Statistics 351 (Fall 2015) September 14, 2015
Prof. Michael Kozdron

Lecture #3: Multivariate Random Variables

Let $(X, Y)'$ be a random vector. (With only two components it is traditional to use X and Y instead of X_1 and X_2.)

Sometimes we are interested in the distribution of just one component.

In the continuous case, we have

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,\mathrm{d}y$$

and

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,\mathrm{d}u = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{X,Y}(x, y)\,\mathrm{d}y.$$

We call $f_X(x)$ the *marginal density (function)* for X (or just *marginal* or just *density*) and $F_X(x)$ the *(marginal) distribution function* for X. Similar formulæ hold for $f_Y(y)$ and $F_Y(y)$.

Definition. The random variables X and Y are *independent* if and only if

$$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y).$$

i.e., iff $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$ in the continuous case.

Exercise. Write down the definition of the marginal mass function and the marginal distribution function in the discrete case.

Example (Chapter 1, Exercise 1.2)**.** Suppose that

$$(X, Y)' = \begin{pmatrix} X \\ Y \end{pmatrix}$$

denotes the coordinates of a dart thrown uniformly at random at a circular dart board. To be specific, suppose that the circle is centred at the origin and has radius 1. We can describe the random vector $(X, Y)'$ by its density function

$$f_{X,Y}(x, y) = \begin{cases} 1/\pi, & \text{if } x^2 + y^2 \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Suppose that $(X, Y)'$ is a point uniformly distributed in the unit disk so that

$$f_{X,Y}(x, y) = \begin{cases} 1/\pi, & \text{if } x^2 + y^2 \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Determine the distribution of X.

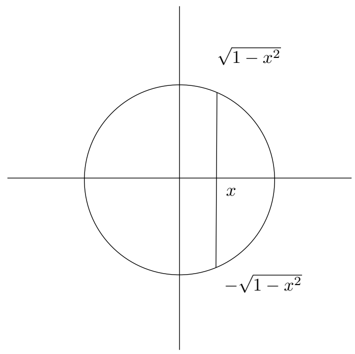

Figure 1: The region $\{x^2 + y^2 \le 1\}$.

Solution. By definition,

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,\mathrm{d}y \quad \text{easy step!} \\ &= \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}\,\mathrm{d}y \quad \text{watch limits!} \end{aligned}$$

That is, if x is fixed between -1 and 1, then y ranges between $-\sqrt{1-x^2}$ and $\sqrt{1-x^2}$. Therefore,

$$f_X(x) = \frac{2}{\pi}\sqrt{1-x^2}, \quad -1 \le x \le 1.$$

Similarly,

$$f_Y(y) = \frac{2}{\pi}\sqrt{1-y^2}, \quad -1 \le y \le 1.$$

We can now ask the following two questions.

- Are X and Y independent?
- Are X and Y uncorrelated?

Clearly, X and Y are NOT independent since $f_X(x) \cdot f_Y(y)$ does NOT equal $f_{X,Y}(x,y)$. It turns out, however, that X and Y are uncorrelated.

Recall. If X, Y are random variables, then

- $\operatorname{Cov}(X,Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y)$, and
- $\operatorname{Corr}(X,Y) = \rho_{X,Y} = \dfrac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}$.

Note that $\rho = \rho_{X,Y}$ is a scale-invariant real number with $-1 \le \rho \le 1$. Also note that in the continuous case,

$$\mathbb{E}(XY) = \iint_{\mathbb{R}^2} xy f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

In order to show that X and Y are uncorrelated, we need to show that $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - (\mathbb{E}X)(\mathbb{E}Y) = 0$. Since $f_{X,Y}(x,y) = 1/\pi$ for $x^2 + y^2 \leq 1$, we have

$$\mathbb{E}(XY) = \iint\limits_{\{x^2+y^2\leq 1\}} xy \cdot \frac{1}{\pi} \cdot \,\mathrm{d}x\,\mathrm{d}y.$$

To compute this double integral, use polar coordinates: $x = r\cos\theta$, $y = r\sin\theta$, $0 \leq r \leq 1$, $0 \leq \theta < 2\pi$, $\mathrm{d}x\,\mathrm{d}y = r\,\mathrm{d}r\,\mathrm{d}\theta$. That is,

$$\iint\limits_{\{x^2+y^2\leq 1\}} xy \cdot \frac{1}{\pi} \cdot \,\mathrm{d}x\,\mathrm{d}y = \frac{1}{\pi}\int_0^1 \int_0^{2\pi} r^3 \cos\theta \sin\theta \,\mathrm{d}r\,\mathrm{d}\theta = \frac{1}{4\pi}\int_0^{2\pi} \cos\theta\sin\theta\,\mathrm{d}r\,\mathrm{d}\theta = 0.$$

Furthermore, we find

$$\mathbb{E}(X) = \int_{-1}^{1} x \cdot \frac{2}{\pi}\sqrt{1-x^2}\,\mathrm{d}x = 0 \quad \text{and} \quad \mathbb{E}(Y) = \int_{-1}^{1} y \cdot \frac{2}{\pi}\sqrt{1-y^2}\,\mathrm{d}y = 0$$

recognizing that the integral of an odd function over a symmetric interval is 0. (Or, one can compute the integrals via first-year calculus substitutions.)