JOINT DISTRIBUTIONS
AN EXTRACT

Jerzy Szulga

Department of Mathematics and Statistics
Auburn University

MATH 5670-6670 FALL 2019

PROBABILITY I

November 6, 2019

Outline

We consider the distribution of random pairs (X, Y), triples (X, Y, Z), or sequences $(X_1, \ldots, X_n)$. For simplicity, we confine to the bivariate case. The common way is through

the **joint cdf** $F(x, y) = \mathrm{P}(X \le x,\ Y \le y))$.

More conveniently:

discrete: **joint pmf** $f(x, y) = \mathrm{P}(X = x,\ Y = y)$, $(x, y) \in S$;

continuous: **joint pdf** $f(x, y) = \dfrac{\partial^2 F(x, y)}{\partial x \partial y}$, $(x, y) \in S$.

The support S is now even more important.

Graphs

An equation $z = f(x, y)$ represents a surface in 3-space over a 2D region S in the xy=plane.

In the discrete case a histogram is commonly used, consisting of narrow blocks over a square base around a point (x, y).
(The upper walls of these blocks form a piecewise constant surface.)

Since 3D graphing requires some skill and training, therefore 2D projections may be very useful.

Marginal distributions

Discrete: $f_X(x) = \sum_y f(x, y)$

Continuous: $f_X(x) = \int_{\mathbb{R}} f(x, y)\, dy.$

The summation or integration involves only y such that $(x, y) \in S$.

Similarly for $f_Y(y)$.

Example 1

Say, in the continuous case the pdf depends on $x^2 + y^2$, e.g.

$$f(x, y) = c\left(1 - x^2 - y^2\right), \ x^2 + y^2 \leq 1.$$

Therefore, it's a surface of revolution. Thus, we may sketch the projection onto the yz-plane, $f(0, y) = 1 - y^2$, $|y| \leq 1$.
Also we sketch the support, the disk $x^2 + y^2 \leq 1$ on the xy-plane.

Here, the constant c is needed to make the volume equal 1. It requires some Calculus 3 to compute it. It turns out that

$$c = \frac{2}{\pi}$$

Example 2a

A spinner is divided into halves, then one of the halves is divided into two quarters. The half is marked by 0 and the quarters are marked by 1 and 2. We spin twice, denoting the outcomes by X, Y.

In lieu of the histogram we present the table of probabilities:

		Y = 0	Y = 1	Y = 2	
X	0	$\frac{4}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{8}{16}$
	1	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{4}{16}$
	2	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{4}{16}$
		$\frac{8}{16}$	$\frac{4}{16}$	$\frac{4}{16}$	1

Summing up along rows/columns we obtain **marginal distributions** of X and Y (iid here)

Example 2b, modified

Now let us record the outcomes only when $X + Y \leq 2$. Skip other.

That is, now $f_1(x, y) = \mathrm{P}(X = x, Y = y | X + Y \leq 2)$.

It helps to list only integer numerators:

$$\begin{array}{cc||c|c|c||c} & & & y & & \\ & & 0 & 1 & 2 & \\ \hline\hline & 0 & 4 & 2 & 2 & \color{blue}{8} \\ \hline x & 1 & 2 & 1 & 1 & \color{blue}{4} \\ \hline & 2 & 2 & 1 & 1 & \color{blue}{4} \\ \hline\hline & & \color{blue}{8} & \color{blue}{4} & \color{blue}{4} & 16 \end{array} \mapsto \begin{array}{cc||c|c|c||c} & & & y & & \\ & & 0 & 1 & 2 & \\ \hline\hline & 0 & 4 & 2 & 2 & \color{blue}{8} \\ \hline x & 1 & 2 & 1 & & \color{blue}{3} \\ \hline & 2 & 2 & & & \color{blue}{2} \\ \hline\hline & & \color{blue}{8} & \color{blue}{3} & \color{blue}{2} & 13 \end{array}$$

By dividing the entries by 13 we obtain the new joint probability distribution table. (This is a "c-problem").

Independent vs. dependent

In the first table $f(x, y)$ we see that

$$f(x, y) = f_X(x) \cdot f_Y(y) \quad \textbf{for each and every } x, y$$

Random variables are **independent**.

In the second table it's not true - RVs are **dependent**.

We see it right away - the support is not rectangular.

Grave Logical Error

"*RVs are independent because the support is rectangular*".

Zero credit in a quiz or exam!

Conditional distribution

In either discrete or continuous case:

$$f(x|y) = \frac{f(x,y)}{f_Y(y)} \quad \text{or} \quad f(y|x) = \frac{f(x,y)}{f_X(x)}.$$

Notice that in the discrete case $f(x|y) = \mathrm{P}(X = x | Y = y)$.

Conditioning

Discrete:

Conditional expectation: $\mathrm{E}\left[X|Y=y\right]=\sum_{x} x\, f(x|y)$

Conditional moments: $\mathrm{E}\left[X^n|Y=y\right]=\sum_{x} x^n\, f(x|y)$,

Etc.

Continuous: s.a.a., just replace $\sum$ by $\int$.

Conditional variance:

$$\mathrm{Var}(X|Y=y)=\mathrm{E}\left[X^2|Y=y\right]-\Big(\mathrm{E}\left[X|Y=y\right]\Big)^2$$

Transformations

Given (X, Y) and a bivariate function $h(x, y)$ we transform $Z = h(X, Y)$. Then, as usual

$$\mathrm{E}\, h(X, Y) = \begin{cases} \displaystyle\sum_x \sum_y h(x, y)\, f(x, y) & \text{if discrete} \\ \displaystyle\iint_{\mathbb{R}^2} h(x, y)\, f(x, y)\, dx\, dy & \text{if continuous} \end{cases}$$

The product $h(x, y) = xy$ yields the **covariance** σ_{XY}

$$\mathrm{Cov}(X, Y) \stackrel{\text{def}}{=} \mathrm{E}\left(X - \mathrm{E}\, X\right)\left(Y - \mathrm{E}\, Y\right) = \mathrm{E}\left(XY\right) - \left(\mathrm{E}\, X\right)\left(\mathrm{E}\, Y\right).$$

Earlier encounters with covariance

Recall the **inner product** (e.g., Linear Algebra MATH 2660):

In a vector space, it is a real scalar-valued mapping $(\mathbf{u}, \mathbf{v})$:

- $(\mathbf{v}, \mathbf{v}) \geq 0$ and $(\mathbf{v}, \mathbf{v}) = 0$ iff $\mathbf{v} = \mathbf{0}$;
- symmetry: $(\mathbf{u}, \mathbf{v}) = (\mathbf{v}, \mathbf{u})$;
- bi-linearity: $(\mathbf{u}, \cdot)$ is linear

Thus, in the vector space of random variables with mean zero and finite variances, the covariance is an inner product.

It yields a norm $||X|| = \sigma_X = \sqrt{\mathrm{Var}(X)}$ and the angle between RVs:

$$\cos\theta = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Covariance algebra

If means are zero, then simply $\mathrm{Cov}(X, Y) = \mathrm{E}\,(XY)$.

Clearly, $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$.

Exercise. Show that

$$\begin{aligned}\mathrm{Cov}(X, Y) &= \frac{\mathrm{Var}(X+Y) - \mathrm{Var}(X-Y)}{4} \\ &= \frac{\mathrm{Var}(X+Y) - \mathrm{Var}(X) - \mathrm{Var}(Y)}{2}.\end{aligned}$$

Hint: That's algebra: e.g., $ab = \frac{(a+b)^2-(a-b)^2}{4} = \frac{(a+b)^2-a^2-b^2}{2}$.

More algebra - one famous inequality

Exercise. Show that $\Big(\mathrm{E}\,(XY)\Big)^2 \le \mathrm{E}\,(X^2)\,\mathrm{E}\,(Y^2)$.

Solution. Obviously, $\mathrm{E}\,(tX - Y)^2 \ge 0$ for every t. Expand the LHS:

$$\mathrm{E}\,(tX - Y)^2 = t^2\,\mathrm{E}\,(X^2) - 2t(\mathrm{E}\,XY) + \mathrm{E}\,(Y^2).$$

This a quadratic function $at^2 + bt + c \ge 0$ for every t, which means that the discriminant $b^2 - 4ac \le 0$, which translates to our inequality.

Even more algebra

In the case of a single root t, when the discriminant is 0, the quadratic function takes the single value 0.

That is, $tX - Y = 0$, i.e., $Y = tX$.

This happens iff $\Big(\mathrm{E}\,(XY)\Big)^2 = \mathrm{E}\,(X^2)\,\mathrm{E}\,(Y^2)$.

In other words, the famous inequality becomes the factual equality iff one variable is a constant multiple of the other.

Correlation coefficient

The standardized covariance is called the **correlation coefficient**

$$\rho = \rho_{X,Y} \stackrel{\text{def}}{=} \mathrm{Cov}\left(\widetilde{X}, \widetilde{Y}\right) = \mathrm{E}\left(\widetilde{X} \cdot \widetilde{Y}\right) = \frac{\mu_{XY} - \mu_X \mu_Y}{\sigma_X \sigma_Y}$$

From the latter exercise $\boldsymbol{-1 \leq \rho \leq 1}$.

Corollary

$\rho = \pm 1$ *iff* $\widetilde{Y} = t\widetilde{X}$ *iff* $Y = aX + b$ *for some* a, b.

That, is the extreme values ± 1 indicate that one random variable is just an affine transformation of another. That's very dependent!

Linear approximation

Standardized random variables involve no shift.

(The presence of shift messes up all formulas.)

Still, we can bring back the shift any time at will.

Suppose that we desire and believe a linear relation:

$$\widetilde{Y} = t\widetilde{X}$$

That may be not true, it might be an error, but nevertheless what is the "best" t that minimizes the error?

Least square error

We use the variance to measure and minimize the error.

$$\mathrm{Var}\left(\widetilde{Y} - t\widetilde{X}\right) = \mathrm{Var}\left(\widetilde{Y}\right) - 2t\mathrm{Cov}\left(\widetilde{Y}, \widetilde{X}\right) + t^2\mathrm{Var}\left(\widetilde{X}\right) = 1 - 2t\rho + t^2.$$

Clearly, the minimum is attained for $t = \rho$.

This is the **regression line**:

$$\widetilde{\mathbf{y}} = \boldsymbol{\rho}\,\widetilde{\mathbf{x}}$$

(after standardization.)

Alternative formulas

Expanding the latter equation:

$$\frac{y-\mu_Y}{\sigma_Y}=\rho\left(\frac{x-\mu_x}{\sigma_x}\right).$$

Rewriting and presenting in the popular form:

$$y=\mu_Y+\rho\frac{\sigma_Y}{\sigma_X}\Big(x-\mu_X\Big)$$

Definition

We say that two random variables X and Y with finite variances are **uncorrelated** if $\mathrm{Cov}(X, Y) = 0$; equivalently, iff $\rho = 0$.

If either variance is infinite, the notion is undefined.

Theorem

Consider random variables with finite variances.

Independent random variables are uncorrelated.

The converse implication is false, with only few exceptions.

Comparison to independence

The null correlation is rather fragile.

If X and Y are uncorrelated, their transforms such as X^2 and Y^2 typically loose this property. In contrast, independence is preserved under arbitrary transforms.

Nevertheless, the weaker notion of zero correlation has a great advantage due to its simplicity.