Probability for Statistics
Problem Sheet 3

The objective here is to practise working with joint distributions of random variables, and the quantities that can be derived from them. All questions should be accessible by the time you have watched all week 4 videos.

1. Let X be an absolutely continuous random variable with range $\mathbb{X} = \mathbb{R}^+$, pdf f_X and cdf F_X.

   (a) Show that
   $$\mathrm{E}(X) = \int_0^\infty [1 - F_X(x)]\ \mathrm{d}x.$$

   (b) Show also that for integer $r \geq 1$,
   $$\mathrm{E}(X^r) = \int_0^\infty r x^{r-1} [1 - F_X(x)]\ \mathrm{d}x.$$

   (c) Find a similar expression for random variables for which $\mathbb{X} = \mathbb{R}$.

2. Consider two absolutely continuous random variables X and Y such that
   $$\Pr(X \leq x \text{ and } Y \leq y) = \left(1 - e^{-x}\right)\left(\frac{1}{2} + \frac{1}{\pi}\tan^{-1} y\right), \text{ for } x > 0 \text{ and } -\infty < y < \infty,$$
   with
   $$\Pr(X \leq x \text{ and } Y \leq y) = 0, \text{ for } x \leq 0.$$
   Find the joint pdf, $f_{X,Y}$. Are X and Y independent ? Justify your answer.

3. Suppose that the joint pdf of X and Y is given by
   $$f_{X,Y}(x, y) = 24xy, \text{ for } x > 0,\ y > 0, \text{ and } x + y < 1,$$
   and zero otherwise. Find

   (a) the marginal pdf of X, f_X,

   (b) the marginal pdf of Y, f_Y,

   (c) the conditional pdf of X given $Y = y$, $f_{X|Y}$,

   (d) the conditional pdf of Y given $X = x$, $f_{Y|X}$,

   (e) the expected value of X,

   (f) the expected value of Y,

   (g) the conditional expected value of X given $Y = y$, and

   (h) the conditional expected value of Y given $X = x$.

   [Hint: Sketch the region on which the joint density is non-zero; remember that the integrand is only non-zero for some part of the integral range.]

4. (Harder) Suppose that X and Y have joint pdf that is constant on the range $\mathbb{X}^{(2)} \equiv (0,1) \times (0,1)$, and zero otherwise. Find the marginal pdf of the random variables $U = X/Y$ and $V = -\log(XY)$, stating clearly the range of the transformed random variable in each case.

   [Hint: For U, you might consider first the joint pdf of (U, X), then obtain the marginal pdf of U. For V, consider the joint pdf of $(V, -\log X)$, then obtain the marginal pdf of V. These choices result in much simpler calculations than those required to derive the joint transformation from (X, Y) to (U, V).]

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

$$f_{X,Y}(x,y) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left(x^2 + y^2\right)\right\}, \quad \text{for } x, y \in \mathbb{R}.$$

   (a) Let the random variable U be defined by $U = X/Y$. Find the pdf of U.

   (b) Suppose now that $S \sim \chi^2_\nu$ is independent of X and Y. (The pdf of S is given by

$$f_S(s) = c(\nu) s^{\nu/2-1} e^{-s/2}, \quad \text{for } s > 0,$$

   where ν is a positive integer and $c(\nu)$ is a normalizing constant depending on ν.) Find the pdf of random variable T defined by

$$T = \frac{X}{\sqrt{S/\nu}}.$$

   This is the pdf of a t random variable with ν degrees of freedom.

For discussion

6. Consider two independent random variables X_1 and X_2, exponentially distributed with rate 1. Suppose we wish to consider the density function of X_1 conditional on the event $\{X_1 = X_2\}$.

   (a) One way to do this is to consider the variable $Z = X_1 - X_2$, and condition on the event $Z = 0$. Find the pdf $f(x_1|z = 0)$.

   (b) Alternatively, one could consider the variable $W = \frac{X_2}{X_1}$, and condition on the event $W = 1$. Find the pdf $f(x_1|w = 1)$.

   (c) Comment on your answers to the two parts above. *(This is an instance of the Borel-Kolmogorov paradox.)*

7. Consider the data in Table 1, taken from Richard Doll's 1950s study of smoking. The table shows per capita consumption of cigarettes in 11 countries in 1930, and the death rates from lung cancer for men in 1950.

   (a) Produce a scatter plot of the data.

   (b) Why does the study compare cigarette consumption in 1930 with lung cancer rates 20 years later?

   (c) Why does the study only consider death rates in men?

   (d) Is it fair to conclude from these data that, on the whole, the higher the rate of smoking in a country in 1930, the higher the death rate from lung cancer in 1950?

   (e) Is it fair to conclude from these data that lung cancer death rates amongst smokers tend to be higher?

Table 1: Data on smoking and lung cancer rates

Country	Cigarette consumption	Deaths per million
Australia	480	180
Canada	500	150
Denmark	380	170
Finland	1100	350
Great Britain	1100	460
Iceland	230	60
Netherlands	490	240
Norway	250	90
Sweden	300	110
Switzerland	510	250
USA	1300	200