

Probability and stats mop-up unit

- ▶ Jointly distributed random variables (spring-break eve)
- ▶ Understanding covariance and correlation (**Today**)
- ▶ Convergence of random variables (Wednesday)
- ▶ The weak law of large numbers (Friday)
- ▶ The central limit theorem (next week Monday)
- ▶ Review for test (next week Wednesday)
- ▶ Test 2 (next week Friday)

Today:

- ▶ Recap of jointly distributed random variables from last time
- ▶ The intuition of correlation from data
- ▶ Deriving and explaining covariance
- ▶ Proofs
- ▶ Deriving and explaining correlation

Let X and Y be random variables. Then the **joint cumulative distribution function** of X and Y is

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$$

If X and Y are discrete, then their **joint probability mass function** is

$$p_{X,Y}(x,y) = P(X = x, Y = y)$$

Moreover,

$$F_{X,Y}(x,y) = \sum_{i \leq x} \sum_{j \leq y} p_{X,Y}(i,j)$$

Let X and Y be random variables. Then the **joint cumulative distribution function** of X and Y is

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$$

If X and Y are continuous, then their **joint probability density function** is $f_{X,Y}$ such that

$$F_{X,Y}(a,b) = \int_{-\infty}^b \int_{-\infty}^a f(x,y) dx dy$$

Moreover,

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

and

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f_{X,Y}(x,y) dx dy$$

Let Ω be a sample space and let A and B be events over that sample space. Then $P(A \cap B)$ is the **joint probability** of events A and B .

Suppose A_0, A_1, \dots, A_{n-1} and B_0, B_1, \dots, B_{m-1} are each partitions of sample space Ω . Then the joint probabilities $P(A_i \cap B_j)$ have the property that

$$\sum_{i=0}^{n-1} \sum_{j=0}^{m-1} P(A_i \cap B_j) = 1$$

In this context, the probability of an individual events, for example $P(A_i)$, is called a **marginal probability**, and can be computed by summing over the joint probabilities with the other partition:

$$P(A_i) = \sum_{j=0}^{m-1} P(A_i \cap B_j)$$

Let X and Y be jointly distributed random variables. The **marginal cumulative distribution function** of X is

$$F_X(a) = F_{X,Y}(a, \infty) = P(X \leq a)$$

If X and Y are discrete, the **marginal probability mass function** of X is

$$p_X(x) = \sum_y p_{X,Y}(x, y)$$

If X and Y are continuous, the **marginal probability density function** of X is

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

Let A and B be events over a sample space. The **conditional probability** of A given B is

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

(This is undefined if $P(B) = 0$.)

Moreover, A and B are **independent** if the following (equivalent) are true:

$$\begin{aligned}P(A | B) &= P(A) \\P(B | A) &= P(B) \\P(A \cap B) &= P(A)P(B)\end{aligned}$$

Let X and Y be jointly distributed random variables. Then, if X and Y are discrete, the **conditional probability mass function** of X , given that $Y = y$, is

$$p_{X|Y}(x|y) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

If X and Y are continuous, the **conditional probability density function** of X , given that $Y = y$, is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Let X and Y be jointly distributed random variables. Then X and Y are **independent** if for all $A, B \subseteq \mathbb{R}$,

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

Or, equivalently, for all $a, b \in \mathbb{R}$,

$$P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b)$$

Or, again equivalently, with the same a, b ,

$$F_{X,Y}(a, b) = F_X(a)F_Y(b)$$

If X and Y are discrete, this is equivalent to, for all x, y

$$p_{X,Y}(x, y) = p_X(x)p_Y(y)$$

If X and Y are continuous, this is equivalent to, for all x, y

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

Let X and Y be jointly distributed random variables and let μ_X and μ_Y be their marginal means. The **covariance** of X and Y is

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

Compare this with the **variance** of random variable X is

$$\text{Var}(X) = E[(X - \mu)^2]$$

If X and Y are independent, then $\text{Cov}(X, Y) = 0$.

Prior Result 1

If X is a continuous random variable, then $E[aX + b] = aE[X] + b$.

Lemma 2

If X and Y are independent, continuous random variables, then $E[XY] = E[X]E[Y]$.

Theorem 3

If X and Y are continuous random variables, then

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

Corollary 4

If X and Y are independent, continuous random variables, then $\text{Cov}(X, Y) = 0$.

Recall that the **variance** of random variable X is

$$\text{Var}(X) = E[(X - \mu)^2]$$

Let X and Y be jointly distributed random variables and let μ_X and μ_Y be their marginal means and σ_X and σ_Y be their marginal standard deviations (square roots of their variances) The **covariance** of X and Y is

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

The **correlation coefficient** of X and Y is

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

If X and Y are independent, then $\text{Cov}(X, Y) = 0$, and hence so is $\rho(X, Y)$.

For next time:

Do the correlation exercise (Jupyter notebook) (due Fri, Mar 20)