Lezione 6 –

Recap

↳ We introduced the Poisson distribution, which the limit of the binomial distribution under some condition.

As the n increase, p increase, while $np = \lambda = \text{const.}$

Poisson : $\overset{\text{is}}{X} \overset{\text{distributed}}{\sim} Poi(\lambda)$ with $\lambda > 0$, when:

$$p_X(x) = e^{-\lambda} \frac{\lambda^x}{x}, \quad x = 0, 1, 2, \ldots$$

$$S = \{ \underset{p_0,}{0}, \underset{p_1}{1}, \underset{p_2,}{2}, \underset{p_3}{3}, \ldots \}$$

$$p_j = \frac{e^{-\lambda} \lambda^j}{j!}$$

$$\underbrace{\sum_{j=0}^{\infty} \frac{e^{-\lambda} \lambda^j}{j!} = 1}$$

$$e^{-\lambda} \underbrace{\sum \frac{\lambda^j}{j!}}_{e^{\lambda}} = e^{-\lambda} e^{\lambda} = 1$$ ⇒ there is a variable that can be distributed like this.

Approximation of the Binomial with the Poisson density

Assume $X \sim Bin(\overset{\#\text{trial}}{n}, p) \Rightarrow X \overset{\text{Approx}}{\sim} Poi(\lambda)$ $\quad \lambda = np$

↳ success probability

↓ under these conditions

Consider n large, $\rightsquigarrow$ $n \geq 50$
good approximation of ∞ according to our theory

p small $\rightsquigarrow$ $np \leq 10$
np is constant

Exercise

Suppose 10% of solar panel produced by a factory are defective.

1) What is the probability that at least 95% of panel picked out will function correctly, with $n = 20$

X_{20} = n° of panel which function out of the 20 extracted.

$X_{20} \sim Bin(20, .9)$ → 1 - probability of being defect, because of how are defined "success".

"success" = "panel is ok" → Probability of "success"

$$P(X_{20} \geq 95\% \cdot 20) = P(X_{20} \geq 19) = \binom{20}{19} p^{19}(1-p) + \binom{20}{20} p^{20}\cancel{(1-p)^{0}}$$

$\simeq .3917$

2) What is the probability that at least 95% of out of $n = 100$ of the panels will function correctly.

exact probability = 0,00576 → using Binomial

Instead of focusing on the number of panels which function, let's focus on the defective ones.

Y_{100} = n° of panels ($n = 100$) that are defective

$Y_{100} \sim Bin(n = 100, \tilde{p} = 0{,}1)$

$\tilde{p}$ = "success" defined as probability its defective.

$\Rightarrow$ at most 5% will be defective, not 95% functioning.
The function is the same, since they are logically equivalent.

$$P(Y_{100} \leq 5\% \cdot 100) = P(Y_{100} \leq 5) = 0{,}00576$$

↳ Using R

$n = 100 \geq 50$
$np = 100 \cdot .1 = 10 \ [\leq 10]$ $\Rightarrow$ we can use Poisson.

$Y_{100} \sim Poi(\lambda = 10)$

$$\Rightarrow = e^{-\lambda}\left(1 + \frac{\lambda}{1} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \frac{\lambda^4}{4!} + \frac{\lambda^5}{5!}\right) = 0{,}0671$$

↳ The relative error is high because np was at the border of the condition.

What if we have to consider many random variables together, instead of only one?
When we have more random variables it isn't unusual for them to be related.

Joint Distribution of random variables

We call joint distribution of the random variables $(X \& Y)$ the value:

AND (logical)

$$P(X \in A, Y \in B) \qquad A, B \in F$$

↳ probability X belongs to A and Y belongs to B

The comma is equivalent to the intersect $\cap$

$$P(a \leq X \leq b, c \leq Y \leq b) = P(\{a < X \leq b\} \cap \{c < Y \leq d\})$$
$$\overset{!}{=} P(\{a < X \leq b\} \text{ AND } \{c < Y \leq d\})$$

In general its difficult to assign couples of random variable.

Two random variables X, Y are couples of a.c.r.v with density $f_{X,Y}$ if there exists $f_{X,Y}: \mathbb{R}^2 \to \mathbb{R}$, where $f_{X,Y} \geq 0$

joint

and $\int_{\mathbb{R}^2} f_{X,Y}(x,y) = 1$, such that:

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

$\hookrightarrow$ Not what we focus on, we focus more on discrete functions.

How do we assign a discrete couple of r.v.

If both X and Y are d.r.v, we need to assign the joint density $d(X,Y)$, that is

$$p_{X,Y}(x,y) = \begin{cases} = P(X = x, Y = y) & \text{for } x,y \in S \text{ support of } (X,Y) \\ = 0 & \text{for } x,y \notin S \end{cases}$$

$\hookrightarrow$

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

Marginal Densities

↳ The density X (or Y) is called marginal density of X (or Y), is computed as:

- when X, Y is a discrete couple:

$$p_X(x) = P(X=x) = P(X=x, y \in \mathbb{R})$$

- where X, Y are a couple of a.c.r.v

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x,y)\, dy \quad \forall x \in \mathbb{R}$$

This can be generalized to n variables.

We will focus on n=2 discrete random variables.

Table of Joint Distribution of a couple (X, Y) of discrete random variables.

X has values $a_1 < a_2 < a_3 < \dots < a_k$

Y " $b_1 < b_2 < b_3 < \dots < b_m$

$$S = \{(a_j, b_l), j=1, \dots, k, l=1, \dots, m)$$

$p_{X,Y}(a_j, b_l)$ → if k and m are not very large we can use a table to write them we can add a colum of marginal density

$X \backslash Y$	b_1	b_2	$\dots$	$\dots$	b_m	p_X
a_1	$p_{X,Y}(a_1,b_1)$	$p_{X,Y}(a_1,b_2)$			$p_{X,Y}(a_1,b_m)$	
a_2	$p_{X,Y}(a_2,b_1)$					
$\dots$						
a_n					$p_{X,Y}(a_n,b_m)$	
p_Y	$p_Y(b_1)$				$p_Y(b_m)$	1

$p_X(a_1) = p_{X,Y}(a_1,b_1) + \dots + p_{X,Y}(a_1,b_m)$

↳ marginal densities of Y

Exercise:

(X,Y) $\quad p_{X,Y}(1,1) = \frac{1}{6}, \; p_{X,Y}(1,2) = \frac{2}{9} \; p_{X,Y}(2,1) = \frac{5}{18}$

$p_{X,Y}(2,2) = \frac{1}{3}$

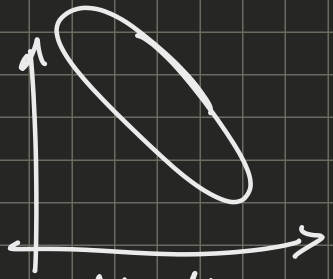
$X \backslash Y$	1	2	p_X
1	$1/6$	$2/9$	$7/18$
2	$5/18$	$1/3$	$11/18$
p_Y	$8/18$	$10/18$	1

We need to calculate all if we have no checked if the sum is equal to 1.

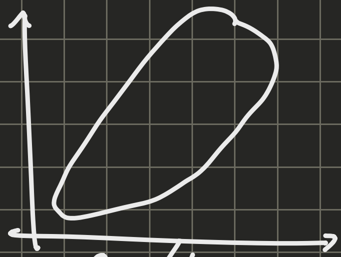
Homework compute $E(X)$, $Var(X)$

$E(Y)$, $Var(Y)$

Many variables will have a codependence, meaning that one tends to.



Negative association (covariance)



Positive association (covariance

Covariance between X and Y

Let X, Y be two r.v.s., the covariance between X and Y is a real number defined as:

$$Cov(X,Y) = E\left[(X - E(X))(Y - E(Y))\right] = E(X \times Y) - (EX)(EY)$$

Discrete

$$\sum_{(x,y)\in S} (x_i - EX)(y_i - EY)\, p_{X,Y}(x_i, y_i)$$

Abs. Cont.

$$\int_{\mathbb{R}^2} (x - EX)(y - EY) f_{X,Y}(x,y)\, dx\, dy$$

Remark

Let (X, Y) be r.v., and a new random variable

$g(X,Y)$ = deterministic transformation of (X, Y), e.g. $\frac{X}{Y}, X^2 \cdot Y, \ldots,$

$$\Rightarrow E\left[g(X,Y)\right] \xrightarrow{\text{Discrete}} \sum_{(x_i, y_j)\in S} g(x_i, y_j)\, p_X(x_i, y_j)$$

$$E\left[g(X,Y)\right] \xrightarrow{\text{Abs. Cont.}} \int_{\mathbb{R}^2} g(x,y) f_{X,Y}(x,y)\, dx\, dy$$

Properties:

$$cov(X,Y) = E\left[(X - EX)(Y - EY)\right] = E(X \cdot Y) - EX \cdot EY$$

↳ Proof:

$$\begin{aligned}
cov(X,Y) &= E[(X-EX)\cdot(Y-EY)] \\
&= E[XY - EX\cdot Y + X\cdot(-EY) + EX\cdot EY] \\
&= E[XY] + E[-EX\cdot Y] + E[X\cdot(-EY)] + E[EX\cdot EY] \\
&= E[XY] - \cancel{EX\cdot EY} - EY\cdot EX + \cancel{EX\cdot EY} \\
&= E[XY] - EX\cdot EY
\end{aligned}$$

↳ $E[EX] = EX$ expectation of a constant is a constant.

Example

X \ Y	1	2	3	px
1	.1	.2	.1	.4
2	.2	.1	.3	.6
pY	.3	.3	.4	1

$$P((X,Y)\in B) = \sum_{(x,y)\in B} p(x,y)$$

$$P(X+Y=3) = 0{,}2 + .2 = .4$$

$$(x,y)\in S \iff x+y=3 \iff (1,2),\ (2,1)$$

$$cov(X,Y) = E(X\cdot Y) - \underbrace{E(X)}_{1{,}6}\cdot\underbrace{E(Y)}_{2.1} = 3{,}4 - 1{,}6\cdot 2{,}1 = 0{,}04$$

$$E(X\cdot Y) = \sum_{(x,y)\in S}(x\cdot y)\,p_{XY}(x,y) = 1\cdot1\cdot.1 + 2\cdot1\cdot.2 + 3\cdot1\cdot.1 + 1\cdot2\cdot.2 + 2\cdot2\cdot.1 + 3\cdot2\cdot.3 = 3{,}4$$

Correlation between

$$\mathrm{Cor}(X,Y) = \rho_{X,Y} = \mathrm{cov}\Bigg(\underbrace{\frac{X-EX}{\sqrt{\mathrm{Var}X}}, \frac{Y-EY}{\sqrt{\mathrm{Var}Y}}}_{\text{Standardised versions of } X \text{ and } Y}\Bigg)$$

↳ norm cov

$$= E\left[\frac{X-EX}{\sqrt{\mathrm{Var}X}} \cdot \frac{Y-EY}{\sqrt{\mathrm{Var}Y}}\right] - \cancel{E\left(\frac{X-EX}{\sqrt{\mathrm{Var}X}}\right)}^{0} \cdot \cancel{E\left[\frac{Y-EY}{\sqrt{\mathrm{Var}Y}}\right]}^{0}$$

Definition of standardised variables.
$E=0$
$\mathrm{Var}=1$

$$= \frac{1}{\sqrt{\mathrm{Var}X}\sqrt{\mathrm{Var}Y}} E\big[(X-EX)\cdot(Y-EY)\big]$$

$$= \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{Var}X \cdot \mathrm{Var}Y}}$$

$\Rightarrow -1 \leq \rho_{x,y} \leq 1$ $\quad |\rho_{x,y}| = 1 \iff \exists\, a,b \in \mathbb{R}$ such that $a \neq 0$ $\quad P(Y = aX+b) = 1$

if $\rho_{x,y} = 1 \Rightarrow a > 0$

$\rho_{x,y} = -1 \Rightarrow a < 0$

The correllation measures the linear correllation between X and Y.