Probability and statistics

Lecture notes from March 6.

Definition 1. *The function $f(x,y)$ is called the density function(density) of the 2-dimensional random variable (X,Y) iff*

1. $f(x,y) \geq 0$, *for* $(x,y) \in \mathbb{R}^2$,
2. $\int_{\mathbb{R}} \int_{\mathbb{R}} f(x,y)\, dydx = 1.$

For the discrete case, the function $p(i,j)$ is called the density (probability) iff

1. $P(X = x_i, Y = y_j) = p_{ij} \geq 0$, *for* $(i,j) \in \mathbb{N}^2$,
2. $\sum_{i\in\mathbb{N}} \sum_{j\in\mathbb{N}} p_{ij} = 1.$

Definition 2. *The marginal density function of a random variable (X,Y) with density $f(x,y)$ are the functions $f_1(x) = \int_{\mathbb{R}} f(x,y)dy$ and $f_2(y) = \int_{\mathbb{R}} f(x,y)dx$.*
For the discrete case, we use the symbols: $p_{i\bullet} = \sum_{j\in\mathbb{N}} p_{ij}$ and $p_{\bullet j} = \sum_{i\in\mathbb{N}} p_{ij}$.

Definition 3.

- *The moment of the order of k (k-th moment) of the random variable X is $m_k = E(X^k)$.*
- *The central moment of the order k (k-th central moment) is equal to $\mu_k = E\Big[(X - EX)^k\Big]$.*
- *For a 2-dimensional random variable (X,Y), the mixed moment of the order (k,l) is $m_{kl} = E(X^k Y^l)$ and $\mu_{kl} = E\Big[(X - EX)^k \cdot (Y - EY)^l\Big]$.*

Comments

Continuous distribution	Discrete distribution
$m_k = \int_{\mathbb{R}} x^k f(x)dx$	$m_k = \sum_{i\in\mathbb{N}} x_i^k p_i$
$\mu_k = \int_{\mathbb{R}} (x - \mathrm{E}X)^k f(x)dx$	$\mu_k = \sum_{i\in\mathbb{N}} (x_i - \mathrm{E}X)^k p_i$
$m_{kl} = \int_{\mathbb{R}} \int_{\mathbb{R}} x^k y^l f(x,y)dydx$	$m_{kl} = \sum_{i,j\in\mathbb{N}} x_i^k y_j^l p_{ij}$

The expected value $\mathrm{E}X$ is m_1, the variance $\mathrm{V}X$ is μ_2, the mixed moment m_{11} is the covariance of the variables X, Y, also denoted as $\mathrm{Cov}(X,Y)$. The symbols $\mathrm{E}X$, $\mathrm{E}(X)$ mean the same (expected value). Similarly, the symbols $\mathrm{V}X$, $\mathrm{V}(X)$ mean variance. For distinction: $\mathrm{E}(X^2)$ is the expected value of the variable X^2, while $\mathrm{E}^2(X)$ or $[\mathrm{E}X]^2$ - the square of the expected value of the variable X.

Definition 4. *Let (X,Y) be a 2-dimensional random variable. Variables (1-dimensional) X,Y are called independent iff*

- $\forall x,y \in \mathbb{R}\;\; f(x,y) = f_1(x) \cdot f_2(y)$ (*vide* def. 2), or
- $\forall i,j \in \mathbb{N}\;\; p_{ij} = p_{i\bullet} \cdot p_{\bullet j}$.

Example:

(i) We consider the function $f(x,y) = \dfrac{3xy}{16}$ defined in the area bounded by the straight lines $y = 0$, $x = 2$ and the curve $y = x^2$. Immediately we obtain

$$\int_0^2 \int_0^{x^2} x \cdot y \, dy \, dx = \frac{16}{3}, \tag{1}$$

i.e. the non-negative function $f(x,y)$ in the selected area can be considered as density. The equation (1) should be written as

$$\int_{\mathbb{R}} \int_{\mathbb{R}} f(x,y) \, dy \, dx = 1,$$

where by default, we assume that $f(x,y) = 0$ anywhere outside the defined area.

(ii) The 2-dimensional cumulative distribution $F(s,t) \equiv F_{XY}(x,y) = \displaystyle\int_{-\infty}^{s} \int_{-\infty}^{t} f(x,y) \, dy \, dx$.

(iii) Let us determine the marginal densities:

$$f_1(x) \equiv f_X(x) = \int_{\mathbb{R}} \frac{3xy}{16} dy = \int_0^{x^2} \frac{3xy}{16} dy = \frac{3x^5}{32}, \quad x \in [0,2],$$
$$f_2(y) \equiv f_Y(y) = \int_{\mathbb{R}} \frac{3xy}{16} dx = \int_{\sqrt{y}}^{2} \frac{3xy}{16} dx = \frac{3(4-y)\,y}{32}, \quad y \in [0,4].$$

Note (in the first equation) that, $y \in [0, x^2]$ for the variable x. Similarly, $x \in [\sqrt{y}, 2]$ for the variable y. In addition, $f_1(x), f_2(x)$ can be considered as densities because they are non-negative (in the defined area) and

$$\int_0^2 f_1(x) \, dx = \int_0^2 \frac{3x^5}{32} dx = 1, \quad \int_0^4 f_2(y) \, dy = \int_0^4 \frac{3(4-y)\,y}{32} dx = 1.$$

Subsequent integrals give the results: $\mathrm{E}X = \displaystyle\int_0^2 x \cdot f_1(x) \, dx = \frac{12}{7}$ and $\mathrm{E}Y = \displaystyle\int_0^4 y \cdot f_2(y) \, dy = 2$ (but it is such addition).

Let's check if the variables X, Y are independent. Let's consider $(x,y) = (1, 1.5)$. Then

$$0 = f(1, 1.5) \neq f_1(1) \cdot f_2(1.5) = \frac{3}{32} \cdot \frac{45}{128}.$$

This means that the variables X, Y are not independent (In short: they are dependent).

Example:

The next example shows that it is much simpler for the discrete variables. Also **for this reason** (there are other reasons) we will pay less attention to these variables. (i) We define the random variable (X,Y) as follows:

$$(X,Y) = \begin{array}{c|ccc|} & y_1 & y_2 & y_3 \\ \hline x_1 & p_{11} & p_{12} & p_{13} \\ x_2 & p_{21} & p_{12} & p_{23} \\ x_3 & p_{31} & p_{32} & p_{33} \\ x_4 & p_{41} & p_{42} & p_{43} \end{array}$$

In the above formula, $p_{ij} \geq 0$ and $\sum\limits_{i,j} p_{ij} = 1$, i.e. for example:

$$(X,Y) = \begin{array}{r|rrr|} X/Y & 2 & 3 & 5 \\ \hline -2 & 0.10 & 0.05 & 0.07 \\ 0 & 0.05 & 0.03 & 0.08 \\ 1 & 0.01 & 0.07 & 0.15 \\ \sqrt{2} & 0.38 & 0.00 & 0.01 \end{array}$$

(ii) Distribution $F(s,t) = \sum\limits_{x_i \leq s,\, y_j \leq t} p_{ij}$.

(iii) Marginal density:

$$(X,Y) = \begin{array}{r|rrr|r} X/Y & 2 & 3 & 5 & p_{1\bullet} \\ \hline -2 & 0.10 & 0.05 & 0.07 & 0.22 \\ 0 & 0.05 & 0.03 & 0.08 & 0.16 \\ 1 & 0.01 & 0.07 & 0.15 & 0.23 \\ \sqrt{2} & 0.38 & 0.00 & 0.01 & 0.39 \\ \hline p_{\bullet j} & 0.54 & 0.15 & 0.31 & 1.00 \end{array}$$

Therefore, the variables X, Y have the marginal distributions:

$$X = \begin{array}{c|cccc} x_i & -2 & 0 & 1 & \sqrt{2} \\ \hline p_{i\bullet} & 0.22 & 0.16 & 0.23 & 0.39 \end{array}, \quad Y = \begin{array}{c|ccc} y_j & 2 & 3 & 5 \\ \hline p_{\bullet j} & 0.54 & 0.15 & 0.31 \end{array}$$

(iv) Independence of the variables X, Y: The variables are dependent because

$$0.01 = p_{31} = \mathrm{P}(X = 1, Y = 2) \neq \mathrm{P}(X = 1) \cdot \mathrm{P}(Y = 2) = p_{3\bullet}\, p_{\bullet 1} = 0.23 \cdot 0.54.$$

Polski	**Angielski**
dystybuanta	cdf (cummulative distribution function)
gęstość	density (pdf)
gęstość brzegowa	marginal density
gęstość warunkowa	conditional density
i.i.d.[1]	i.i.d. [2]
kowariancja	covariance
rozkład	distribution
rozkład ciągły	continuous distribution

[1](zmienne) niezależne o tym samym rozkładzie

[2]independent and of identical distribution (variables)

Polski	**Angielski**
rozkład dyskretny	discrete distribution
wariancja	variance
wartość oczekiwana	expected value
zmienna losowa	random variable
zmienne niezależne	independent variables

Z poważaniem,
Witold Karczewski