpha_1 + \alpha_2, \beta) = \frac{\beta^{\alpha_1 + \alpha_2}}{\Gamma(\alpha_1 + \alpha_2)} x^{\alpha_1 + \alpha_2 - 1} e^{-\beta x}
$$

   Hence, the result is proved.

(c) If X and Y are independent Gaussian random variables with respective parameters $(\mu_1,\ \sigma_1^2)$ and $(\mu_2,\ \sigma_2^2)$, then prove that $X+Y$ is also a Gaussian random variable with parameters $(\mu_1+\mu_2,\ \sigma_1^2+\sigma_2^2)$. Recall that a Gaussian random variable has a density of the form

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2}}.$$

Let, $Z = X+Y$. Therefore, we have,

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx \\ &= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{\infty} e^{-\frac{(x-\mu_1)^2}{2{\sigma_1}^2}} e^{-\frac{(z-x-\mu_2)^2}{2{\sigma_2}^2}} dx \end{aligned}$$

how

Taking $a = x-\mu_1$ and $b = z-\mu_1-\mu_2$, expressing the argument of the exponent in terms of a, b we get

$$\begin{aligned} f_Z(z) &= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{b^2}{2({\sigma_1}^2+{\sigma_2}^2)}} \left(\frac{2\pi{\sigma_1}^2{\sigma_2}^2}{{\sigma_1}^2+{\sigma_2}^2}\right)^{1/2} \\ &= \frac{1}{\sqrt{2\pi({\sigma_1}^2+{\sigma_2}^2)}} e^{-\frac{(z-\mu_1-\mu_2)^2}{2({\sigma_1}^2+{\sigma_2}^2)}} \end{aligned}$$

This completes the proof.

8. Conditional distribution of random variable

(a) Suppose that the joint density of X and Y is given by

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

Find $P\{X>1|Y=y\}$.
We first obtain the conditional density of X given that $Y=y$.

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{X,Y}(x,y)}{f_Y(y)} \\ &= \frac{e^{-x/y}e^{-y}/y}{\int_0^{\infty} e^{-x/y}e^{-y}/y dx} = \frac{e^{-x/y}}{y} \end{aligned}$$

Hence

$$\begin{aligned} P\{X>1|Y=y\} &= \int_1^{\infty} \frac{e^{-x/y}}{y} dx = -e^{-x/y}\Big|_0^{\infty} \\ &= e^{-1/y}. \end{aligned}$$

(b) If X and Y are independent Poisson random variables with respective parameters λ_1 and λ_2. Prove that the conditional distribution of X given that $X+Y=n$ is a binomial distribution. A Poisson random variable has a pmf as

$$f(k;\lambda) = P\{X=k\} = \frac{\lambda^k e^{-\lambda}}{k!}.$$

We calculate the conditional probability mass function of X given that $X+Y=n$ as follows:

$$\begin{aligned} P\{X=k|X+Y=n\} &= \frac{P\{X=k, X+Y=n\}}{P\{X+Y=n\}} = \frac{P\{X=k, Y=n-k\}}{P\{X+Y=n\}} \\ &= \frac{P\{X=k\}P\{Y=n-k\}}{P\{X+Y=n\}} \end{aligned}$$

where the last equality follows from the assumed independence of X and Y. It can be easily proven that $X+Y$ has a Poisson distribution with parameter $\lambda_1+\lambda_2$, we see that the preceding equals

$$\begin{aligned}P\{X=k|X+Y=n\} &= \frac{\lambda_1^k e^{-\lambda_1}}{k!}\frac{\lambda_2^{(n-k)}e^{-\lambda_2}}{(n-k)!}\left[\frac{(\lambda_1+\lambda_2)^n e^{-(\lambda_1+\lambda_2)}}{n!}\right]^{-1} \\ &= \frac{n!}{(n-k)!k!}\frac{\lambda_1^k\lambda_2^{n-k}}{(\lambda_1+\lambda_2)^n} = \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}\end{aligned}$$

In other words, the conditional distribution of X given that $X+Y=n$ is the binomial distribution with parameters n and $\frac{\lambda_1}{\lambda_1+\lambda_2}$.

(c) Rajan goes to the bank to make a deposit, and is equally likely to find 0 or 1 customer ahead of him. The times of service of these customers are independent and exponentially distributed with parameter λ. What is the CDF of Rajan's waiting time? Recall that a Exponential random variable has a density of the form

$$f(x;\lambda)=\lambda e^{-\lambda x} \quad x\geq 0.$$

Let X be the waiting time and Y be the number of customers found. For $x<0$, we have $F_X(x)=0$. For $x\geq 0$,

$$F_X(x)=P\{X\leq x\}=\frac{1}{2}(P\{X\leq x \mid Y=0\}+P\{X\leq x \mid Y=1\})$$

We have

$$\begin{aligned}P\{X\leq x \mid Y=0\} &= 1 \\ P\{X\leq x \mid Y=1\} &= 1-e^{-\lambda x}\end{aligned}$$

Thus

$$F_X(x)=\frac{1}{2}(2-e^{-\lambda x}) \quad x\geq 0.$$

9. Joint probability distribution of functions of random variables

(a) X and Y have joint density function

$$f(x,y)=\frac{1}{x^2y^2} \quad x\geq 1\ y\geq 1$$

Compute the joint density function of $U=XY,\ V=X/Y$. What are the marginal densities?

If $u=xy, v=x/y$, then $J=\begin{vmatrix} y & x \\ \frac{1}{y} & \frac{-x}{y}\end{vmatrix}=\frac{-2x}{y}$, and $x=\sqrt{uv}, y=\sqrt{u/v}$. Hence

$$f_{U,V}(u,v)=\frac{1}{2v}f_{X,Y}(\sqrt{uv},\sqrt{u/v})=\frac{1}{2u^2v}, \quad u\geq 1,\ \frac{1}{u}<v<u$$

To evaluate marginal densities,

$$f_U(u)=\int_{1/u}^{u}\frac{1}{2u^2v}dv=\frac{1}{u^2}\log u, \quad u\geq 1.$$

For $v>1$

$$f_V(v)=\int_{v}^{\infty}\frac{1}{2u^2v}du=\frac{1}{2v^2}, \quad v>1$$

For $v<1$

$$f_V(v)=\int_{1/2}^{\infty}\frac{1}{2u^2v}du=\frac{1}{2}, \quad 0<v<1.$$

(b) Let X be exponentially distributed with mean 1. Once we observe the experimental value x of X, we generate a normal random variable Y with zero mean and variance $x+1$. What is the joint PDF of X and Y?

We have $f_X(x) = e^{-x}$, for $x \geq 0$, and

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi(x+1)}} e^{-y^2/2(x+1)}.$$

Thus

$$f_{X,Y}(x,y) = f_X(x) f_{Y|X}(y|x) = e^{-x} \frac{1}{\sqrt{2\pi(x+1)}} e^{-y^2/2(x+1)},$$

for all $x \geq 0$ and all y.