Slides
Probability Theory
Chapter 5

Ch. 5:

- **Joint and conditional distributions.**
- **Correlation and independence**
- **Distribution of the sum $X + Y$.**

Definition and properties of Expectation and Variance

Discrete

$$\mu = E(X) = \sum_x xP(X = x)$$

$$\sigma^2 = var(X) = E(X - \mu)^2$$

and $\sigma = \sqrt{var(X)}$

Formula for calculations:

$$var(X) = E(X^2) - \mu^2$$

$$E(aX + b) = aE(X) + b$$

$$var(aX + b) = a^2 var(X)$$

$$\sigma_{aX+b} = |a|\sigma_X$$

If the distribution is symmetrical about $x = c$, **then** $E(X) = c$

$var(X) \geq 0$ and $\sigma_X \geq 0$

2 rolls with a dice: **joint distribution** of X = "# one`s" and $Y$ = "# six`s"

Probability function
$\boldsymbol{P(X = x \text{ and } Y = y)}$

y / x	**0**	**1**	**2**	$\boldsymbol{P(X = x)}$
0	16/36	8/36	1/36	**25/36**
1	8/36	2/36	0	**10/36**
2	1/36	0	0	**1/36**
$\boldsymbol{P(Y = y)}$	**25/36**	**10/36**	**1/36**	**1**

The **marginal** distributions: $P(X = x)$ and $P(Y = y)$

It follows that $\boldsymbol{E(X) = E(Y) = \frac{1}{3}}$

And after computing $E(X^2)$ **we find:**

$$\boldsymbol{var(X) = var(Y)} = E(X^2) - (EX)^2 = \boldsymbol{\frac{10}{36}}$$

The **joint probability function**

of 2 discrete r.v. X and Y is given by:

$\boldsymbol{P}(\boldsymbol{X} = \boldsymbol{x} \text{ and } \boldsymbol{Y} = \boldsymbol{y})$, with (x, y) in $S_X \times S_Y$

(defined by a in table or a formula)

Requirements:

1. $P(X = x \text{ and } Y = y) \geq 0$

2. $\sum_{(x,y) \text{ in } S_X \times S_Y} P(X = x \text{ and } Y = y)$

Marginal probability functions:

$$P(X = x) = \sum_{y \text{ in } S_Y} P(X = x \text{ and } Y = y)$$

and

$$P(Y = y) = \sum_{\boldsymbol{x \text{ in } S_X}} P(X = x \text{ and } Y = y)$$

Distribution of functions of X and Y, such as $X + Y$

For example: $X + Y = 2$

$\boldsymbol{P(X + Y = 2)}$

$= \mathbf{1/36 + 2/36 + 1/36}$

$= \mathbf{4/36}$

x \ y	0	1	2
0	16/36	8/36	**1/36**
1	8/36	**2/36**	0
2	**1/36**	0	0

$Z = X + Y$ has a distribution

$\boldsymbol{z}$	**0**	**1**	**2**	Total
$\boldsymbol{P(Z = z)}$	**16/36**	**16/36**	**4/36**	1

$E(X + Y) = E(Z) = \frac{2}{3}$

Note that: $\boldsymbol{E(X + Y) = E(X) + E(Y)} = ⅓ + ⅓$

But: $\frac{\mathbf{16}}{\mathbf{36}} = \boldsymbol{var(X + Y) \neq var(X) + var(Y) = 2 \times \frac{10}{36}}$

Independence of discrete r.v.`s X and Y

- Recall the independence of events:

$$\boldsymbol{P(A \cap B) = P(A) \times P(B)}$$

- r.v.`s X and Y are independent if all events $A = \{X = x\}$ and $B = \{Y = y\}$ are independent.

- Definition:

  X and Y are **independent (ind.)** if

$$\boldsymbol{P(X = x \text{ and } Y = y) = P(X = x) \cdot P(Y = y)}$$

  for all x in S_X and y in S_Y.

- **Dependence** is simply proven by showing that the equality does not hold for one specific pair (x, y).

- **Independence** is usually **assumed**
  - e.g. if experiments are obviously independent.

Conditional distributions

E.g. the number of 1`s given the number of 6`s:

Suppose we have no 6 in two rolls: $Y = 0$,

$$\boldsymbol{P(X = 0|Y = 0) = \frac{P(X = 0 \text{ and } Y = 0)}{P(Y = 0)} = \frac{16/36}{25/26} = \frac{16}{25}}$$

The **conditional distribution of** X, given $Y = 0$:

$\boldsymbol{x}$	**0**	**1**	**2**	***Total***
$\boldsymbol{P(X = x\|Y = 0)}$	**16/25**	**8/25**	**1/25**	***1***

Conditional expectation of X, given $Y = 0$:

$$\boldsymbol{E(X|Y = 0) = \sum_x x \cdot P(X = x|Y = 0)}$$

$$\boldsymbol{= 0 \cdot \frac{16}{25} + 1 \cdot \frac{8}{25} + 2 \cdot \frac{1}{25} = \frac{2}{5}}$$

The **relation** of (unconditional) expectation $\boldsymbol{E(X)}$ and conditional expectations $\boldsymbol{E(X|Y=y)}$

y	$E(X\|Y=y)$	$P(Y=y)$	$E(X\|Y=y)\cdot P(Y=y)$
0	$\frac{2}{5}$	$\frac{25}{36}$	$\frac{10}{36}$
1	$\frac{1}{5}$	$\frac{10}{36}$	$\frac{2}{36}$
2	**0**	$\frac{1}{36}$	**0**
			Total $= \frac{1}{3} = E(X)$

So: $E(X) = \sum_y E(X|Y=y)\cdot P(Y=y)$

Or: $E(X) = E[E(X|Y)]$

Conditional probability functions

$$P(X = x | Y = y) = \frac{P(X = x \text{ and } Y = y)}{P(Y = y)}$$

for any (fixed) value of y.

Conditional expectation:

- $E(X|Y = y) = \sum_{x \text{ in } SX} x \cdot P(X = x|Y = y)$
- $E(X|Y)$ is a **r.v.** that attains the values $E(X|Y = y)$ with accompanying probabilities $P(Y = y)$.
- $E(X) = E[E(X|Y)]$

In the case of **independence** of X and Y:

$$P(X = x|Y = y) = P(X = x)$$

and $E(X|Y = y) = E(X)$

Example conditional distributions with formulas

Information about selling pairs of shoes by a web shop:

- $N =$ "# ordered pairs on a day" $\sim Poisson\ (\mu = 8)$
- 40% of the shoes are returned to the web shop (60% sold).
- $X =$ "# sold pairs on that day"

What is **a**. $P(X = 0 \text{ and } N = 5)$

b. $E(X)$

c. $P(X = 0)$ and the distribution of X

d. $P(N = 5|X = 0)$

a. X has, given $N = n$, a $B(n, 0.6)$-distr.:

$$P(X = x|N = n) = \binom{n}{x} 0.6^x 0.4^{n-x},\ x = 0, 1, \ldots, n$$

$$P(X = 0 \text{ and } N = 5) = P(X = 0|N = 5) \cdot P(N = 5)$$

$$= (0.4)^5 \cdot \frac{8^5}{5!} e^{-8}$$

Continuation example conditional distributions

b. $E(X|N=n) = np = 0.6n$, and therefore $E(X|N) = 0.6N$

$E(X) = E[E(X|N)] = E[0.6N] = 0.6 \cdot E(N) = 4.8$

c. $P(X=0) = \sum_{n=0}^{\infty} P(X=0 \text{ and } N=n)$

$= \sum_{n=0}^{\infty} P(X=0|N=n) \cdot P(N=n)$

$= \sum_{n=0}^{\infty} (0.4)^n \cdot \frac{8^n}{n!} e^{-8}$

$= e^{-8} \cdot \sum_{n=0}^{\infty} \frac{3.2^n}{n!}$ *use the Taylor series* $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$

$= e^{-8} \cdot e^{3.2} = e^{-4.8}$

This is $P(X=0)$ for a Poisson (μ = 4.8) – distribution

Similarly $P(X=x) = \sum_{n=x}^{\infty} P(X=x|N=n) \cdot P(N=n)$

which shows that $X \sim Poisson(\mu = 4.8)$

d. $P(N=5|X=0) = \frac{P(X=0 \text{ and } N=5)}{P(X=0)} = \frac{\frac{3.2^5}{5!} e^{-8}}{e^{-4.8}} = \frac{3.2^5}{5!} e^{-3.2}$

Measures for the relation of *X* and *Y*

<u>Definition</u> the **covariance** of *X* and *Y* is:

$$\boldsymbol{cov}(\boldsymbol{X},\boldsymbol{Y}) = \boldsymbol{E}(\boldsymbol{X} - \mu_{\boldsymbol{X}})(\boldsymbol{Y} - \mu_{\boldsymbol{Y}})$$

Direct calculation with the joint distribution:

$$\mathbf{cov}(X, Y) = \sum_x \sum_y (x-\mu_X)(y-\mu_Y) \cdot P(X=x \text{ and } Y=y)$$

Calculation with the **formula**:

$$\boldsymbol{cov}(\boldsymbol{X},\boldsymbol{Y}) = \boldsymbol{E}(\boldsymbol{XY}) - \boldsymbol{E}(\boldsymbol{X}) \cdot \boldsymbol{E}(\boldsymbol{Y})$$

where:

$$\boldsymbol{E}(\boldsymbol{XY}) = \sum_x \sum_y xy \cdot P(X=x \text{ and } Y=y)$$

Other **properties of the covariance**

- $cov(X, Y) = cov(Y, X)$
- $cov(X, X) = var(X)$
- $cov(aX + b, Y) = a \cdot cov(X, Y)$

"the covariance depends on the unit of measurement"

- $cov(X, Y + Z) = cov(X, Y) + cov(X, Z)$

- $var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$

and

$$var(X - Y) = var(X) + var(Y) - 2cov(X, Y)$$

the correlation coefficient: $\rho(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$

Example 1: $\rho(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{-\frac{2}{36}}{\sqrt{\frac{10}{36}} \cdot \sqrt{\frac{10}{36}}} = -\frac{1}{5}$

"*X* and *Y* have a weak, negative correlation"

Properties: $-1 \le \rho \le 1$ **(ρ is a "normalized" measure)**

Interpretation of the values of ρ w.r.t. the relation of *X* and *Y*:

- $\rho = 1$ → **strict linear, positive relation**
- $\rho = -1$ → **strict linear, negative relation**
- $\rho = 0$ → **no linear relation**
- $|\rho| > 0.9$ → **strong linear relation**

Furthermore:

$\rho(aX, Y) = \rho(X, Y)$ **(a > 0)**

$\rho(X + b, Y) = \rho(X, Y)$

"ρ is a normalized measure of linear relation: not depending on unit of measurement or place".

Variance of the sum for a sequence of r.v.`s:

$$var(\sum_i X_i) = \sum_i var(X_i) + \sum\sum_{i \neq j} cov(X_i, X_j)$$

$$\boldsymbol{n^2 \quad = \quad n \qquad\qquad + \qquad n(n\text{-}1)}$$

$\boldsymbol{n^2}$ **= the number of terms** $\boldsymbol{(X_i, X_j)}$

Properties for **independent r.v.`s:**

1. $\boldsymbol{E(XY) = EX \cdot EY}$

2. $\boldsymbol{cov(X, Y) = 0}$ and $\rho(\boldsymbol{X, Y}) = \boldsymbol{0}$

   "X and Y are not correlated"

3. $\boldsymbol{var(X \pm Y) = var(X) + var(Y)}$

   and $\boldsymbol{var(\sum_i X_i) = \sum_i var(X_i)}$

Properties of functions $g(X, Y)$

1. $E g(X,Y) = \sum_x \sum_y g(x,y) \cdot P(X = x \text{ and } Y = y)$

e.g.: $E(XY) = \sum_x \sum_y xy \cdot P(X = x \text{ and } Y = y)$

2. $E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$

This rule holds for all (independent and dependent) variables!

Convolution sum:

for discrete and independent X and Y we have:

$$P(X + Y = k) = \sum_i P(X = i) P(Y = k - i)$$

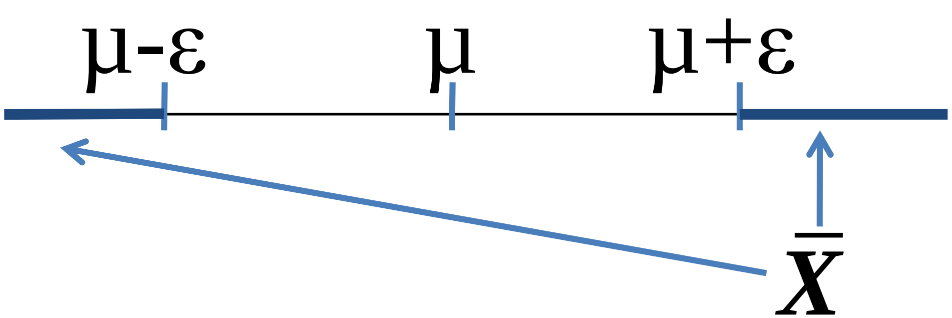
Application: if X and Y are independent and both have a Poisson distribution with "means" μ_1 and μ_2 resp., then $X + Y$ has a Poisson distribution as well with mean $\mu_1 + \mu_2$

The sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (or $\overline{X}_n$)

If $X_1, X_2, \dots, X_n$ **are independent and all have the same distribution** with expectation μ and variance σ^2, then:

1. $E(\overline{X}) = \mu$ and $var(\overline{X}) = \frac{\sigma^2}{n}$
2. **The weak law of large numbers**:

For each $\varepsilon > 0$: $\lim_{n\to\infty} P(|\overline{X}_n - \mu| \geq \varepsilon) = 0$



For large n the probability of the event in the diagram approaches 0.