Probability Theory

Joint Distribution of RVs

- In real life, we are often interested in two (or more) random variables at the same time. For example,
  - we might measure the height and weight of an object, or
  - frequency of exercise and rate of heart disease in adults,
  - level of air pollution and rate of respiratory illness in cities,
  - number of Facebook friends and age of Facebook members
- Joint distribution allows us to compute probabilities of events involving both variables and understand the relationship between the variables.

Joint Distribution of Discrete RVs

- Suppose X and Y are two discrete random variables.
  - X takes values $\{x_1, x_2, . . . , x_n\}$ and Y takes values $\{y_1, y_2, . . . , y_m\}$. The ordered pair (X, Y) take values in the product$\{(x_1, y_1),(x_1, y_2), . . .(x_n, y_m)\}$.
- The joint probability mass function (joint pmf) of X and Y is the function $p(x_i, y_j)$ giving the probability of the joint outcome $X = x_i$, $Y = y_j$.

Joint probability mass function must satisfy two properties:

1. $0 \leq p(x_i, y_j) \leq 1$
2. The total probability is 1.

$$\sum_{i=1}^{n} \sum_{j=1}^{m} p(xi, yj) = 1$$

$X \backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	y_m
x_1	$p(x_1, y_1)$	$p(x_1, y_2)$	$\cdots$	$p(x_1, y_j)$	$\cdots$	$p(x_1, y_m)$
x_2	$p(x_2, y_1)$	$p(x_2, y_2)$	$\cdots$	$p(x_2, y_j)$	$\cdots$	$p(x_2, y_m)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_i	$p(x_i, y_1)$	$p(x_i, y_2)$	$\cdots$	$p(x_i, y_j)$	$\cdots$	$p(x_i, y_m)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	
x_n	$p(x_n, y_1)$	$p(x_n, y_2)$	$\cdots$	$p(x_n, y_j)$	$\cdots$	$p(x_n, y_m)$

Joint Distribution of Discrete RVs

Q1. Roll two dice. Let X be the value on the first die and let T be the total on both dice. Draw the joint probability table.

$X\backslash T$	2	3	4	5	6	7	8	9	10	11	12
1	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	0
2	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0
3	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0
4	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0
5	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0
6	0	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36

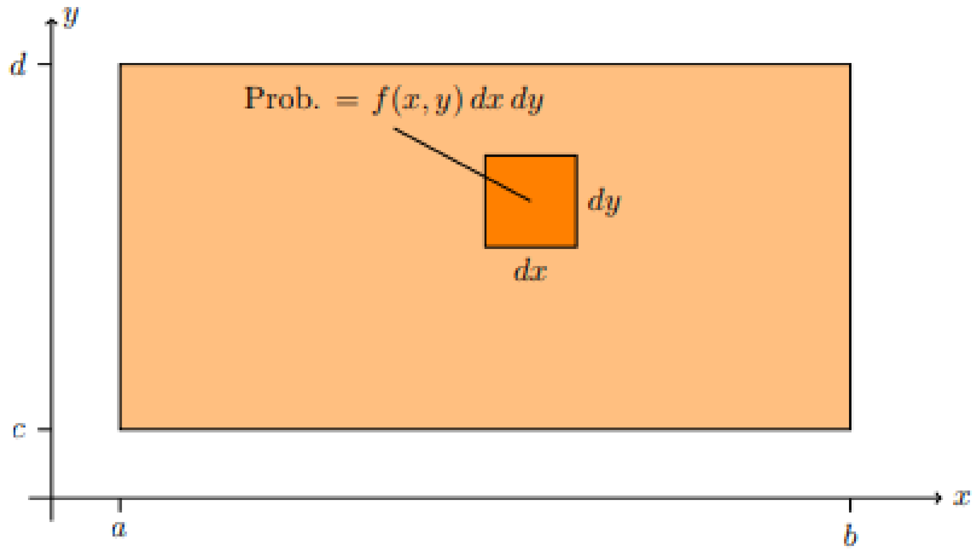
Q2. Roll two dice. Let X be the value on the first die and let Y be the value on the second die. Then both X and Y take values 1 to 6 and the joint pmf is p(i, j) = 1/36 for all i and j between 1 and 6. Draw the Joint probability table and find the probability of event B= ‘ X-Y ≥ 2’.

Joint Distribution of Continuous RVs

If X takes values in [a, b] and Y takes values in [c, d] then the pair (X, Y) takes values in the product [a, b] × [c, d].

- The joint probability density function (joint pdf) of X and Y is a function f(x, y) giving the probability density at (x, y).
- That is, the probability that (X, Y) is in a small rectangle of width dx and height dy around (x, y) is f(x, y)dxdy.
- A joint PDF must satisfy:
  1. $0 \leq f(x, y)$
  2. The total probability is 1.

$$\int_c^d \int_a^b f(x, y)dxdy = 1$$



Joint Cumulative Distributions RVs

Suppose X and Y are jointly-distributed random variables. We will use the notation '$X \leq x, Y \leq y$' to mean the event '$X \leq x$ and $Y \leq y$'. The joint cumulative distribution function (joint cdf) is defined as

$$F(x, y) = P(X \leq x, Y \leq y)$$

Continuous case: If X and Y are continuous random variables with joint density $f(x, y)$ over the range $[a, b] \times [c, d]$ then the joint cdf is given by the double integral

$$F(x, y) = \int_c^y \int_a^x f(u, v)\, du\, dv.$$

To recover the joint pdf, we differentiate the joint cdf. Because there are two variables we need to use partial derivatives:

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

Discrete case: If X and Y are discrete random variables with joint pmf $p(x_i, y_j)$ then the joint cdf is give by the double sum

$$F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} p(x_i, y_j).$$

Joint Distribution of Continuous RVs

Q3. Let X & Y both take values in [0,1] with density f(x, y) = 4xy.

i. Show f(x, y) is a valid joint PDF,

ii. Visualize the event A = 'X < 0.5 and Y > 0.5' and find its probability.

To show $f(x, y)$ is a valid joint pdf we must check that it is positive (which it clearly is) and that the total probability is 1.

$$\text{Total probability} = \int_0^1 \int_0^1 4xy \, dx \, dy = \int_0^1 \left[2x^2 y\right]_0^1 dy = \int_0^1 2y \, dy = 1. \quad \text{QED}$$

The event A is just the upper-left-hand quadrant. Because the density is not constant we must compute an integral to find the probability.

$$P(A) = \int_0^{.5} \int_{.5}^1 4xy \, dy \, dx = \int_0^{.5} \left[2xy^2\right]_{.5}^1 dx = \int_0^{.5} \frac{3x}{2} \, dx = \boxed{\frac{3}{16}}.$$

Q4. Let X & Y both take values in [0,1] with density f(x, y) = 4xy. Find Joint CDF of X and Y.

Marginal Density RVs

Given a joint density for X and Y, we define the marginal density of X to be

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

and the marginal density of Y to be

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

As usual, we restrict the integral to the region where f is positive when that is not the entire plane.

Example Consider

$$f(x, y) = \begin{cases} x + y & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

The marginal density of X is given by

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y)\, dy \\ &= \int_0^1 x + y\, dy \\ &= x + 1/2 \end{aligned}$$

Marginal Distributions RVs

Q5. Suppose (X, Y) takes values on the unit square [0, 1] × [0, 1] with joint pdf $f(x, y) = \frac{3}{2}(x^2 + y^2)$.
Find the marginal pdf $f_X(x)$ and use it to find P(X < 0.5).

$$f_X(x) = \int_0^1 \frac{3}{2}(x^2 + y^2)\, dy = \left[\frac{3}{2}x^2 y + \frac{y^3}{2}\right]_0^1 = \boxed{\frac{3}{2}x^2 + \frac{1}{2}}.$$

$$P(X < 0.5) = \int_0^{0.5} f_X(x)\, dx = \int_0^{0.5} \frac{3}{2}x^2 + \frac{1}{2}\, dx = \left[\frac{1}{2}x^3 + \frac{1}{2}x\right]_0^{0.5} = \boxed{\frac{5}{16}}.$$

Independence in RVs

- Events A and B are independent if P(A ∩ B) = P(A)P(B).
- The joint distribution (or density or mass) of Independent RVs is the product of the marginals.

Definition: Jointly-distributed random variables X and Y are independent if their joint cdf is the product of the marginal cdf's:

$$F(X,Y) = F_X(x)F_Y(y).$$

For discrete variables this is equivalent to the joint pmf being the product of the marginal pmf's.:

$$p(x_i, y_j) = p_X(x_i)p_Y(y_j).$$

For continous variables this is equivalent to the joint pdf being the product of the marginal pdf's.:

$$f(x, y) = f_X(x)f_Y(y).$$

Independence in RVs

Example 12. For discrete variables independence means the probability in a cell must be the product of the marginal probabilities of its row and column. In the first table below this is true: every marginal probability is $1/6$ and every cell contains $1/36$, i.e. the product of the marginals. Therefore X and Y are independent.

$X\backslash Y$	1	2	3	4	5	6	$p(x_i)$
1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
3	1/36	1/36	1/36	1/36	1/36	1/36	1/6
4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
5	1/36	1/36	1/36	1/36	1/36	1/36	1/6
6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
$p(y_j)$	1/6	1/6	1/6	1/6	1/6	1/6	1

Example 13. For continuous variables independence means you can factor the joint pdf or cdf as the product of a function of x and a function of y.

(i) Suppose X has range $[0, 1/2]$, Y has range $[0, 1]$ and $f(x, y) = 96x^2y^3$ then X and Y are independent. The marginal densities are $f_X(x) = 24x^2$ and $f_Y(y) = 4y^3$.

(ii) If $f(x, y) = 1.5(x^2+y^2)$ over the unit square then X and Y are not independent because there is no way to factor $f(x, y)$ into a product $f_X(x)f_Y(y)$.

(iii) If $F(x, y) = \frac{1}{2}(x^3y + xy^3)$ over the unit square then X and Y are not independent because the cdf does not factor into a product $F_X(x)F_Y(y)$.