# Math 525: Lecture 23

April 10, 2018

## 1 Joint distribution

In previous lectures, we have dealt with only one-dimensional distribution functions (i.e., $F(x) = \mathbb{P}\{X \le x\}$). In this lecture, we extend our findings to higher dimensions.

**Definition 1.1.** The *joint distribution* of two random variables $X$ and $Y$ (defined on the same probability space) is the function $F : \mathbb{R} \times \mathbb{R} \to [0,1]$ defined by

$$F(x,y) = \mathbb{P}\left\{X \le x, Y \le y\right\}.$$

If we wish to be explicit, we may write $F_{XY}$.

*Remark* 1.2. Similarly, we can define $F(x_1, \ldots, x_n) = \mathbb{P}\{X_1 \le x_1, \ldots, X_n \le x_n\}$ as the joint distribution of $n$ random variables. Since this extension is trivial, we work solely with the two dimensional case for ease of notation.

Recall that distribution functions satisfy some distinctive properties. Let's now extend those to joint distribution functions.

**Proposition 1.3.** *Let $y$ be arbitrary. Then,*

1. *$x \mapsto F(x,y)$ is nondecreasing.*
2. *$x \mapsto F(x,y)$ is right-continuous.*
3. $\lim_{x\to-\infty} F(x,y) = 0$.
4. $\lim_{x\to\infty} F(x,y) = \mathbb{P}\{Y \le y\}$.

*Proof.* Let $B^y = \{Y \le y\}$ and for each $x$, let $A_x = \{X \le x\}$.

1. Since $A_{x_1} \subset A_{x_2}$ whenever $x_1 \le x_2$, then $A_{x_1} \cap B^y \subset A_{x_2} \cap B^y$ and hence $F(x_1, y) = \mathbb{P}(A_{x_1} \cap B^y) \le \mathbb{P}(A_{x_2} \cap B^y) = F(x_2, y)$.
2. Let $x$ be arbitrary and $(x_n)_{n\ge1}$ be a sequence with $x_n \downarrow x$. Then,

$$A_{x_1} \cap B^y \supset A_{x_2} \cap B^y \supset \cdots$$

By continuity of measure,

$$\lim_{n\to\infty} F(x_n, y) = \lim_{n\to\infty} \mathbb{P}(A_{x_n} \cap B^y) = \mathbb{P}\left(\bigcap_{n\geq 1} (A_{x_n} \cap B^y)\right)$$
$$= \mathbb{P}\left(\left(\bigcap_{n\geq 1} A_{x_n}\right) \cap B^y\right) = \mathbb{P}\left(A_x \cap B^y\right) = F(x, y).$$

3. Note that $A_{-1} \supset A_{-2} \supset \cdots$ and hence by continuity of measure,

$$\lim_{n\to\infty} F(-n, y) = \lim_{n\to\infty} \mathbb{P}(A_{-n} \cap B^y) = \mathbb{P}(\emptyset) = 0.$$

4. Note that $A_1 \subset A_2 \subset \cdots$ and hence by continuity of measure,

$$\lim_{n\to\infty} F(n, y) = \lim_{n\to\infty} \mathbb{P}(A_n \cap B^y) = \mathbb{P}(B^y).$$ □

Note that the last point tells us that we can retrieve the distribution function of $Y$ by taking limits in $x$. That is, $F(\infty, y) \equiv \lim_{x\to\infty} F(x, y) = F_Y(y)$, where $F_Y(y) = \mathbb{P}\{Y \leq y\}$ defines the distribution function of $Y$. We call $F(\infty, y)$ the *marginal distribution* of $Y$. The act of taking limits in $x$ is called *marginalization* (i.e., "summing out" the variable you are not interested in).

**Corollary 1.4.** *Let $F_Y$ be the distribution function of $Y$. If $F_Y \neq 0$, then, $x \mapsto F(x, y)/F_Y(y)$ is a distribution function.*

Indeed, we can go a little further in our claim:

$$\frac{F(x, y)}{F_Y(y)} = \frac{\mathbb{P}\left\{X \leq x, Y \leq y\right\}}{\mathbb{P}\{Y \leq y\}} = \mathbb{P}(X \leq x \mid Y \leq y).$$

*Proof.* We proved that $x \mapsto F(x, y)/F_Y(y)$ is nondecreasing, right-continuous, its limit as $x \to -\infty$ is zero and its limit as $x \to \infty$ is

$$\lim_{x\to\infty} \frac{F(x, y)}{F_Y(y)} = \frac{F_Y(y)}{F_Y(y)} = 1.$$ □

Using the same ideas as above (involving continuity of measure) we can prove a few more facts about joint distribution functions. We state them without proof.

**Proposition 1.5.**

1. *$(x, y) \mapsto F(x, y)$ is nondecreasing with respect to the element-wise partial order (i.e., $F(x_1, y_1) \leq F(x_2, y_2)$ whenever $x_1 \leq x_2$ and $y_1 \leq y_2$).*

2. *$\lim_{x,y\to\infty} F(x, y) = 1$.*

As usual, we can define $F(x-,y) = \lim_{t\uparrow x} F(t,y)$, $F(x,y-) = \lim_{s\uparrow y} F(x,s)$, and $F(x-,y-) = \lim_{t\uparrow x, s\uparrow y} F(t,s)$. We use the notation $x \uparrow y$ to mean that $x$ converges to $y$ *strictly* from below (i.e., $x < y$ and $x \to y$).[1]

**Proposition 1.6.** *Let $x_1 < x_2$ and $y_1 < y_2$. Then,*

1. $\mathbb{P}\{X \leq x_2, Y < y_2\} = F(x_2, y_2-)$.
2. $\mathbb{P}\{X < x_2, Y \leq y_2\} = F(x_2-, y_2)$.
3. $\mathbb{P}\{X < x_2, Y < y_2\} = F(x_2-, y_2-)$.
4. $\mathbb{P}\{x_1 < X \leq x_2, y_1 < Y \leq y_2\} = F(x_2,y_2) - F(x_1,y_2) - F(x_2,y_1) + F(x_1,y_1)$.
5. $\mathbb{P}\{X = x_1, Y = y_1\} = F(x_1,y_1) - F(x_1-,y_1) - F(x_1,y_1-) + F(x_1-,y_1-)$.

**Definition 1.7.** If $X$ and $Y$ are discrete random variables, their *joint probability mass function* $p : \mathbb{R} \times \mathbb{R} \to [0,1]$ (written $p_{XY}$ if we wish to be explicit) is defined by

$$p_{XY}(x,y) = \mathbb{P}\{X = x, Y = y\}.$$

In this case, the corresponding joint distribution function is given by

$$F_{XY}(x,y) = \sum_{u \leq x, v \leq y} p_{XY}(u,v).$$

Note that this sum is well defined since even though the set $A = \{(u,v) \colon u \leq x, v \leq y\}$ is uncountable, there are only countably many elements $(u,v) \in A$ such that $p_{XY}(u,v)$ is nonzero.

# 2 Absolutely continuous distributions

**Definition 2.1.** We say $F$ is *absolutely continuous* if it can be written

$$F(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(u,v)\,du\,dv$$

for some integrable function $f$. We call $f$ the joint density of $X$ and $Y$ (sometimes written $f_{XY}$ to be explicit).

From the definition,

$$\mathbb{P}\{x_1 < X \leq x_2, y_1 < Y \leq y_2\} = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f(x,y)\,dx\,dy.$$

More generally, for any Borel measurable set $A \subset \mathbb{R}^2$,

$$\mathbb{P}((X,Y) \in A) = \int\int_A f(x,y)\,dx\,dy.$$

[1] Walsh uses the symbol $\upuparrows$ for this.

Moreover, by the fundamental theorem of calculus, if $f$ is continuous, then

$$f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y).$$

As with marginal distributions, we can also define *marginal densities*:

**Proposition 2.2.** *Suppose $X$ and $Y$ have a joint density $f_{XY}$. Then, $Y$ has a density $f_Y$ satisfying*

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx.$$

*Proof.* We obtain the desired result by showing that $F_Y$ is absolutely continuous with integrand $f_Y$:

$$F_Y(y) = \mathbb{P}\{-\infty < X < \infty, Y \le y\} = \int_{-\infty}^{y} \int_{-\infty}^{\infty} f_{XY}(u,v)dudv = \int_{-\infty}^{y} f_Y(v)dv. \qquad \square$$

The above shows that if $X$ and $Y$ have a joint density, then $Y$ (and also $X$) also has a density. The converse is not true in general:

**Example 2.3.** Let $X \sim U[0,1]$ and $Y = X$ ($X$ and $Y$ are not independent). Of course, both of these variables have the density $f_X = f_Y = 1$. If they had a joint density $f_{XY}$, its probability mass would have to be concentrated on the diagonal $\{(x,y)\colon x = y\}$. However, the area of this diagonal is zero, and hence the integral $\int_0^1 \int_0^1 f_{XY}(x,y)dxdy$ would also be zero, a contradiction.

**Proposition 2.4.** *Suppose $X$ and $Y$ have a continuous joint density $f_{XY}$. Then, they are independent if and only if $f_{XY}(x,y) = f_X(x)f_Y(y)$ at all points $x$ and $y$.*

Continuity is not actually required in the above if we are willing to replace "at all" by "at almost all" since absolutely continuous functions are differentiable almost everywhere (those of you who have taken/are taking measure theory might have seen this).

*Proof.* Suppose $X$ and $Y$ are independent, so that $F_{XY}(x,y) = F_X(x)F_Y(y)$ for all $x$ and $y$. Then, for all $x$ and $y$,

$$\begin{aligned} f_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y}\left[F_X(x)F_Y(y)\right] = \left(\frac{\partial}{\partial x}F_X(x)\right)\left(\frac{\partial}{\partial y}F_Y(y)\right) \\ = f_X(x)f_Y(y). \end{aligned}$$

Suppose $f_{XY}(x,y) = f_X(x)f_Y(y)$ for all $x$ and $y$. Then, for all $x$ and $y$

$$\begin{aligned} F_{XY}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{XY}(u,v)dudv &= \int_{-\infty}^{y}\int_{-\infty}^{x} f_X(u)f_Y(v)dudv \\ &= \left(\int_{-\infty}^{x} f_X(u)du\right)\left(\int_{-\infty}^{y} f_Y(v)dv\right) = F_X(x)F_Y(y). \qquad \square \end{aligned}$$

## 3 Covariances and correlations

So far, we know that two random variables can either be independent or not. This is not particularly useful when we want to make statistical inferences.

For example, if $X$ is someone's muscle mass and $Y$ is the average amount of protein they eat per day, we certainly do not expect $X$ and $Y$ to be independent. We expect that $X$ to be large if $Y$ is large. That is, we expect $X$ and $Y$ to be "positively correlated".

Before we define correlation, let's define the covariance:

**Definition 3.1.** Let $X$ and $Y$ be square integrable random variables. The *covariance* of $X$ and $Y$ is

$$\operatorname{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}X)(Y - \mathbb{E}Y)\right].$$

Note that the covariance above is well-defined since by Cauchy-Schwarz,

$$|\operatorname{Cov}(X, Y)| \leq \sqrt{\mathbb{E}\left[(X - \mathbb{E}X)^2\right]\mathbb{E}\left[(Y - \mathbb{E}Y)^2\right]} = \sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)} < \infty. \tag{1}$$

Note also that $\operatorname{Cov}(X, X) = \operatorname{Var}(X)$. The correlation is just the normalized covariance:

**Definition 3.2.** Let $X$ and $Y$ be square integrable random variables each having nonzero variance. The *correlation* of $X$ and $Y$ is

$$\operatorname{Cor}(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)}\sqrt{\operatorname{Var}(Y)}}.$$

Note that in the above, the correlation is undefined when either $X$ and $Y$ have variance zero. However, this is not a particularly interesting case, since a random variable with variance zero (i.e., a random variable which "does not vary") is actually deterministic:

**Proposition 3.3.** *Let $X$ be a square integrable random variable with $\operatorname{Var}(X) = 0$. Then, $X = \mathbb{E}X$ a.s.*

*Proof.* This follows directly from the fact that $\operatorname{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = 0$ implies that $(X - \mathbb{E}[X])^2 = 0$ a.s. □

**Proposition 3.4.** *If $X$ and $Y$ are square-integrable, then*

1. $-1 \leq \operatorname{Cor}(X, Y) \leq 1$.
2. $\operatorname{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.
3. $\operatorname{Cor}(X, Y) = 1 \iff Y = aX + b$ *a.s. for some constants* $a > 0$ *and* $b$.
4. $\operatorname{Cor}(X, Y) = -1 \iff Y = aX + b$ *a.s. for some constants* $a < 0$ *and* $b$.
5. *The correlation is the covariance of the normalized random variables:*

$$\operatorname{Cor}(X, Y) = \operatorname{Cov}\left(\frac{X}{\sqrt{\operatorname{Var}(X)}}, \frac{Y}{\sqrt{\operatorname{Var}(Y)}}\right).$$

6. *If $X$ and $Y$ are independent, then* $\operatorname{Cor}(X,Y) = \operatorname{Cov}(X,Y) = 0$.

*Proof.*

1. By (1).
2. This follows from

$$\operatorname{Cov}(X,Y) = \mathbb{E}\left[XY - X\mathbb{E}Y - Y\mathbb{E}X + \mathbb{E}X\mathbb{E}Y\right] = \mathbb{E}\left[XY\right] - \mathbb{E}X\mathbb{E}Y. \tag{2}$$

3. Suppose $Y = aX + b$ for some $a > 0$ and $b$. Then,

$$\begin{aligned}\operatorname{Cov}(X,Y) &= \operatorname{Cov}(X, aX+b) = \mathbb{E}\left[X\left(aX+b\right)\right] - \mathbb{E}X\mathbb{E}\left[aX+b\right] \\ &= a\mathbb{E}\left[X^2\right] + b\mathbb{E}\left[X\right] - \mathbb{E}X\left(a\mathbb{E}\left[X\right] + b\right) = a\left(\mathbb{E}\left[X^2\right] - \left(\mathbb{E}X\right)^2\right) = a\operatorname{Var}(X).\end{aligned}$$

Moreover, $\operatorname{Var}(Y) = \operatorname{Var}(aX+b) = a^2\operatorname{Var}(X)$. Therefore,

$$\operatorname{Cor}(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{a^2\operatorname{Var}(X)}\sqrt{\operatorname{Var}(X)}} = \frac{a\operatorname{Var}(X)}{a\operatorname{Var}(X)} = 1.$$

Now, suppose $\operatorname{Cor}(X,Y) = 1$. This means that (1) holds with equality, which only occurs if

$$Y - \mathbb{E}Y = a\left(X - \mathbb{E}X\right) \qquad \text{a.s.}$$

for some constant $a$ (we proved this in an early lecture). Therefore,

$$Y = aX \underbrace{-a\mathbb{E}X + \mathbb{E}Y}_{b} \qquad \text{a.s.}$$

4. Similar to the above.
5. This follows by two applications of (2):

$$\begin{aligned}\operatorname{Cov}\left(\frac{X}{\sqrt{\operatorname{Var}(X)}}, \frac{Y}{\sqrt{\operatorname{Var}(Y)}}\right) &= \mathbb{E}\left[\frac{XY}{\sqrt{\operatorname{Var}(X)}\sqrt{\operatorname{Var}(Y)}}\right] - \frac{\mathbb{E}X\mathbb{E}Y}{\sqrt{\operatorname{Var}(X)}\sqrt{\operatorname{Var}(Y)}}. \\ &= \frac{\mathbb{E}\left[XY\right] - \mathbb{E}X\mathbb{E}Y}{\sqrt{\operatorname{Var}(X)}\sqrt{\operatorname{Var}(Y)}} = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)}\sqrt{\operatorname{Var}(Y)}} = \operatorname{Cor}(X,Y).\end{aligned}$$

6. By (2). □

**Definition 3.5.** If $\operatorname{Cov}(X,Y) = 0$, we say $X$ and $Y$ are *orthogonal*.

While independence implies orthogonality, the converse is not true in general.