Section 2.1: Joint, Marginal, and Conditional Distributions

Consider random variables $X_1, X_2, \ldots, X_n$.

We want to study these RVs together.
It is more convenient to use vector notations:

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

Joint distributions (discrete case)

When all variables are discrete, their joint distribution is described by the joint probability function

$$p(\boldsymbol{x}) = p(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n).$$

To be a valid PF, $p(\boldsymbol{x})$ must satisfy

$$p(\boldsymbol{x}) \geq 0 \quad \text{for all } \boldsymbol{x} \text{ and } \sum_{\boldsymbol{x}} p(\boldsymbol{x}) = 1.$$

Joint distributions (continuous case)

When all variables are continuous, their joint distribution is described by the joint density function

$$f(\boldsymbol{x}) = f(x_1, x_2, \ldots, x_n)$$

with the property that

$$P(\boldsymbol{X} \in A) = \int_A f(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n$$

integrating over A

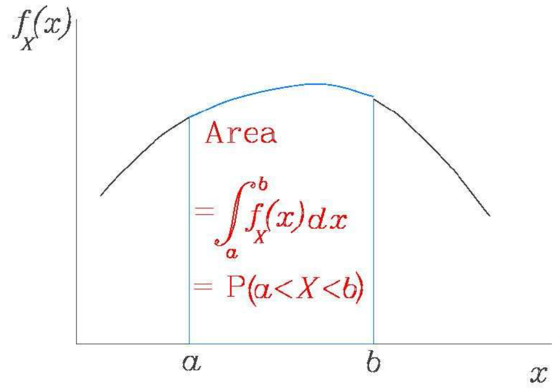
for any measureable set A.

To be a valid PDF, $f(\boldsymbol{x})$ must satisfy $\boxed{f(\boldsymbol{x}) \geq 0}$ for all $\boldsymbol{x}$ and

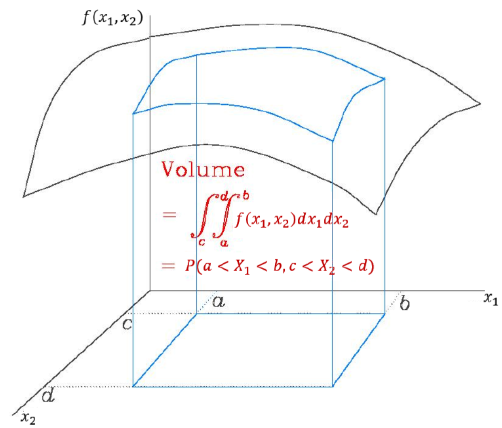
$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n = 1.$$

A note about $P(\boldsymbol{X} \in A)$:

- In the case of a univariate continuous RV X, $\int_a^b f(x)dx = P(a < X < b)$ is the area under the curve $f(x)$ between $x = a$ and $x = b$.

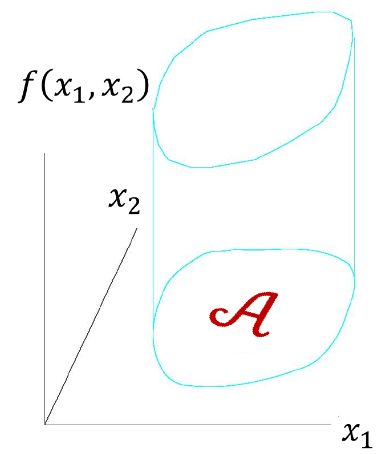


- In the case of bivariate continuous RV $\boldsymbol{X} = (X_1, X_2)^\top$, $\int_a^b \int_c^d f(x_1, x_2)dx_1dx_2 = P(a < X_1 < b, c < X_2 < d)$ corresponds to the volume between the surface $f(x_1, x_2)$ and the x_1x_2-plan bounded by the planes $x_1 = a, x_1 = b, x_2 = x$, and $x_2 = d$.



Joint distributions (continuous case)

- But the volume under a PDF does not necessarily need to sit on a rectangle aligned with the x_1 and x_2 axes.
- In general, A is an arbitrarily-shaped area in the $x_1 x_2$-plane:



- Hence we write $P(\boldsymbol{X} \in A) = \int_A f(x_1, x_2) dx_1 dx_2$.

For RVs of all types, the joint CDF is defined by

$$F(\boldsymbol{x}) = F(x_1, x_2, \ldots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n)$$

$$= F\left(\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \le \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right)$$

In the discrete case, this can be written as

$$P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n) = \sum_{y_1 \le x_1} \sum_{y_2 \le x_2} \cdots \sum_{y_n \le x_n} p_{\boldsymbol{X}}(y_1, y_2, \ldots, y_n).$$

In the continuous case, this can be written as

$$P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f_{\boldsymbol{X}}(y_1, y_2, \ldots, y_n) dy_1 dy_2 \ldots dy_n.$$

Example 2.1 (Discrete joint PF)

An urn contains one red, two white, and three blue balls. Two balls are withdrawn at random without replacement. Describe the joint probability function of the number of red and the number of blue balls drawn.

Let R be the number of red balls

Let B be the number of blue balls

We want $P\left(\begin{bmatrix} R \\ B \end{bmatrix}\right)$

There are $C^{6}_{2} = 15$ possible outcomes.

total six balls

two balls are withdrawn

1 Red	2 White	3 blue
R	W_1 W_2	B_1 B_2 B_3

$(RW_1)(RW_2)(RB_1)\ldots$

Probability of each $= \frac{1}{15}$

$P(R=0, B=0) = P(W_1, W_2) = \frac{1}{15}$

$P(R=1, B=0) = P(R, W_1) + P(R, W_2) = \frac{2}{15}$

$\vdots$

$P(R=1, B=2) = 0$

B \ R	0	1
0	$\frac{1}{15}$	$\frac{2}{15}$
1	$\frac{6}{15}$	$\frac{3}{15}$
2	$\frac{3}{15}$	0

$F(r,b) = P(R \le r, B \le b)$ for $r = 0, 1$ & $b = 0, 1, 2$

B \ R	$r<0$	$0 \le r < 1$	$0 < r \le 1$	$r > 1$
$b<0$	0	0	0	
$0 \le b < 1$	0			
$0 \le b < 2$	0			
$0 \le b < 3$	0			
$b > 3$				

$F(3,-2) = P(X_1 \le 3, X_2 \le -2)$
$= P(X_1 = 1, X_2 = -2) + P(X_1 = 2, X_2 = -2)$

$F(1,0) = P(R \le 1, B = 0) = P(R=0, B=0) + P(R=1, B=0)$
$= \frac{1}{15} + \frac{2}{15} = \frac{3}{15}$

Exercise: Describe the joint CDF of the above bivariate RV.

Example 2.1 Solutions

Example 2.1 Solutions (continue)

Example 2.2 (Continuous joint PF)

A random vector (X_1, X_2) has joint PDF

$$f(X_1, X_2) = \frac{1}{5}(x_1 + 2x_2), \quad 0 < x_1 < 1, 0 < x_2 < 2$$

(a) Find $P(0.5 < X_1 < 1, 0 < X_2 < 1)$

(b) Find the joint CDF of (X_1, X_2).

(a) $P(0.5 < X_1 < 1, 0 < X_2 < 1)$

$= \int_{0.5}^{1} \int_{0}^{1} f(x_1, x_2)\, dx_2\, dx_1$

$= \frac{1}{5} \int_{0.5}^{1} \int_{0}^{1} x_1 + 2x_2\, dx_2\, dx_1$

$= \frac{1}{5} \int_{0.5}^{1} \left[x_1 x_2 + x_2^2 \right]_0^1 dx_1 = \frac{1}{5} \int_{0.5}^{1} x_1 + 1\, dx_1 = \frac{1}{5}\left[\frac{1}{2}x_1^2 + x_1 \right]_{0.5}^{1} = \frac{1}{5}\left(\frac{1}{2} + 1 - \left(\frac{1}{2}\right)^3 + \frac{1}{2} \right) = \frac{7}{40}$

Exercise: Show that the volume under the joint PDF is 1.

Example 2.2 Solutions

Example 2.2 Solutions (continue)

Example 2.2 Solutions (continue)

Example 2.2 Solutions (continue)

Example 2.2 Solutions (continue)

If we have a joint PF (or PDF) of $\boldsymbol{X} = (\boldsymbol{X}_1^\top, \boldsymbol{X}_2^\top)^\top$, where $\boldsymbol{X}_1 = (X_1, X_2, \ldots, X_r)^\top$ and $\boldsymbol{X}_2 = (X_{r+1}, X_{r+2}, \ldots, X_n)^\top$, we can recover the PF (or PDF) for the RVs $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$. These are called marginal distributions.

- In the discrete case

$$p(\boldsymbol{x}_1) = \sum_{\boldsymbol{x}_2} p(\boldsymbol{x}_1, \boldsymbol{x}_2)$$

$$= \sum_{x_{r+1}} \sum_{x_{r+2}} \cdots \sum_{x_n} p(x_1, x_2)$$

- In the continuous case

$$f(\boldsymbol{x}_1) = \int_0^\infty \int_0^\infty \ldots \int_0^\infty f(x_1, x_2, \ldots x_n)\, dx_{r+1}\, dx_{r+2}\, \ldots\, dx_n$$

Example 2.3 (Marginal PF)

Consider the urn example again. Find the marginal PF for each of the number of red balls and the number of blue balls.

		R = 0	R = 1	Sum
B	0	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$
	1	$\frac{6}{15}$	$\frac{3}{15}$	$\frac{9}{15}$
	2	$\frac{3}{15}$	0	$\frac{3}{15}$
Sum		$\frac{10}{15}$	$\frac{5}{15}$	$\frac{15}{15} = 1$

$$p(b) = \begin{cases} \frac{3}{15} & \text{if } b = 0 \\ \frac{9}{15} & \text{if } b = 1 \\ \frac{3}{15} & \text{if } b = 2 \\ 0 & \text{otherwise} \end{cases}$$

$$p(r) = \begin{cases} \frac{9}{15} & \text{if } r = 0 \\ \frac{5}{15} & \text{if } r = 1 \\ 0 & \text{otherwise} \end{cases}$$

Example 2.4 (Marginal PDF)

Consider again the random vector (X_1, X_2) with joint PDF

$$f(X_1, X_2) = \frac{1}{5}(x_1 + 2x_2), \quad 0 < x_1 < 1, 0 < x_2 < 2$$

Find the marginal PDF of X_1 and X_2.

$$f(x_1) = \int_0^2 f(x_1, x_2)\, dx_2$$

$$= \frac{1}{5}\int_0^2 (x_1 + 2x_2)\, dx_2$$

$$= \frac{1}{5}\left[x_1 x_2 + x_2^2\right]_0^2$$

$$= \frac{1}{5}(2x_1 + 4) \text{ for } 0 < x_1 < 1$$

We can also derive the PF (or PDF) for $\boldsymbol{X}_1$ when the other variables $\boldsymbol{X}_2$ has a particular value x_2. This is called a conditional distribution.

- In the discrete case

$$p(\boldsymbol{x}_1|\boldsymbol{x}_2) = \frac{p(\boldsymbol{x}_1, \boldsymbol{x}_2)}{p(\boldsymbol{x}_2)}$$

provided that $p(\boldsymbol{x}_2) > 0$.

- In the continuous case

$$f(\boldsymbol{x}_1|\boldsymbol{x}_2) = \frac{f(\boldsymbol{x}_1, \boldsymbol{x}_2)}{f(\boldsymbol{x}_2)}$$

provided that $f(\boldsymbol{x}_2) > 0$.

Remarks about conditional distributions

The conditional density does not, strictly speaking, produce conditional probabilities. For example,

$$\int_{\boldsymbol{a}}^{\boldsymbol{b}} f(\boldsymbol{x}_1|\boldsymbol{x}_2)d\boldsymbol{x}_1$$

is not

$$P(\boldsymbol{a} < \boldsymbol{X}_1 < \boldsymbol{b}|\boldsymbol{X}_2 = \boldsymbol{x}_2)$$

because

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

is defined only for $P(B > 0)$.

But, for a continuous RV, $P(\boldsymbol{X}_1 = \boldsymbol{x}_1) = 0$. However, if $f(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is a continuous function of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, it can be checked that

$$\int_{\boldsymbol{a}}^{\boldsymbol{b}} f(\boldsymbol{x}_1|\boldsymbol{x}_2)d\boldsymbol{x}_1 = \lim_{\epsilon \to 0} P(\boldsymbol{a} < \boldsymbol{X}_1 < \boldsymbol{b}|\boldsymbol{x}_2 - \epsilon < \boldsymbol{X}_2 < \boldsymbol{x}_2 + \epsilon)$$

Example 2.5 (Conditional PF)

Consider the urn example again. Find the conditional PF for the number of red balls given that there is one blue ball.

$$P(R|B=1) = \begin{cases} \frac{2}{3} & \text{if } r=0 \\ \frac{1}{3} & \text{if } r=1 \\ 0 & \text{otherwise} \end{cases}$$

		R = 0	R = 1	sum
B	0	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$
	1	$\frac{6}{15}$	$\frac{3}{15}$	$\frac{9}{15}$
	2	$\frac{3}{15}$	0	$\frac{3}{15}$
	sum	$\frac{10}{15}$	$\frac{5}{15}$	$\frac{15}{15}$

Scale each row by the sum

		R = 0	R = 1
B	0	$\frac{1}{3}$	$\frac{2}{3}$
	1	$\frac{2}{3}$	$\frac{1}{3}$
	2	1	0

Example 2.6 (Conditional PDF)

Consider again the random vector (X_1, X_2) with joint PDF

$$f(X_1, X_2) = \frac{1}{5}(x_1 + 2x_2), \quad 0 < x_1 < 1, 0 < x_2 < 2$$

Find the conditional PDF of X_1 given X_2 and the conditional PDF of X_1 given $X_2 = 1$.

$$f(x_1 \mid x_2) = \frac{f(x_1, x_2)}{f(x_2)} = \frac{\cancel{\frac{1}{5}}(x_1 + 2x_2)}{\cancel{\frac{1}{5}}(\frac{1}{2} + 2x_2)} = \frac{x_1 + 2x_2}{\frac{1}{2} + 2x_2}$$

$$f(x_1 \mid x_2 = 1) = \frac{x_1 + 2(1)}{\frac{1}{2} + 2(1)} = \frac{x_1 + 2}{\frac{5}{2}} = \frac{2}{5}(x_1 + 2) \; \boxed{\text{for } 0 < x_1 < 1}$$

important to include the support.

Section 2.2: Independence, Covariance, Correlation

The random variables $X_1, X_2, \ldots, X_n$ are independent if and only if their joint PF (or PDF) satisfies the following condition:

- In the discrete case

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2)\ldots p(x_n)$$

for all $x_1, x_2, \ldots, x_n$.

- In the continuous case

$$f(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2)\ldots f(x_n)$$

for all $x_1, x_2, \ldots, x_n$.

Example 2.7 (Independence)

Consider the urn example again. Is the number of red balls selected independent of the number of blue balls selected?

$P(B=0, R=0) = \frac{1}{15}$

		R		
		0	1	sum
B	0	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$
	1	$\frac{6}{15}$	$\frac{3}{15}$	$\frac{9}{15}$
	2	$\frac{3}{15}$	0	$\frac{3}{15}$
	sum	$\frac{10}{15}$	$\frac{5}{15}$	$\frac{15}{15}$

$P(B=0)\,P(R=0)$

$= \frac{3}{15} \times \frac{5}{15} \neq P(B=0, R=0)$

$\therefore$ they are not independent

Example 2.8 (Independence)

Consider again the random vector (X_1, X_2) with joint PDF

$$f(X_1, X_2) = \frac{1}{5}(x_1 + 2x_2), \quad 0 < x_1 < 1, 0 < x_2 < 2$$

Are X_1 and X_2 independent?

Covariance and correlation

seeing if there is a linear relationship or not

Definition

If X_1 and X_2 are random variables with $E(X_i) = \mu_i$ and $\text{var}(X_i) = \sigma_i^2$ for $i = 1, 2$, then the covariance is

$$\text{cov}(X_1, X_2) = \sigma_{12} = E\left[(X_1 - \mu_1)(X_2 - \mu_2)\right]$$

and the correlation coefficient is

$$\text{cor}(X_1, X_2) = \rho_{12} = \frac{\sigma_{12}}{\sigma_1 \sigma_2}.$$

When $X_1 = X_2$, we have $\text{cov}(X_1, X_1) = \text{var}(X_1)$.

$\mathrm{cov}(3X_1, 4X_2) = E[3X_1 4X_2] - E(3X_1)E(2X_2)$

When computing the covariance, it is often easier to use this formula:

$$\boxed{\mathrm{cov}(X,Y) = E[XY] - E(X)E(Y)}$$

(Note that this reduces to the variance formula when $Y = X$).

Proof: $\mathrm{cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$

$= E[XY - E[X]Y - E(Y)X + E[X]E(Y)]$

$= E[XY] - E[X]E[Y] - \cancel{E[Y]E[X]} + \cancel{E[X]E[Y]}$

$= E[XY] - E(X)E(Y)$

Covariance, correlation and independence

Theorem 2.2.1 Correlation and independence

If X and Y are independent, then

$$\text{cov}(X, Y) = \text{cor}(X, Y) = 0.$$

But the converse is not true, i.e., $\text{cov}(X, Y) = \text{cor}(X, Y) = 0$ does not imply that X and Y are independent.

Proof: $\text{cov}(X, Y) = E[(X - E[X])(Y - E[X])]$

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

Example 2.9 (Covariance and correlation)

Consider the urn example again.

(a) What is the covariance of the number of red and the number of blue balls drawn? Why is the covariance negative?

negatively correlation

because the ~~cord~~ correlation between blue & red balls.

		R		
		0	1	
B	0	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{3}{15}$
	1	$\frac{6}{15}$	$\frac{3}{15}$	$\frac{9}{15}$
	2	$\frac{3}{15}$	0	$\frac{3}{15}$
		$\frac{10}{15}$	$\frac{5}{15}$	$\frac{15}{15}$

E(x, ~~X~~x2)) = 1.8

$$cov(B,R) = E[BR] - E[B]E[R]$$

Now $E[BR] = 1 \times \frac{3}{15} + 2 \times \frac{0}{15} = \frac{3}{15}$

(Note that the only cases of interest for $E[BR]$ are when $B, R \neq 0$. So we only need to look at $(B = 1, R = 1)$ and $(B = 2, R = 1)$.)

Also $E[B] = 1 \times \frac{9}{15} + 2 \times \frac{3}{15} = \frac{15}{15} = 1.$

$E[R] = 1 \times \frac{5}{15} = \frac{5}{15} = \frac{1}{3}$

So $cov(BR) = \frac{3}{15} - 1 \times \frac{1}{15} = -\frac{2}{15}.$

(b) What is the correlation of the number of red and the number of blue balls drawn? Why is the covariance negative?

By definition, $\rho_{BR} = cor(B,R) = \frac{\text{cov}(B,R)}{\sigma} = \frac{\sigma_{BR}^2}{\sigma_B \sigma_R} = \frac{-\frac{2}{15}}{\sqrt{\frac{6}{15}\left(\frac{2}{9}\right)}} = -\frac{1}{\sqrt{5}}$. (see below)

		R		
		0	1	**Sum**
B	0	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{8}{15}$
	1	$\frac{6}{15}$	$\frac{3}{15}$	$\frac{9}{15}$
	2	$\frac{3}{15}$	0	$\frac{3}{15}$
	Sum	$\frac{10}{15}$	$\frac{5}{15}$	

From part (a), we have $\text{cov}(B,R) = -\frac{2}{15}$.

{from part (a)}

Now $\sigma_B^2 = \text{var}(B)$

$$
\begin{aligned}
&= E(B^2) - E(B)^2 \\
&= \left(0^2 \times \frac{8}{15} + 1^2 \times \frac{9}{15} + 2^2 \times \frac{3}{15}\right) - 1^2 \\
&= \frac{21}{15} - 1 \\
&= \frac{6}{15}
\end{aligned}
$$

Similarly, $\sigma_R^2 = \frac{2}{9}$.

Example 2.10 (Covariance and correlation)

Consider the random vector (R, S) with joint PDF

$$f(r,s) = 4rs, \quad 0 < r < 1, 0 < s < 1$$

(a) Determine $\text{cov}(R, S)$ using the covariance formula.
(b) Determine $\text{cor}(R, S)$.

(a) $\text{cov}(R,S) = E[RS] - E[R]E[S]$

$E[RS] = \int_0^1 \int_0^1 f(r,s)\, rs \, ds\, dr$

$= \int_0^1 \int_0^1 rs \times 4rs \, ds\, dr$

$= \int_0^1 \int_0^1 r^2 s^2 \, ds\, dr$

$= \frac{4}{9}$

$E[R] = \int_0^1 r f(r)\, dr$

marginal distribution

$f(r) = \int_0^1 f(r,s)\, ds$

$= \int_0^1 4rs\, ds = [2rs^2]_0^1 = 2r$

$E[R] = \int_0^1 r\,2r\,dr = \int_0^1 2r^2 dr = [\frac{2}{3}r^3]_0^1 = \frac{2}{3}$

$f(s)$.

$\Rightarrow \text{cov}(R,S) = \frac{4}{9} - (\frac{2}{3})(\frac{2}{3}) = \frac{4}{9} - \frac{4}{9} = 0$

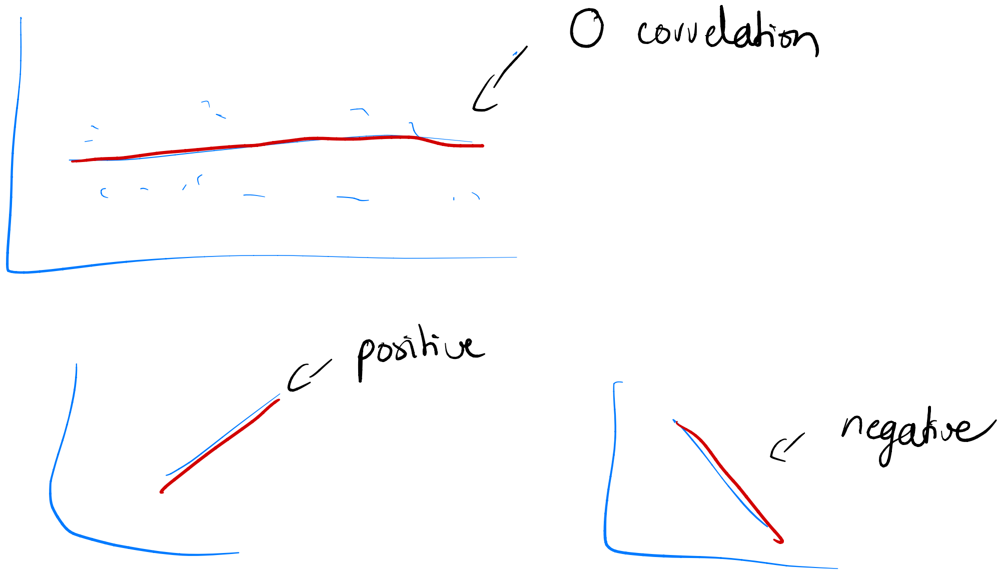
(b) If coveration = 0 $\Rightarrow$ correlation = 0

because $\text{cor}(X_1, X_2) = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$ $\leftarrow$ cov $= \sigma_{12}$

$\therefore$ cov $= \sigma_{12} = 0$

$\Rightarrow \text{cor}(X_1, X_2) = 0$

Example 2.10 (Solutions)



Example 2.11 (Counter-example)

Let $X_1 \sim U(-1,1)$. Define X_1 to be explicitly related to X_2 such that $X_2 = X_1^2$ (i.e. we know there is a relationship between X_1 and X_2, but the relationship is not linear). Show that $\text{cov}(X_1, X_2) = 0$.

Example 2.11 Solutions

Properties of variance, covariance, and correlation

Let X and Y be random variables and a and b be constants:

- $\text{cov}(X, Y) = \text{cov}(Y, X)$ and $\text{cor}(X, Y) = \text{cor}(Y, X)$
- $\text{var}(aX + b) = a^2\text{var}(X)$
- $SD(aX + b) = |a|SD(X)$
- $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y)$
- $\text{var}(aX + bY) = a^2\text{var}(X) + b^2\text{var}(Y) + 2ab\text{cov}(X, Y)$

Example 2.12 (Property of covariance)

Proof this useful result:
If X, Y, and Z are random variables, then

$$\text{cov}(X + Y, Z) = \text{cov}(X, Z) + \text{cov}(Y, Z)$$

$$\begin{aligned}
\text{cov}(X+Y, Z) &= E[(X+Y)Z] - E[X+Y]E[Z] \\
&= E[XZ + YZ] - (E[X] + E[Y])E[Z] \\
&= E[XZ] + E[YZ] - E[X]E[Z] - E[Y]E[Z] \\
&= \underbrace{(E[XZ] - E[X]E[Z])}_{\text{cov}(X,Z)} + \underbrace{(E[YZ] - E[Y]E[Z])}_{\text{cov}(Y,Z)} \\
&= \text{cov}(X,Z) + \text{cov}(Y,Z)
\end{aligned}$$

$\therefore$ proved

Section 2.3: Moments of multivariate RVs

Theorem 2.3.1 Multivariate moments

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a random vector and let $g(X_1, X_2, \ldots, X_n)$ be a real-valued function. Then the expected value $E[g(X_1, X_2, \ldots, X_n)]$ is given by

- in the discrete case: $E[g(x)] = \sum_x g(x)p(x)$ -> single (univariate)

$$\sum_{x_1}\sum_{x_2}\cdots\sum_{x_n} g(x_1, x_2, \ldots, x_n)\, p(x_1, x_2, \ldots, x_n)$$

- in the continuous case:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1, x_2, \ldots, x_n)\, f(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n$$

provided it exists.

Let $a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix}$, $g(x) = \begin{bmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_k(x) \end{bmatrix}$ then $E[\tilde{a}^T g(x)] = a^T E[g(x)]$

Linear combinations

Theorem 2.3.2 Expectation of linear combinations

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a random vector and let $g_1(\boldsymbol{X}), g_2(\boldsymbol{X}), \ldots, g_k(\boldsymbol{X})$ be real-valued functions and $a_1, a_2, \ldots, a_k$ be real constants. Then

$$\begin{aligned} & E[a_1 g_1(\boldsymbol{X}) + a_2 g_2(\boldsymbol{X}) + \ldots + a_k g_k(\boldsymbol{X})] \\ = \; & a_1 E[g_1(\boldsymbol{X})] + a_2 E[g_2(\boldsymbol{X})] + \ldots + a_k E[g_k(\boldsymbol{X})] \end{aligned}$$

univariate

$E[ag(x) + bg(x)] = aE[g(x)] + bE[g(x)]$

Corollary 1

$$E[a_1 X_1 + a_2 X_2 + \ldots + a_k X_k] = a_1 E(X_1) + a_2 E(X_2) + \ldots + a_k E(X_k)$$

This result does not require the assumption of independence.

Proof of Theorem 2.3.2

$$E[a_1 g_1(x) + a_2 g_2(x) + \cdots + a_k g_k(x)]$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}[a_1 g_1(x) + a_2 g_2(x) + \cdots + a_k g_k(x)] f(x) dx_1 dx_2 \cdots dx_k$$

$$= a_1 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g_1(x) f(x) dx_1 dx_2 \cdots dx_k$$

$$+ a_2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g_2(x) f(x) dx_1 dx_2 dx_3 \cdots dx_k$$

$$\vdots$$

$$+ a_k \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g_k(x) f(x) dx_1 dx_2 \cdots dx_k$$

$$= a_1 E[g_1(x)] + a_2 E[g_2(x)] + \cdots + a_k E[g_k(x)]$$

Theorem 2.3.3 Covariance of linear combinations

Let $X_1, X_2, \ldots, X_n$ be random variables such that

$$E(X_i) = \mu_i \quad \text{and} \quad \text{cov}(X_i, X_j) = \sigma_{ij}.$$

Let $a_1, a_2, \ldots, a_k$ and $b_1, b_2, \ldots, b_n$ be constants. Then

$$\text{cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{n} b_j X_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j \sigma_{ij}.$$

Let $a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$ and $b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$,

$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$, then $\text{cov}(a^T X, b^T X)$

$= a^T \text{cov}(X, X) b$

$= a^T \text{Var}(X) b$

Corollary 2

$$\text{var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \sigma_{ij}.$$

Furthermore, if $X_1, X_2, \ldots, X_n$ are independent, then

$$\text{var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \sigma_{ii}.$$

Proof of Theorem 2.3.3

$$\mathrm{cov}(a^T X, b^T X) = E[(a^T X - E(a^T X))(b^T X - E(b^T X))]$$

$$= E[(\underbrace{a^T X) - a^T E(X)}_{\text{common factor}})(b^T X - b^T E(X))]$$

$$= E[a^T(X - E(X))\, b^T(X - E(X))]$$

$$= E[a^T (X - E(X))(X - E(X))^T b]$$

$$= a^T \underbrace{E[(X - E(X))(X - E(X))^T]}_{Var(X)} b$$

$$= a^T Var(X) b$$

Corollary 3
For random variables X_1 and X_2, the correlation coefficient satisfies

$$|\rho_{12}| \leq 1$$

with equality if and only if X_1 satisfy a linear relationship with probability 1.

$0 \leq var(x_1 - t x_2) = \sigma_1^2 + t^2 \sigma_2^2 - 2\sigma_{12} t$

$p(t)$

Proof of Corollary 3 (continue)

Case 1: no real roots

$\Delta < 0$

$\iff 4\sigma_{12}^2 - 4\sigma_1\sigma_2 < 0$

$\iff \dfrac{\sigma_{12}^2}{\sigma_1^2\sigma_2^2} < 1$

$\iff \rho_{12}^2 < 1$

Case 2: 1 real roots

$\Delta = 0 \Rightarrow \rho_{12} = \pm 1$

$p(t_0) = 0 \quad var(X_1 - t_0 X_2) = 0$

$x_1 - t_0 x_2 = c$ without probability 1.

$\Rightarrow X_1 = t_0 X_2 + c$

Example 2.15 (Trivariate case)

Let X_1, X_2, and X_3 be RVs with respective means 3, 4, and 5, and respective variances 6, 7, and 8. Let their covariance be

$$\mathrm{cov}(X_1, X_2) = 0, \quad \mathrm{cov}(X_1, X_3) = -1, \quad \text{and} \quad \mathrm{cov}(X_2, X_3) = 2.$$

Consider the RV $Y = X_1 - 2X_2 + 3X_3$. Determine

(a) $E(Y)$

$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$, $a = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$, then $Y = a^T X$. $E[X] = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$

$\therefore E[Y] = E[a^T X]$
$= a^T E[X]$
$= [1 \ -2 \ 3]\begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$
$= 3 - 8 + 15 = 10$

(b) $\mathrm{var}(Y)$

$\mathrm{var}(X) = \begin{bmatrix} 6 & 0 & -1 \\ 0 & 7 & 2 \\ -1 & 2 & 8 \end{bmatrix}$

$\therefore \mathrm{var}(Y) = \mathrm{var}(a^T X)$
$= a^T \mathrm{var}(X) a = [1 \ -2 \ 3] \ \mathrm{var}(X) \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$
$= 76$

Conditional expectations are simply expectations computed from a conditional distribution.

- in the discrete case:

$$E\left[g(\boldsymbol{X})|\boldsymbol{y}\right] = \sum_{\boldsymbol{x}} g(\boldsymbol{x})\, p(\boldsymbol{x}|\boldsymbol{y})$$

- in the continuous case:

$$E\left[g(\boldsymbol{X})|\boldsymbol{y}\right] = \int_{-\infty}^{\infty} g(\boldsymbol{x})\, f(\boldsymbol{x}|\boldsymbol{y}) d\boldsymbol{x}$$

The expected value of the random variable $E(Y|X)$ is equal to the expected value of Y, that is

law of total expectation

$$E_X\left[E(Y|X)\right] = E_Y(Y)$$

Proof: $E_X[E(Y|X)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(y|x)\,dy\, f(x)\,dx$

$$= \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} \underbrace{f(y|x) f(x)}_{f(x,y)}\,dy\,dx$$

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

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

Example 2.16 (Conditional expectation)

An insect lays Y number of eggs, where Y has a Poisson distribution with parameter λ. If the probability of each egg surviving is p, then what is the expected number of surviving eggs?

Let Y = number of eggs laid

X = number of eggs survived

Then $Y \sim Pois(\lambda)$ and $X|Y=y \sim B(y,p)$

$$E_X(X) = E_Y[E(X|Y)]$$
$$= E_Y[Yp] \quad \text{as } X|Y \sim B(y,p)$$
$$= pE_Y[Y] \quad E(X) = \lambda \text{ as } Y \sim Pois(\lambda)$$
$$= p\lambda$$

Let X = # of loss

Y = # of accident

$X \sim Exp(1.25)$

$X|Y=y \sim B(y, 0.25)$

Example 2.17 (Conditional expectation)

Consider two RVs X and Y where

$$f(x|y) = \frac{x+y}{\frac{1}{2}+y} \quad \text{and} \quad f(y) = y + \frac{1}{2} \quad \text{for } 0 < x < 1, 0 < y < 1.$$

Find $E\left[Y|X = \frac{1}{3}\right]$.

First: find $f(y|x)$ then: calculate $E[y|x]$

$$f(y|x) = \frac{f(x,y)}{f(x)} = \frac{f(x|y)f(y)}{f(x)} = \frac{f(x|y)f(y)}{\int_0^1 f(x,y)\,dy} = \frac{\left(\frac{x+y}{\cancel{\frac{1}{2}+y}}\right)\cancel{\left(y+\frac{1}{2}\right)}}{\int_0^1 (x+y)\,dy} = \frac{x+y}{\left[xy+\frac{1}{2}y^2\right]_0^1} = \frac{x+y}{x+\frac{1}{2}}$$

$$\begin{aligned} E[Y|X=\tfrac{1}{3}] &= \int_0^1 y f(y|x=\tfrac{1}{3})\,dy \\ &= \int_0^1 y\left(\frac{\frac{1}{3}+y}{\frac{1}{3}+\frac{1}{2}}\right)dy \\ &= \frac{6}{5}\int_0^1 \left(\frac{1}{3}y+y^2\right)dy \\ &= \frac{6}{5}\left[\frac{1}{6}y^2+\frac{1}{3}y^3\right]_0^1 \\ &= \frac{6}{5}\left(\frac{1}{6}+\frac{1}{3}\right) \\ &= \frac{3}{5} \end{aligned}$$

Example 2.17 Solution

In a similar way, the conditional variance is simply the variance computed from a conditional distribution.

- In the discrete case:

$$\mathrm{var}\left(X|y\right) = \sum_{x}\left[x - E\left(x|y\right)\right]^{2}\ p(x|y)$$

- In the continuous case:

$$\mathrm{var}\left(X|y\right) = \int_{-\infty}^{\infty}\left[x - E\left(x|y\right)\right]^{2}\ f(x|y)d\boldsymbol{x}$$

(Conditional) variance formula

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

Law of total variance

$$\mathrm{var}(Y) = E\left[\mathrm{var}(Y|X)\right] + \mathrm{var}\left[E(Y|X)\right]$$

Example 2.18 (Conditional variance)

Let X and Y be continuous RVs with joint PDF

$$f(x, y) = e^{-y}, \quad \text{for } 0 < x < y < \infty$$

What is the conditional variance of Y given $X = x$?

Need 1: $f(x)$ 2: $f(y|x)$ 3: $E(Y|X)$ 4: $E(Y^2|X)$

Example 2.18 Solution

Definition

Let $X_1, X_2, \ldots X_n$ be RVs, then the joint MGF (provided it exists) is given by

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = M_{\boldsymbol{X}}(t_1, t_2, \ldots, t_n) = E\left[e^{t_1 X_1 + t_2 X_2 + \ldots + t_n X_n}\right] = E\left[e^{\boldsymbol{t}^\top \boldsymbol{X}}\right]$$

If $X_1, X_2, \ldots X_n$ are mutually independent, then the joint MGF satisfies

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = M_{X_1}(t_1) M_{X_2}(t_2) \ldots M_{X_n}(t_n)$$

provided it exists.

Definition

Let $X_1, X_2, \ldots X_n$ be RVs. The cross moment of order r is given by

$$\mu(r_1, r_2, \ldots, r_n) = E\left[X_1^{r_1} X_2^{r_2} \ldots X_n^{r_n}\right]$$

where $r_1, r_2, \ldots, r_n > 0$ and $r = \sum_{i=1}^{n} r_i$.

Cross moments and MGF

Let $\boldsymbol{X} = (X_1, X_2, \ldots X_n)^\top$ be an random vector that possess a joint MGF $M_{\boldsymbol{X}}(\boldsymbol{t})$. Then

$$\mu(r_1, r_2, \ldots, r_n) = \left.\frac{\partial^r \; M_{\boldsymbol{x}}(t_1, t_2, \ldots, t_n)}{\partial t_1^{r_1} \partial t_2^{r_2} \ldots \partial t_n^{r_n}}\right|_{t_1=0, t_2=0, \ldots, t_n=0}$$

where the derivative on the right-hand side is the rth order partial derivative of $M_{\boldsymbol{X}}(\boldsymbol{t})$ evaluated at the point $\boldsymbol{t} = \boldsymbol{0}$.

Example 2.19 (Cross moment)

The joint MGF of a bivariate standard normal random vector $\boldsymbol{X}$ is

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = e^{\frac{1}{2}\boldsymbol{t}^\top \boldsymbol{t}} = e^{\frac{1}{2}(t_1^2+t_2^2)}.$$

Find the second cross moment $E[X_1 X_2]$ of $\boldsymbol{X}$.