L7

Basic Statistics for Economists

Spring 2020

Department of Statistics

Stockholm University

Earlier, from L5 and L6

Random variables

- Discrete and continuous
- Probability functions vs. density functions
  - $P_X(x) = P(X = x)$ vs. $f_X(x) \neq P(X = x)$
- Cumulative probability functions: $F_X(x) = P(X \leq x)$
- Expected values and variances
- Linear combination and standardizing
- **Binomial** distribution (Bernoulli) and **Normal** distribution
- Direct calculation and tables

Binomial – special case

- The table gives values for $\boldsymbol{n}$ up to and including $\mathbf{20}$
  - What to do if $n > 20$?
  - Direct calculation using $P_X(x)$ is very laborious (without computer)
  - Approximation will be handled in L8, read section NCT 5.4

- Table gives values of $\boldsymbol{p}$ up to an including $\mathbf{0,5}$.
  - What if $p > 0{,}5$?
  - Tip: create $Y = n - X =$ number of failures, number of 0's
  - Take for example … (from L5)

Exercise 1

- In a corporation, 60% of the employees are positive to a particular change.
- Suppose that we have a random sample of four people and that we wish to measure the number of people who are positive to the change, in our sample.

a) Construct the probability function for the r.v. $X =$ ”Number of positive people in the sample.”

b) Calculate expected value and variance of X.

c) What is the probability of at most two positive in the sample?

d) What is the probability of at least two positive?

Exercise 1, solution

- Number of positive $= X \sim Bin(n, p) = Bin(4; 0{,}6)$, $S_X = \{0, 1, 2, 3, 4\}$

a) Probability function for X:

$$\boldsymbol{P_X(x)} = \binom{n}{x} P^x (1-P)^{n-x} = \binom{\mathbf{4}}{\boldsymbol{x}} \cdot \mathbf{0{,}6}^{\boldsymbol{x}} \cdot \mathbf{0{,}4}^{\mathbf{4}-\boldsymbol{x}}$$

b) Expected value and variance for X:

$$\boldsymbol{\mu_X} = \boldsymbol{nP} = 4 \cdot 0{,}6 = 2{,}4$$

$$\boldsymbol{\sigma_X^2} = \boldsymbol{nP(1-P)} = 4 \cdot 0{,}6 \cdot 0{,}4 = \mathbf{0{,}96}$$

=BINOM.FÖRD(2;4;0,6;**1**)

c) The probability of at most two positive in the sample $= \boldsymbol{F_X(2)}$

$$= P_X(0) + P_X(1) + P_X(2) = [\text{complement}] = 1 - (P_X(3) + P_X(4))$$

Stockholm University

Exercise 1, cont.

c) $= 1 - \left[\binom{4}{3} \cdot 0{,}6^3 \cdot 0{,}4^1\right] - \left[\binom{4}{4} \cdot 0{,}6^4 \cdot 0{,}4^0\right] = \mathbf{0,5248}$

$$\binom{4}{3} = \frac{4!}{3!\,1!} = \frac{1 \cdot 2 \cdot 3 \cdot 4}{1 \cdot 2 \cdot 3 \cdot 1} = 4 \qquad \binom{4}{4} = \frac{4!}{4!\,0!} = \frac{1 \cdot 2 \cdot 3 \cdot 4}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 1} = 1$$

=BINOM.FÖRD(3;4;0,6;**0**)   =BINOM.FÖRD(4;4;0,6;**0**)

Alternative with table:

- The table does not work directly since $\boldsymbol{P = 0,6 > 0,5}$
- Define the **linear combination** $\boldsymbol{Y = n - X}$
- $"X \sim Bin(n;P)" + "Y = n - X" \Rightarrow$ $\boxed{\boldsymbol{Y \sim Bin(n; 1 - P)}}$ $\boxed{\boldsymbol{P_X(x) = P_Y(n - x)}}$

Stockholm University

Exercise 1, cont.

c) ”Transform” the event; use Y instead of X

List the outcomes of X and Y list them in rows:

$\boldsymbol{x}$	0	1	2	3	4
$\boldsymbol{y = n - x}$	4	3	2	1	0

$\boldsymbol{P(X \leq 2)} = [\text{circle}] \quad = [\text{equivalent to}] = \boldsymbol{P(Y \geq 2)}$

$= [\text{complement}] = 1 - P(Y \leq 1) = 1 - F_Y(1) = [\text{use table}]$

$= 1 - 0{,}4752 = 0{,}5248$

d) Do it yourself! Answer: $P(X \geq 2) = P(Y \leq 2) = 0{,}8208$

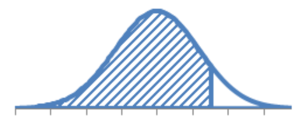
Stockholm University

Crib for normal table

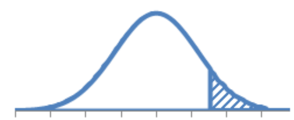
- Found in Mondo under ”Formulary and Distribution Tables”
- Draw and transform the sought probability into a form that allows you to use the table

Anta $Z \sim N(0,1)$ och att $a > 0$, dvs. ett positivt tal. Vi har fyra möjligheter:

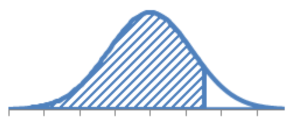
1. $\boldsymbol{P(Z \leq a)} = F_Z(a) \Rightarrow$ tabellen



2. $\boldsymbol{P(Z > a)} =$ [komplementet] $= 1 - P(Z \leq a) = 1 - F_Z(a) \Rightarrow$ tabellen



$= \mathbf{1} -$



- There are two more cases…

Stockholm University

Today

- ***BIVARIATE DISTRIBUTIONS (Discrete and continuous)***
  - Simultaneous distributions, (density functions), cumulative distribution functions
  - Marginal distributions
  - Conditional distributions
  - Independent variabels
  - Covariance and correlation
  - Mean and variance of linear functions of random variables (linear combinations)

Two or more r.v. studied together

Why multiple random variables at once?

- **Studying relationships**
  - When there is a (assumed) relationship between different variables, models can be created
  - You can use one to make predictions about the other
- We will focus on discrete r.v. and we start from ***cross tables*** but one can use functions. This also works for continuous r.v.
- **Numerical measures**
  - ***covariance*** and ***correlation***

Simultaneous distributions and functions

- **Discrete bivariate probability functions:**
  - $\boldsymbol{P_{X,Y}(x,y) = P(X = x \cap Y = y)}$ a probability (for given values x and y)
  - A table or function of x and y which gives us simultaneous probabilities for different combinations of x and y.

- **Continuous bivariate density function:**
  - $\boldsymbol{f_{X,Y}(x,y) \neq P(X = x \cap Y = y)}$ **not a probability**
  - A function of x and y that gives us simultaneous ”densities” for different combinations of x and y.

- **Cumulative probability function:**
  - $\boldsymbol{F_{X,Y}(x,y) = P(X \leq x \cap Y \leq y)}$ a probability (for given values x and y)
  - A table or function of x and y that gives us simultaneous cumulative probabilities for different combinations of x and y.

Example: dice rolls

- (X, Y) = number of dots, roll 1 vs. roll 2.
- **Simultaneous probability function**: $P_{X,Y}(x, y) = 1/36$

$P_{X,Y}(x,y)$	y = 1	2	3	4	5	6	$P_Y(y)$
x = 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
$P_X(x)$	1/6	1/6	1/6	1/6	1/6	1/6	1

$$P_{X,Y}(3,4) = \frac{1}{36}$$

Stockholm University

Example: dice rolls, cont.

- (X, Y) = number of dots, roll 1 vs. roll 2.
- **Cumulative probability function**: $F_{X,Y}(x, y) = xy/36$

$P(X \leq 3 \cap Y \leq 4)$ motsvarar $3 \cdot 4 = 12$ möjliga utfall, vart and ett med sannolikhet $= 1/36$

$F_{X,Y}(x,y)$	y = 1	2	3	4	5	6	$F_Y(y)$
x = 1	1/36	2/36	3/36	4/36	5/36	6/36	1/6
2	2/36	4/36	6/36	8/36	10/36	12/36	2/6
3	3/36	6/36	9/36	12/36	15/36	18/36	3/6
4	4/36	8/36	12/36	16/36	20/36	24/36	4/6
5	5/36	10/36	15/36	20/36	25/36	30/36	5/6
6	6/36	12/36	18/36	24/36	30/36	36/36	6/6
$F_X(x)$	1/6	2/6	3/6	4/6	5/6	6/6	

$$F_{X,Y}(3,4) = \frac{3 \cdot 4}{36} = \frac{12}{36} = \frac{1}{3}$$

Compare L4 p. 37

Bivariate probability function

- Suppose that X can take values $S_X = \{0,1,\dots,C\}$
- Suppose that Y can take values $S_Y = \{0,1,\dots,D\}$

	$Y = 0$	…	D	Sum
$X = 0$	$P_{X,Y}(0,0)$	…	$P_{X,Y}(0,D)$	$P_X(0)$
1	$P_{X,Y}(1,0)$	…	$P_{X,Y}(1,D)$	$P_X(1)$
⋮	⋮		⋮	⋮
C	$P_{X,Y}(C,0)$	…	$P_{X,Y}(C,D)$	$P_X(C)$
Sum	$P_Y(0)$	…	$P_Y(D)$	**1**

Simultaneous distribution of X <u>and</u> Y

Marginal distribution of X

Marginal distribution of Y

Stockholm University

Compare L4 p. 38

Marginal distributions

Column sum: $P_Y(y) = \sum_{x=0}^{C} P_{X,Y}(x,y)$ **Sum over x to get $P_Y(y)$**

	$Y=0$	...	D	Summa
$X=0$	$P_{X,Y}(0,0)$	...	$P_{X,Y}(0,D)$	$P_X(0)$
1	$P_{X,Y}(1,0)$	...	$P_{X,Y}(1,D)$	$P_X(1)$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
C	$P_{X,Y}(C,0)$	...	$P_{X,Y}(C,D)$	$P_X(C)$
Summa	$P_Y(0)$	...	$P_Y(D)$	**1**

Marginal distribution for Y

Row sum:

$$P_X(x) = \sum_{y=0}^{D} P_{X,Y}(x,y)$$

sum over y to get $P_X(x)$

Stockholm University

Conditional distribution

- Condition Y on a particular value of x, i.e. $Y|X = x$

1. Select the row that corresponds to $X = x$, t.ex. $X = 1$

	$Y = 0$	…	D	Sum
$X = 1$	$P_{X,Y}(1,0)$	…	$P_{X,Y}(1, D)$	$P_X(1)$

2. Divide all probabilites by $P_X(1)$

	$Y = 0$	…	D	Sum
$X = 1$	$\frac{P_{X,Y}(1,0)}{P_X(1)}$	…	$\frac{P_{X,Y}(1, D)}{P_X(1)}$	$\frac{P_X(1)}{P_X(1)} = 1$

3. The results are the conditional probabilities of Y given $X = 1$

	$Y = 0$	…	D	Sum
$X = 1$	$P_{Y\|X=1}(0)$	…	$P_{Y\|X=1}(D)$	1

Conditional probability distributions

- For discrete r.v. with probability functions $P_X(x)$, $P_Y(y)$ and $P_{X,Y}(x,y)$ the **conditional** probability functions are defined:

$$P_{Y|X}(y|x) = \frac{P_{X,Y}(x,y)}{P_X(x)} \qquad P_{X|Y}(x|y) = \frac{P_{X,Y}(x,y)}{P_Y(y)}$$

- Corresponding expressions for continuous density functions:

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

- **Compare to the formula** $\boldsymbol{P(A|B) = \frac{P(A \cap B)}{P(B)}}$ (F4 p. 25)

Stockholm University

Conditional expected values and variances

- Consider a bivariate distribution. Once you have a **conditional distribution** this distribution is a **function of <u>one</u> variable**.
- Condition on $X = 1$. We get a function of only y:

$$P_{Y|X}(y|1) = \frac{P_{X,Y}(1,y)}{P_X(1)}$$

x has been substituted for the number 1

- For this new distribution we can calculate expected value and variance:

$$\boldsymbol{\mu_{Y|X=1}} = \sum_{y \in S_Y} y P_{Y|X=1}(y|1) \qquad \boldsymbol{\sigma^2_{Y|X=1}} = \sum_{y \in S_Y} \left(y - \mu_{Y|X=1}\right)^2 P_{Y|X=1}(y|1)$$

Stockholm University

Exercise 2

- From the following table, calculate the simultaneous probabilities for (X, Y):

X \ Y	0	1	2
0	1/9	2/9	1/9
1	2/9	2/9	0
2	1/9	0	0

a) What is the probability of the event: $X \leq 1 \cap Y \geq 1$?

b) Calculate the **marginal distributions** for X and Y and **expected value** and **variance** of X and Y.

c) State the **conditional distribution** and **expected value** and **variance** of X **given that** $Y = 1$.

Exercise 2, solution

a) The probability that $X \leq 1 \cap Y \geq 1$:

X \ Y	0	1	2
0	1/9	2/9	1/9
1	2/9	2/9	0
2	1/9	0	0

- Circle the area and calculate the sum! (the cells are pairwise disjoint)

$$P(X \leq 1 \cap Y \geq 1) = \frac{2+2+1+0}{9} = \mathbf{\frac{5}{9}}$$

Stockholm University

Verify this yourself by doing the calculations!

Exercise 2, cont.

b) Sum over columns for X and rows for Y:

X \ Y	0	1	2	P(x)
0	1/9	2/9	1/9	4/9
1	2/9	2/9	0	4/9
2	1/9	0	0	1/9
P(y)	4/9	4/9	1/9	1

Marginal distribution of X

$\mu_X = \mathbf{\frac{2}{3}}$, $\sigma_X^2 = \mathbf{\frac{4}{9}}$

(same as for Y in this case)

Marginal distribution of Y

$$\mu_Y = 0 \cdot \frac{4}{9} + 1 \cdot \frac{4}{9} + 2 \cdot \frac{1}{9} = \frac{6}{9} = \mathbf{\frac{2}{3}}$$

$$\sigma_Y^2 = 0^2 \cdot \frac{4}{9} + 1^2 \cdot \frac{4}{9} + 2^2 \cdot \frac{1}{9} - \left(\frac{2}{3}\right)^2 = \mathbf{\frac{4}{9}}$$

Exercise 2, cont.

Verify this yourself by doing the calculations!

c) Pick the column for which $Y = 1$ and divide each probability by $P(Y = 1)$:

X \ Y	0	1	2	P(x)
0	1/9	2/9	1/9	4/9
1	2/9	2/9	0	4/9
2	1/9	0	0	1/9
P(y)	4/9	4/9	1/9	1

x	0	1	2	
$P(x\|1)$	$\frac{2/9}{4/9} = \mathbf{\frac{1}{2}}$	$\frac{2/9}{4/9} = \mathbf{\frac{1}{2}}$	$\frac{0/9}{4/9} = \mathbf{0}$	$\frac{4/9}{4/9} = 1$

$$\mu_{X|Y=1} = \mathbf{\frac{1}{2}}$$

$$\sigma^2_{X|Y=1} = \mathbf{\frac{1}{4}}$$

Compare to $\mu_X = \mathbf{\frac{2}{3}}$, $\sigma_X^2 = \frac{4}{9}$

Independent

The probability of one is not affected by what happens to the other. The conditional distribution of one is equal to the marginal distribution regardless of the outcome of the other.

- Two r.v. are **independent** if the probability function (density functions) of one is unaffected by conditioning on the other.
- I.e. if

$$\boldsymbol{P_{Y|X}(y|x) = P_Y(y)} \qquad\qquad \boldsymbol{P_{X|Y}(x|y) = P_X(x)}$$

for all combinations of x and y we say that X and Y are **independent**.

- Compare to L4 p. 30:

$P(A|B) = P(A)$ and $P(B|A) = P(B) \Leftrightarrow A$ and B are independent

- Independence also means that conditional expected values and variances are equal to the expected values and variances of the marginal distributions.

Stockholm University

Example

- What about the following r.v. X and Y? Independent or not?

X \ Y	0	1	2	P(x)
0	1/9	2/9	1/9	4/9
1	2/9	2/9	0	4/9
2	1/9	0	0	1/9
P(y)	4/9	4/9	1/9	1

- They are dependent if one can find a cell where the value is different from the product of the row and column margins.
- **Here: Dependent!**
- (e.g. $P(2,2) = 0$, while $P(X = 2) \cdot P(Y = 2) = \frac{1}{81}$)

Compare L3 p. 13

Covariance

- Measure of two random variables vary together, **linear relationships**
- Can be negative, positive, or zero.
- Not always easy to interpret.

$$\sigma_{XY} = Cov(X,Y) = \sum_{x \in S_x} \sum_{y \in S_y} (x - \mu_X)(y - \mu_Y) P_{XY}(x,y)$$

alternative formula: $$= \sum_{x \in S_x} \sum_{y \in S_y} xyP(x,y) - \mu_X \mu_Y$$

Stockholm University

compare F3 p. 14

Correlation

- ***Correlation*** is covariance scaled to the interval [-1,1]
- Independent of scale, easier to compare pairs of variables
- Absolute values close to 1, i.e. near -1 or +1, indicate strong ***linear*** relationships
- Values close to zero indicate weak linear relationships

Definition:

$$\rho_{XY} = Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Exercise 3 (same r.v. as exercise 2)

- Calculate covariance and correlation for X and Y:

X \ Y	0	1	2	P(x)
0	1/9	2/9	1/9	4/9
1	2/9	2/9	0	4/9
2	1/9	0	0	1/9
P(y)	4/9	4/9	1/9	1

Previous results:

$\mu_X = \mu_Y = \frac{2}{3}$

$\sigma_X = \sigma_Y = \sqrt{4/9} = \frac{2}{3}$

- Nine combinations of x and y:

$$\begin{aligned}\boldsymbol{\sigma_{XY}} = &\left(0 \cdot 0 \cdot \tfrac{1}{9}\right) + \left(0 \cdot 1 \cdot \tfrac{2}{9}\right) + \left(0 \cdot 2 \cdot \tfrac{1}{9}\right)\\ &+ \left(1 \cdot 0 \cdot \tfrac{2}{9}\right) + \left(1 \cdot 1 \cdot \tfrac{2}{9}\right) + \left(1 \cdot 2 \cdot \tfrac{0}{9}\right)\\ &+ \left(2 \cdot 0 \cdot \tfrac{1}{9}\right) + \left(2 \cdot 1 \cdot \tfrac{0}{9}\right) + \left(2 \cdot 2 \cdot \tfrac{0}{9}\right) - \frac{2}{3} \cdot \frac{2}{3}\end{aligned}$$

$$= -\frac{2}{9}$$

$$\boldsymbol{\rho_{XY}} = -\frac{2/9}{4/9} = -\frac{1}{2}$$

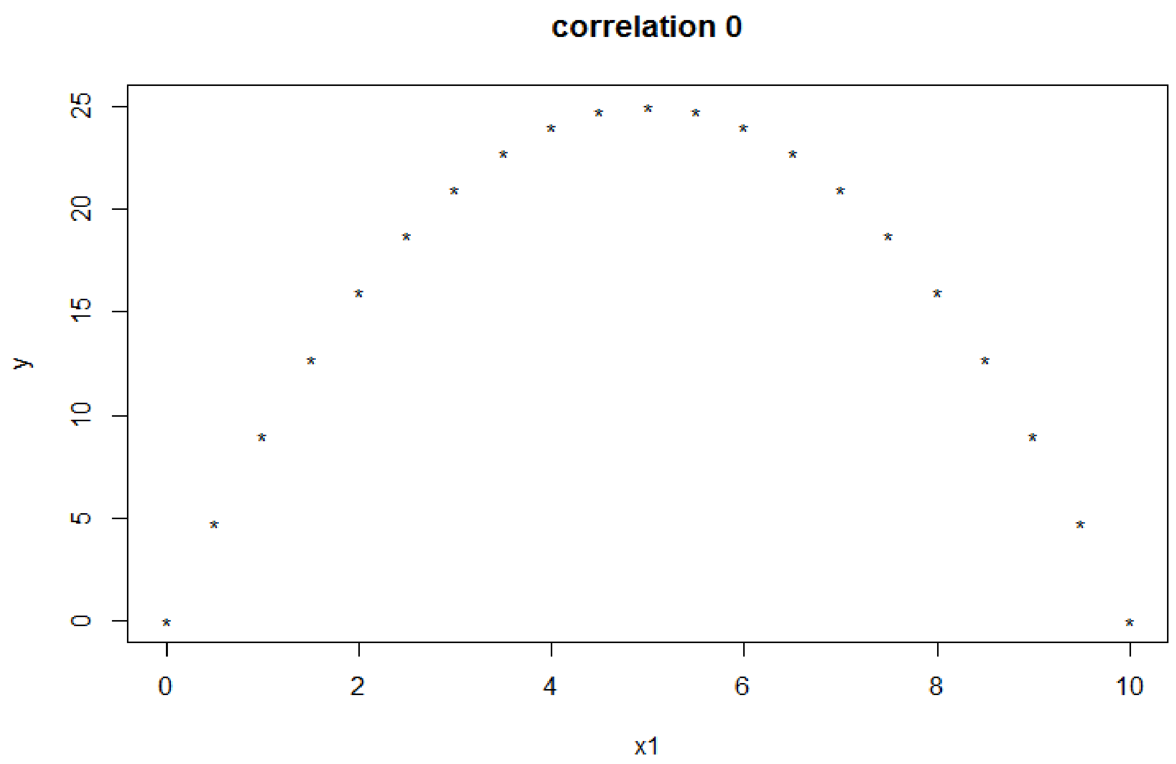
Stockholm University

Independence and correlation

- if X and Y are **independent** $\Rightarrow \boldsymbol{\rho_{XY}} = 0$
  - if $\boldsymbol{\rho_{XY}} \neq 0 \Rightarrow$ linearly dependent, strongly or weakly
- if $\boldsymbol{\rho_{XY}} = 0 \;\Rightarrow\;$?
  - You cannot say if X and Y are independent, or not
  - The relationship could be non-linear and strong; correlation measures linear relationships
  - To be sure you need further analysis

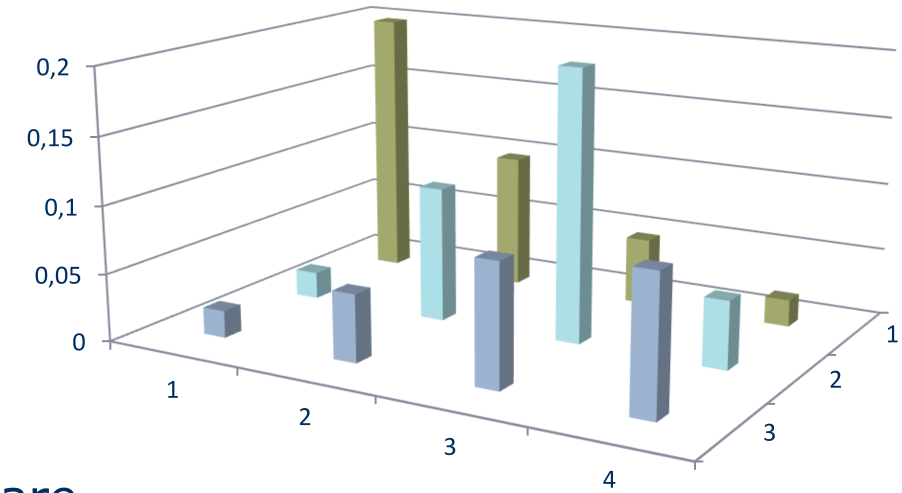
Example

Dependent random variables $\boldsymbol{X}, \boldsymbol{Y}$ with zero correlation. ($P(X = x) = \frac{1}{21}$ for all x; $Y = X^2$)



Two random variables, graphically

- Each bar represents a combination of x and y
- The height of each pair represents the probability of that event

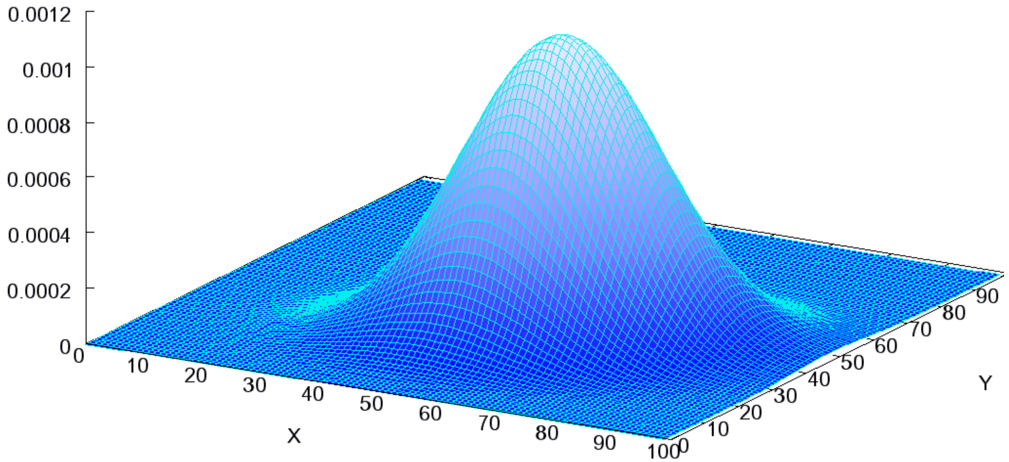


- Difficult to read and compare
- Not easy to draw in three dimensions, even worse in higher dimensions.

Stockholm University

The bivariate normal distribution

- Five parameters: expected values $\boldsymbol{\mu_X}$, $\boldsymbol{\mu_Y}$; variances $\boldsymbol{\sigma_X^2}$, $\boldsymbol{\sigma_Y^2}$; and covariance $\boldsymbol{\sigma_{XY}}$ (alternatively, the correlation $\boldsymbol{\rho_{XY}}$)
- The marginal distributions of X and Y are normally distributed.
- Conditional distributions $X|Y$ and $Y|X$ are normally distributed



(Source: Wikipedia, https://en.wikipedia.org/wiki/Joint_probability_distribution)

Stockholm University

Linear combinations and linear combinations

Calculation for **expected values** – both discrete and continuous

- Simple linear combination of a r.v. X with expected value μ_X
  - $E(c) = \boldsymbol{c}$ (the expected value of a constant is equal to that constant)
  - $E(aX) = \boldsymbol{a\mu_X}$
  - $E(aX + c) = \boldsymbol{a\mu_X + c}$

- Linear combinations of two r.v. X and Y w/ expected values μ_X, μ_Y
  - $E(aX + bY + c) = \boldsymbol{a\mu_X + b\mu_Y + c}$

Stockholm University

Linear combinations and linear combinations

Calculation rules for **variances** – both discrete and continuous

- Simple linear combination of a r.v. X with variance $\boldsymbol{\sigma_X^2}$
  - $Var(c) = \mathbf{0}$ (a constant does not vary)
  - $Var(aX) = \boldsymbol{a^2\sigma_X^2}$
  - $Var(aX + c) = \boldsymbol{a^2\sigma_X^2}$ (the constant c shifts all values, but not the spread)

- Linear combination of two r.v. X and Y with expected values μ_X and μ_Y
  - $Var(aX + bY + c) = \boldsymbol{a^2\sigma_X^2 + b^2\sigma_Y^2 + 2abCov(X, Y)}$

Stockholm University

Exercise 4

Portfolio theory

- Let X be the return of a particular investment and Y the return of another, both with the same probability function:

x, y (%)	$P_X(x)$	$P_Y(y)$
-8	0,25	0,25
2	0,50	0,50
12	0,25	0,25
Summa	1,00	1,00

a) Calculate expected return and variance (risk) for both investments.

Create the portfolio $\boldsymbol{W = 0,5X + 0,5Y}$.

b) Calculated expected return and variance (risk) for the portfolio when $\boldsymbol{\rho_{XY} = 0;\ 0,5;\ 1;\ -0,5;}$ and $\boldsymbol{-1}$.

Stockholm University

Verify by doing the calculations yourself!

Exercise 4, solution

a) Expected value for X and Y (they have identical sample spaces and distributions):

$$\boldsymbol{\mu_X} = \boldsymbol{\mu_Y} = (-8 \cdot 0{,}25) + (2 \cdot 0{,}5) + (12 \cdot 0{,}25) = \mathbf{2}$$

Variance (risk):

$$\boldsymbol{\sigma_X^2} = \boldsymbol{\sigma_Y^2} = (10^2 \cdot 0{,}25) + (0^2 \cdot 0{,}5) + (10^2 \cdot 0{,}25) = \mathbf{50}$$

b) Expected value for the portfolio $\boldsymbol{W}$:

$$\boldsymbol{\mu_W} = E(aX + bY + c) = E(0{,}5X + 0{,}5Y + 0) = 0{,}5\mu_X + 0{,}5\mu_Y = \mathbf{1} + \mathbf{1} = \mathbf{2}$$

Variance (risk):

$$\boldsymbol{\sigma_W^2} = Var(aX + bY + c) = Var(0{,}5X + 0{,}5Y + 0) = \boldsymbol{a^2\sigma_X^2} + \boldsymbol{b^2\sigma_Y^2} + \boldsymbol{2abCov}(\boldsymbol{X}, \boldsymbol{Y})$$

$$= 0{,}5^2 \cdot 50 + 0{,}5^2 \cdot 50 + 2 \cdot 0{,}5 \cdot 0{,}5 \cdot Cov(X, Y) = \mathbf{25} + \mathbf{0{,}5} \cdot \boldsymbol{Cov}(\boldsymbol{X}, \boldsymbol{Y})$$

Stockholm University

Verify by doing the calculations yourself!

Exercise 4, cont.

- Variance: $\sigma_W^2 = 25 + 0{,}5 \cdot Cov(X,Y)$
- Correlation: $\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$
- Covariance: $Cov(X,Y) = \sigma_{X,Y} = \rho_{X,Y}\sigma_X\sigma_Y = \rho_{X,Y} \cdot \sqrt{50}\sqrt{50} = 50\rho_{X,Y}$

$\rho_{X,Y}$	-1	-0,5	0	0,5	1
$\sigma_{X,Y}$	-50	-25	0	25	50
σ_W^2	0	12,5	25	37,5	50

Risk eliminated; I am guaranteed 2% return

Reducing risk

No reduction of risk compared to only choosing one of the investment

Stockholm University

Exercise 5

- $\boldsymbol{X}$ = return of portfolio A, $\boldsymbol{\mu_X = 5}$ (expectation) and $\boldsymbol{\sigma_X^2 = 3^2}$ (risk).
- $\boldsymbol{Y}$ = return of portfolio B, $\boldsymbol{\mu_Y = 6}$ (expectation) and $\boldsymbol{\sigma_Y^2 = 6^2}$ (risk).
- Suppose that X and Y are both **normally distributed**.

a) What is the probability that A gives negative return?

b) What is the probability that B gives negative return?

c) Calculate the probability that A gets better return than B.
   - Which additional assumption do you have to make to calculate the probability in c)?
   - Is this a good assumption?

Stockholm University

Exercise 5, solution

a) ”A gives negative return” $= X < 0$

$$\boldsymbol{P(X < 0)} = P\left(Z < \frac{0-5}{3}\right) \approx P(Z < -1{,}67) = 1 - F_Z(1{,}67) = 1 - 0{,}95254 = \mathbf{0{,}04746}$$

b) ” A gives negative return” $= Y < 0$

$$\boldsymbol{P(Y < 0)} = P\left(Z < \frac{0-6}{6}\right) = P(Z < -1{,}00) = 1 - F_Z(1{,}00) = 1 - 0{,}84134 = \mathbf{0{,}15866}$$

c) ”A give better return than B” $= \boldsymbol{X > Y} \quad \Leftrightarrow \quad \boldsymbol{X - Y > 0}$

Define $\boldsymbol{W} = aX + BY + c = \boldsymbol{X - Y}$ i.e. a linear combination with

$a = 1,\ b = -1,\ c = 0$

Exercise 5, cont.

c) Expected value: $\boldsymbol{E(W) = \mu_X - \mu_Y = 5 - 6 = -1}$

Variance: $\boldsymbol{Var(W)} = a^2 Var(X) + b^2 Var(Y) + 2ab Cov(X, Y)$

$= 1^2 \cdot \sigma_X^2 + (-1)^2 \cdot \sigma_Y^2 + 2 \cdot 1 \cdot (-1) \cdot Cov(X, Y)$

$= 3^2 + 6^2 - 2Cov(X, Y) = \boldsymbol{45 - 2Cov(X, Y)}$

Nothing is said about the covariance. If we assume **independence** between X and Y we get $\boldsymbol{Cov(X, Y) = 0}$ and $\boldsymbol{Var(W) = 45}$.

$$\boldsymbol{P(X - Y > 0)} = P(W > 0) = P\left(Z > \frac{0 - (-1)}{\sqrt{45}}\right) \approx P(Z > 0{,}15) = 1 - P(Z < 0{,}15)$$

$$= 1 - F_Z(0{,}15) = 1 - 0{,}55962 = 0{,}44038 \approx \boldsymbol{0{,}44}$$

Stockholm University

Summary

Two random variables (or many)

Probability functions - discrete
Density functions - continuous
Cumulative probability functions
- **simultaneous** and **marginal**

expected values, variances,
covariance and **correlations**

Independence

Conditioning - functions, expected values and variances

Linear combinations – add together several r.v. to make a new r.v. and its own **expected value** and **variance**

Next time

Sampling theory

- Many (n) independent r.v. $\boldsymbol{X_i}$ from the same distribution, i.e. a **sample** $= \{\boldsymbol{X_1}, \boldsymbol{X_2}, \dots, \boldsymbol{X_n}\}$
- Sample statistics which by definition are r.v. will be calculated, e.g. the sample mean $\overline{\boldsymbol{X}}$.
- What is the distribution of $\overline{\boldsymbol{X}}$?

Central Limit Theorem (CLT)

- The reason that the normal distribution occurs frequently
- An application of CLT: approximation of the binomial distribution

Stockholm University