Probability and Statistics

Alvin Lin

Probability and Statistics: January 2017 - May 2017

Jointly Distributed Random Variables

Let X and Y be two discrete random variables defined on the same sample space S of an experiment. The joint pmf $p(x, y)$ is defined for any numbers x and y by:

$$p(x, y) = P(X = x \; and \; Y = y)$$

The function $p(x, y)$ satisfies the following properties.

1. $p(x, y) \geq 0$ for any x, y
2. $\sum_x \sum_y p(x, y) = 1$

Marginal PMF

Marginal pmf of X:

$$p_X(x) = \sum_{y:p(x,y)>0} p(x, y)$$

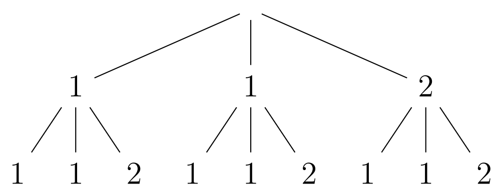
Jointly Distributed Continuous Random Variables

Let X and Y be continuous random variables. A joint pdf $f(x, y)$ satisfies the following conditions.

1. $P((X, Y) \in A) = \int_A \int f(x, y) \, \mathrm{d}x \, \mathrm{d}y$
2. $f(x, y) \geq 0 \; for \; any \; x \; and \; y$
3. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, \mathrm{d}x \, \mathrm{d}y = 1$

Example

A hat contains two balls marked 1 and one ball marked 2. Select a ball randomly, and then with replacement select a ball randomly. Let X be the number on the first ball and Y be the number on the second ball. Let $p(x, y)$ be the joint pmf of X and Y. Let $p(x)$ and $p(y)$ be the marginal pmf of X and Y respectively.



$$\begin{aligned} p_x(1) &= \sum_y p(1, y) \\ &= p(1,1) + p(1,2) \\ &= \frac{4}{9} + \frac{2}{9} = \frac{6}{9} \\ p_x(2) &= \sum_y p(2, y) \\ &= p(2,1) + p(2,2) \\ &= \frac{2}{9} + \frac{1}{9} = \frac{3}{9} \end{aligned}$$

$$\begin{aligned} p_y(1) &= \sum_x p(x, 1) \\ &= p(1,1) + p(2,1) \\ &= \frac{4}{9} + \frac{2}{9} = \frac{6}{9} \\ p_y(2) &= \sum_x p(x, 2) \\ &= p(1,2) + p(2,2) \\ &= \frac{2}{9} + \frac{1}{9} = \frac{3}{9} \end{aligned}$$

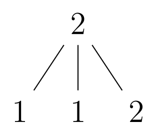
x	y	$p_x(x)$	$p_y(y)$	$p_x(x) \cdot p_y(y)$	$p(x, y)$
1	1	$\frac{6}{9}$	$\frac{6}{9}$	$\frac{36}{81} = \frac{4}{9}$	$\frac{4}{9}$
1	2	$\frac{6}{9}$	$\frac{3}{9}$	$\frac{18}{81} = \frac{2}{9}$	$\frac{2}{9}$
2	1	$\frac{3}{9}$	$\frac{6}{9}$	$\frac{18}{81} = \frac{2}{9}$	$\frac{2}{9}$
2	2	$\frac{3}{9}$	$\frac{3}{9}$	$\frac{9}{81} = \frac{1}{9}$	$\frac{1}{9}$

$$p_x(1) = P(X = 1) = \frac{6}{9}$$

X and Y are independent. Find the conditional probability function of Y given that $X = 2$.

$$p_{y|x}(y|2) = \frac{p(2, y)}{p_x(2)} = \begin{cases} \frac{3}{3} & , y = 1 \\ \frac{1}{3} & , y = 2 \\ 0 & , otherwise \end{cases}$$

Suppose we have $p_{y|x}(1|2)$. This represents the probability that $Y = 1$ given $X = 2$. Of the sample space, the new subspace consists of:



$$p_{y|x}(1|2) = P(Y = 1 \textit{ given that } X = 2) = \frac{2}{3}$$

Double Integral

If $f(x) \geq 0$ for any $x \in \mathbb{R}$, $\int_1^3 f(x)\, \mathrm{d}x$ is the area under the curve $y = f(x)$, $x \in [1, 3]$. If $f(x, y) \geq 0$ for any $(x, y) \in \mathbb{R}^2$, $\int_1^2 \int_1^3 f(x, y)\, \mathrm{d}x\, \mathrm{d}y$ is the volume under the surface $\delta = f(x, y)$ where:

$$(x, y) \in A = \{(x, y) \in \mathbb{R}^2 \mid 1 \leq x \leq 3 \textit{ and } 1 \leq y \leq 2\}$$

Suppose $f(x, y)$ is the joint pdf uniformly distributed on the rectangle A, then:

$$f(x, y) = \begin{cases} c & , (x, y) \in A \\ 0 & , otherwise \end{cases}$$

How do we find the value of the constant c?

$$
\begin{aligned}
1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\ \mathrm{d}x\ \mathrm{d}y \\
&= \int\int_A f(x,y)\ \mathrm{d}x\ \mathrm{d}y \\
&= \int\int_A c\ \mathrm{d}x\ \mathrm{d}y \\
&= c\int\int_A\ \mathrm{d}x\ \mathrm{d}y \\
&= c\Big[area\ of\ A\Big] \\
&= c(2-1)(3-1) \\
c &= \frac{1}{2}
\end{aligned}
\qquad
\begin{aligned}
1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\ \mathrm{d}x\ \mathrm{d}y \\
&= \int\int_A f(x,y)\ \mathrm{d}x\ \mathrm{d}y \\
&= \int\int_A c\ \mathrm{d}x\ \mathrm{d}y \\
&= c\int\int_A\ \mathrm{d}x\ \mathrm{d}y \\
&= c\int_1^2\Bigg[\int_1^3\ \mathrm{d}x\Bigg]\ \mathrm{d}y \\
&= c\int_1^2\Bigg[x\Bigg]_{x=1}^{x=3}\ \mathrm{d}y \\
&= c\int_1^2\Bigg[3-1\Bigg]\ \mathrm{d}y \\
&= c(3-1)\int_1^2\ \mathrm{d}y \\
&= c(3-1)\Bigg[y\Bigg]_{y=1}^{y=2} \\
&= c(3-1)(2-1) \\
c &= \frac{1}{2}
\end{aligned}
$$

The volume under the curve is numerically equal to the area of A even though they are dimensionally different because the surface is flat and its height is equal to 1.

Example

Annie and Alvie have agreed to meet between 5:30PM and 6:00PM for dinner at a local health-food restaurant. Let $X = Annie's\ arrivaltime$ and $Y = Alvie's\ arrival\ time$. Suppose X and Y are independent with each uniformly distributed on the interval [5:30,6:00] ($[5.5, 6]$). What is the joint pdf of X and Y?

Let $A = \{(x,y) \in \mathbb{R}^2 \mid 5.5 \leq x \leq 65.5 \leq y \leq 6\}$.

$$f(x,y) = \begin{cases} c & ,(x,y) \in A \\ 0 & ,otherwise \end{cases}$$

$$
\begin{aligned}
1 &= \int_{-\infty} \infty \int_{-\infty} \infty f(x,y)\ \mathrm{d}x\ \mathrm{d}y \\
&= \int \int_A f(x,y)\ \mathrm{d}x\ \mathrm{d}y \\
&= \int \int_A c\ \mathrm{d}x\ \mathrm{d}y \\
&= c\Big[area\ of\ A\Big] \\
&= c \cdot \frac{1}{2} \cdot \frac{1}{2} \\
c &= 4
\end{aligned}
$$

$$
f(x,y) = \begin{cases} 4 & ,(x,y) \in A \\ 0 & ,otherwise \end{cases}
$$

What is the probability that both arrive between 5:15 and 5:45?

The area of overlap is only between 5:30 and 5:45, so the area is $\frac{1}{4} \cdot \frac{1}{4}$.

$$
\frac{overlap}{total} = \frac{\frac{1}{4} \cdot \frac{1}{4}}{\frac{1}{2} \cdot \frac{1}{2}} = \frac{1}{4}
$$

If the first one to arrive will wait only 10 minutes before leaving to eat elsewhere, what is the probability that Annie arrives before Alvie and that they have dinner at the health-food restaurant? We can write these conditions as follows:

$$|x - y| \leq \frac{1}{6}$$

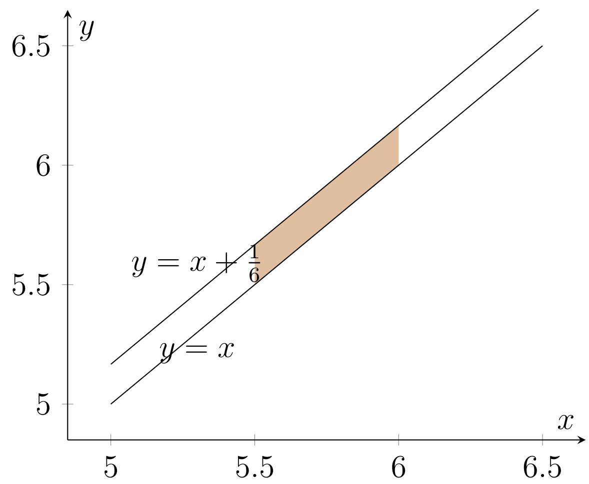
$$x \leq y$$

$$-\frac{1}{6} \leq x - y \leq \frac{1}{6}$$

$$y \leq x + \frac{1}{6}$$

$$x - \frac{1}{6} \leq y$$

$$y \geq x - \frac{1}{6}$$



The area of this region (trapezoidal) is:

$$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} - \frac{1}{2}(\frac{1}{2} - \frac{1}{6})^2 = \frac{1}{8} - \frac{1}{18}$$

$$P = \frac{area}{total} = \frac{\frac{1}{8} - \frac{1}{18}}{\frac{1}{2} \cdot \frac{1}{2}}$$

You can find all my notes at `http://omgimanerd.tech/notes`. If you have any questions, comments, or concerns, please contact me at alvin@omgimanerd.tech