Probability for Statistics
Unseen Problem 4

1. Suppose that X and Y are absolutely continuous random variables with joint pdf given by

$$f_{X,Y}(x,y) = cx(1-y), \text{ for } 0 < x < 1 \text{ and } 0 < y < 1,$$

and zero otherwise, for some constant c.

   (a) Are X and Y independent random variables?

   (b) Find the value of c.

   (c) Find $\Pr(X < Y)$.

   (a) Write the joint support of the random vector (X, Y) *as* $(0,1)\times(0,1)$. *Then*

$$f_{X,Y}(x,y) = cx(1-y), \qquad 0 < x < 1, \quad 0 < y < 1,$$

   so that

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad \textit{with support equal to} \quad \mathbf{X}\times\mathbf{Y},$$

   where $\mathbf{X}$ *and* $\mathbf{Y}$ *are the supports of* X *and* Y *respectively, and*

$$f_X(x) = c_1 x \qquad \textit{and} \qquad f_Y(y) = c_2(1-y) \tag{1}$$

   for some constants satisfying $c_1c_2 = c$. *Hence, the two conditions for independence are satisfied in (1), and* X *and* Y *are independent.*

   (b) We must have

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}x\mathrm{d}y = 1 \qquad \textit{so} \qquad c^{-1} = \int_0^1\int_0^1 x(1-y)\,\mathrm{d}x\mathrm{d}y$$

   and as

$$\int_0^1\int_0^1 x(1-y)\,\mathrm{d}x\mathrm{d}y = \left\{\int_0^1 x\,\mathrm{d}x\right\}\left\{\int_0^1 (1-y)\,\mathrm{d}y\right\} = \frac{1}{2}\times\frac{1}{2} = \frac{1}{4},$$

   we have $c = 4$.

   (c) Set $A = \{(x,y) : 0 < x < y < 1\}$. *Because we must integrate only over the region where* $x < y$, *we first fix a* y *and integrate* $\mathrm{d}x$ *on the range* $(0, y)$, *and then integrate* $\mathrm{d}y$ *on the range* $(0,1)$, *that is*

$$\begin{aligned}\Pr(X<Y) &= \int\int_A f_{X,Y}(x,y)\,\mathrm{d}x\mathrm{d}y = \int_0^1\left\{\int_0^y 4x(1-y)\,\mathrm{d}x\right\}\mathrm{d}y = \int_0^1\left\{\int_0^y x\,\mathrm{d}x\right\}4(1-y)\,\mathrm{d}y \\ &= \int_0^1 2y^2(1-y)\,\mathrm{d}y = \left[\frac{2}{3}y^3 - \frac{1}{2}y^4\right]\Bigg|_0^1 = \frac{1}{6}.\end{aligned}$$

2. Let X be a 2×2 symmetric matrix with random entries. Suppose $X_{11}, X_{22} \sim N(0,1)$ and $X_{12} \sim N(0,\frac{1}{2})$, with all mutually independent. Let the eigenvalues of X be λ_1 and λ_2. Find the distribution of the eigenvalue spacing $|\lambda_1 - \lambda_2|$.

The eigenvalues satisfy the characteristic polynomial

$$\lambda^2 - (X_{11} + X_{22})\lambda + X_{11}X_{22} - X_{12}^2 = 0,$$

so that

$$\lambda = \frac{X_{11} + X_{22} \pm \sqrt{(X_{11} + X_{22})^2 - 4(X_{11}X_{22} - X_{12}^2)}}{2}.$$

This then gives

$$|\lambda_1 - \lambda_2| = \sqrt{(X_{11} + X_{22})^2 - 4(X_{11}X_{22} - X_{12}^2)} = \sqrt{(X_{11} - X_{22})^2 + 4X_{12}^2}$$

Now $X_{11} - X_{22} \sim N(0, 2)$ *so that* $Z_1 := \frac{1}{\sqrt{2}}(X_{11} - X_{22}) \sim N(0, 1)$ *and* $Z_1^2 \sim \chi_1^2$. *Similarly* $Z_2 = \sqrt{2}X_{12} \sim N(0, 1)$ *so* $Z_2^2 \sim \chi_1^2$.

Now

$$|\lambda_1 - \lambda_2| = \sqrt{2Z_1^2 + 2Z_2^2}$$

but $Z_1^2 + Z_2^2$ *is the sum of two independent* χ_1^2 *variables and so is* χ_2^2. *This is the same as a* $\Gamma(1, \frac{1}{2})$ *variable (using the rate parameterization of the gamma distribution), so that using the scale family property of the gamma distribution,* $2\left(Z_1^2 + Z_2^2\right) \sim \Gamma(1, \frac{1}{4})$. *This is just an exponential variable with* $\lambda = \frac{1}{4}$.

So the problem reduces to finding the density of $W = \sqrt{Y}$ *for* $Y \sim \text{EXPONENTIAL}(\lambda)$.

$$\Pr(W \leq w) = \Pr(Y \leq w^2) = 1 - \exp(-\lambda w^2), \qquad w > 0,$$

so

$$f_W w = \frac{d}{dw}\Pr(W \leq w) = 2\lambda w \exp(-\lambda w^2), \qquad w > 0.$$

Hence the probability density for the eigenvalue spacing is

$$f_W w = \frac{w}{2}\exp\left(-\frac{w^2}{4}\right), \qquad w > 0.$$