L6

Basic Statistics For Economist

Spring 2020

Department of Statistics

Stockholm University

Last time

Bayes' theorem

- Reverse conditional prob.: $P(E_i|A) = \dfrac{P(A|E_i)P(E_i)}{P(A)} = \dfrac{P(A|E_i)P(E_i)}{\sum_k P(A|E_k)P(E_k)}$

Discrete random variables

- Probability function: $P(X = x) = P_X(x)$
- Cumulative probability functions: $P(X \leq x) = F_X(x)$
- Expected value μ_X and variance σ_X^2: how to calculate
- Linear combination $Y = a + bX$ and standardization $Z = \dfrac{X - \mu_X}{\sigma_X}$

Bernoulli experiment, Bernoulli distribution, 0-1 outcome

- parameter: P = probability of success = $P(X = 1)$

Stockholm University

Last time

Many Bernoulli experiments = **Binomial distribution = $\boldsymbol{Bin(n, P)}$**

- Sum of n independent Bernoulli with identical P
- parameters n and P
- $Bernoulli = Bin(1, P)$
- $\mu_X = E(X) = nP$, $\sigma_X^2 = Var(X) = nP(1 - P)$
- Probability function: $P(x) = \binom{n}{x} P^x (1 - P)^{n-x}$
- Calculate probabilities
  - Direct calculation using the probability function
  - Using table $F(x) = P(X \leq x)$; P(x)=F(x)-F(x-1)

Stockholm University

Today

- ***Continuous random variables***
  - Probability functions for **continuous** r.v.
  - **The density function** and the cumulative probability function
  - Expected value and variance of continuous r.v.
  - A particularly common distribution:

    The normal distribution
  - Standardized normal distribution
  - Calculate probabilities using a table
  - Approximate Binomial with a Normal (Ch. 5.4 is left for F8)

Continuous variables

- A **continuous variable** can *map* to **any value in an interval** (or many intervals). A discrete variable can only map to a *discrete* collection of values.

Example:

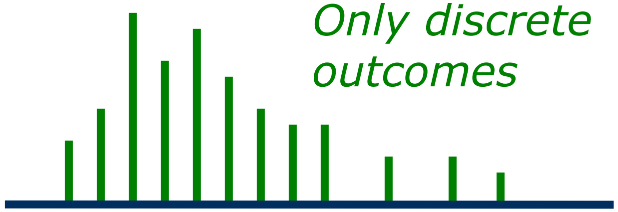
- The number of units that need repairing among 10 chosen ones is a discrete variable.
  - Can be listed: ex. $S_X = \{0,1,2,\dots,10\}$
- The time it takes to repair a unit is a continuous variable.
  - Cannot be listed
  - Continuous interval: e.g. $S_X = (0 \leq x \leq 100)$

From L5

Random variables, continued

Random variables

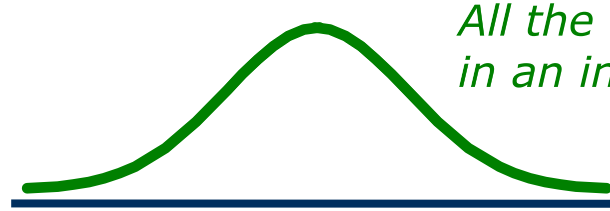
Discrete



Only discrete outcomes

The bars represent the probabilities of the different outcomes

Continuous



All the outcomes in an interval

The curve represents … ?

The so called ***Density Function***

Example of continuous random variables

Let $\boldsymbol{X}$ = "the repair time of a unit"

- Sample space: $\boldsymbol{S_X = \{0 \le X \le 100\}}$
- $\boldsymbol{S_X}$ = the elementary outcomes (disjoint) of the experiment are numbers, but they cannot be listed!
- We may define events as **sub intervals** of $\boldsymbol{S_X}$:

$A = [0,2]$	$0 \le X \le 2$	$X \le 2$
$\bar{A} = (2,100]$	$2 < X \le 100$	$X > 2$
$B = [2,100]$	$2 \le X \le 100$	$X \ge 2$
$C = \{11\}$	$X = 11$, i.e. exactly $= 11$	

Notice! For <u>continuous variables</u> it <u>does not matter</u> whether the inequality is strict or not!

Stockholm University

Cumulative distribution functions

A cumulative distribution function

$$\boldsymbol{F_X(x) = P(X \leq x)}$$

Probability that $\boldsymbol{X}$ does not excced the value $\boldsymbol{x}$ expressed as a function of $\boldsymbol{x}$.

The probability of an event = the probability of a subinterval of $\boldsymbol{S_X}$

Calculated as

$$\boldsymbol{P(a \leq X \leq b)} = P(X \leq b) - P(X < a) = \boldsymbol{F_X(b) - F_X(a)}$$

where $a < b$.

Stockholm University

The density function for a continuous r.v.

- The probability density function *(abbreviated pdf)*
- Denoted $\boldsymbol{f_X(x) \neq P(X = x)}$
- The density function is typically a continuous "curve" but its values are not probabilities but "densities" as a function of $\boldsymbol{x}$.

1. A density function has the property $\boldsymbol{f_X(x) \geq 0}$ for all $\boldsymbol{x \in S_X}$
2. <u>The area under the curve</u> of $\boldsymbol{f_X(x)}$ over all of $\boldsymbol{S_X}$ must be $= \mathbf{1}$
3. Let a and b be two points in $\boldsymbol{S_X}$ such that $a < b$. Then the probability that $\boldsymbol{X}$ takes a value beteen a and b = the area under the curve between these points: $\boldsymbol{P(a < X < b) = \int_a^b f_X(x)dx}$

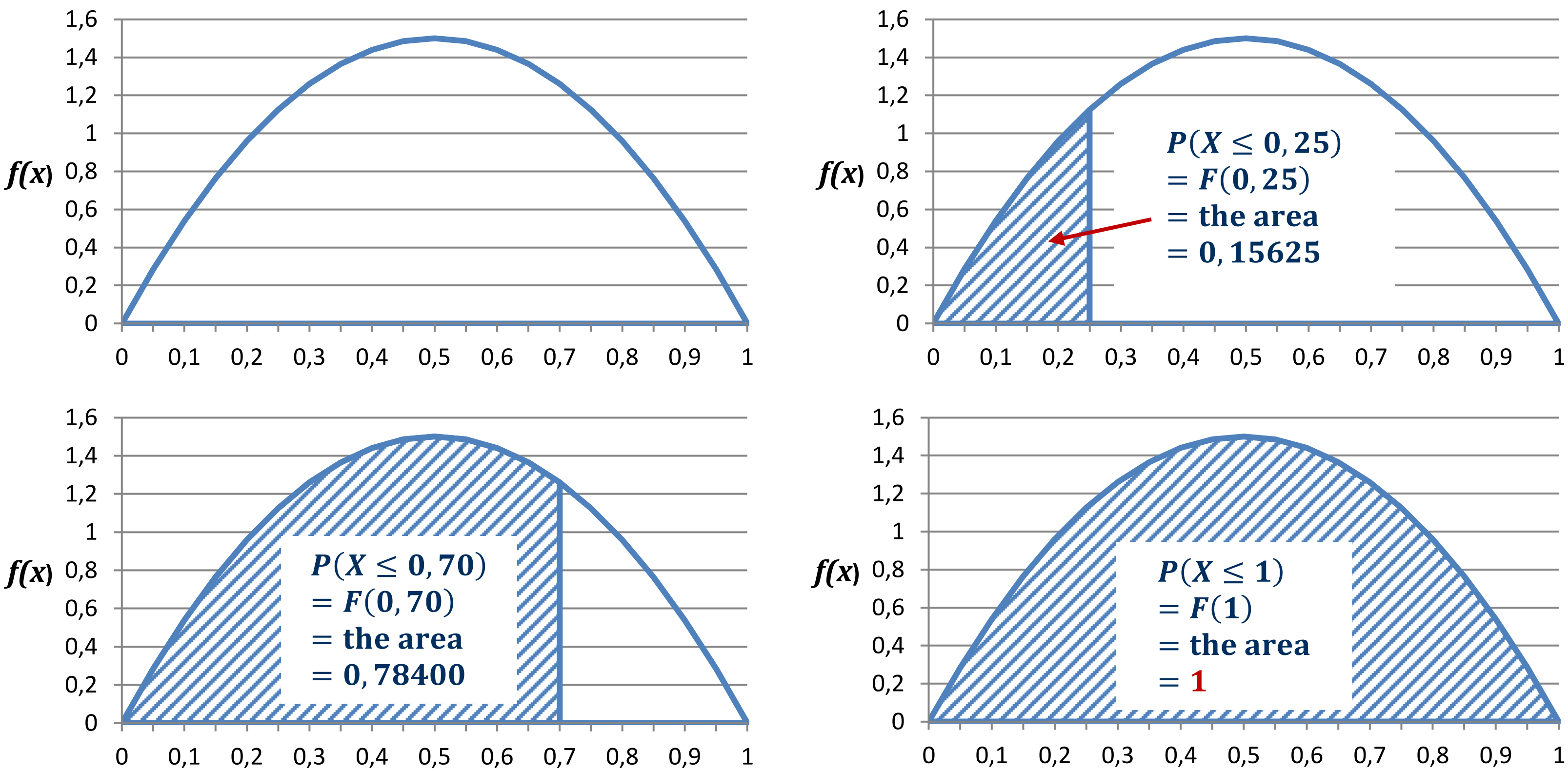
Stockholm University

Notice! $f(x)$ may be > 1

*But the **areas** under $f(x)$ must be ≤ 1*

Graphical illustration

- A simple density function: $f_X(x) = 6x - 6x^2$, $S_X = (0,1)$



Stockholm University

Graphical illustration, cont.

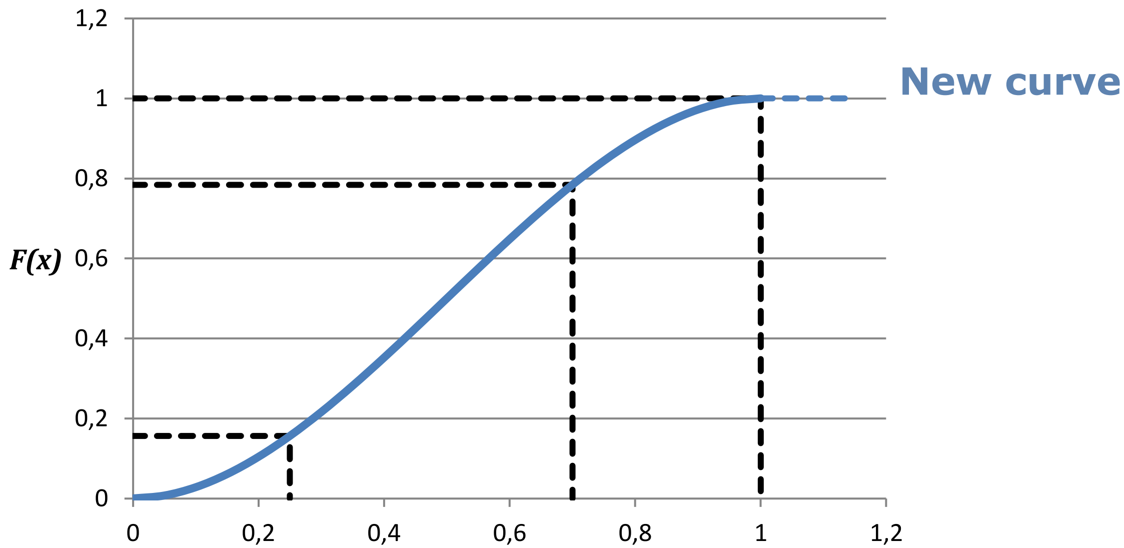
You do not need to know how I got $F_X(x)$ from $f_X(x)$.

- The cumulative probability function: $F_X(x) = 3x^2 - 2x^3$

$\boldsymbol{F_X(1) = 1}$

$\boldsymbol{F_X(0,70) = 0,784}$

$\boldsymbol{F_X(0,25) = 0,15625}$

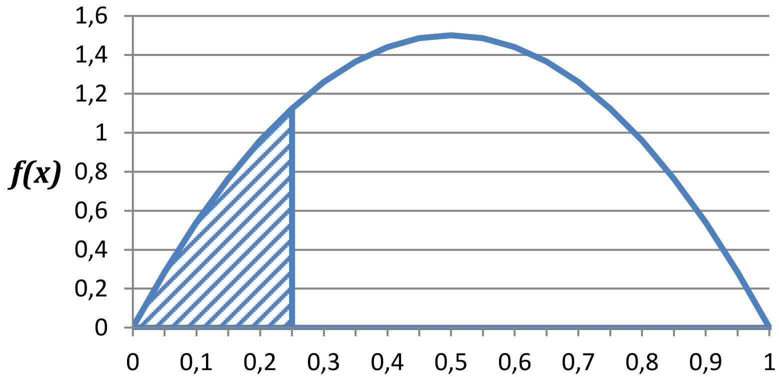


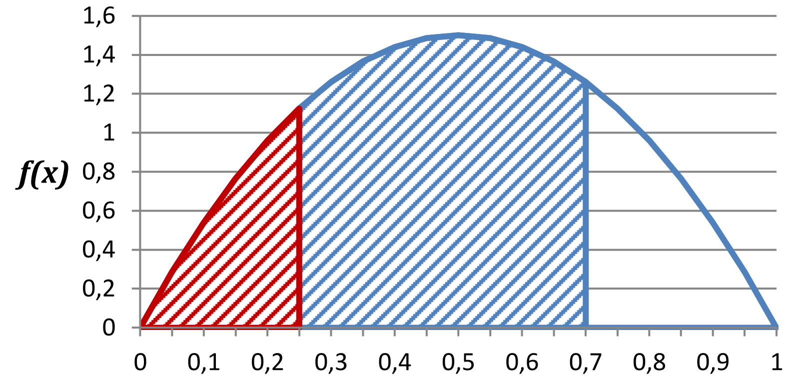
$\boldsymbol{x = 0,25}$ $\boldsymbol{x = 0,70}$ $\boldsymbol{x = 1}$

Stockholm University

Graphic illustration, cont.

- Calculate probabilities: $P(0{,}25 \leq X \leq 0{,}70) = F_X(0{,}70) - F_X(0{,}25)$

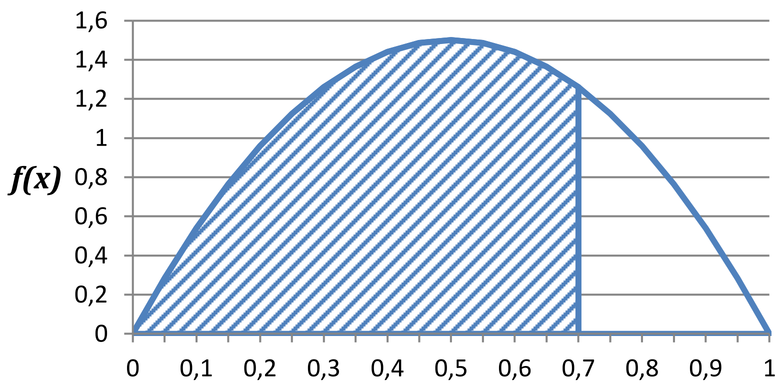




$\boldsymbol{F_X(0,70)}$

$\boldsymbol{-F_X(0,25)}$

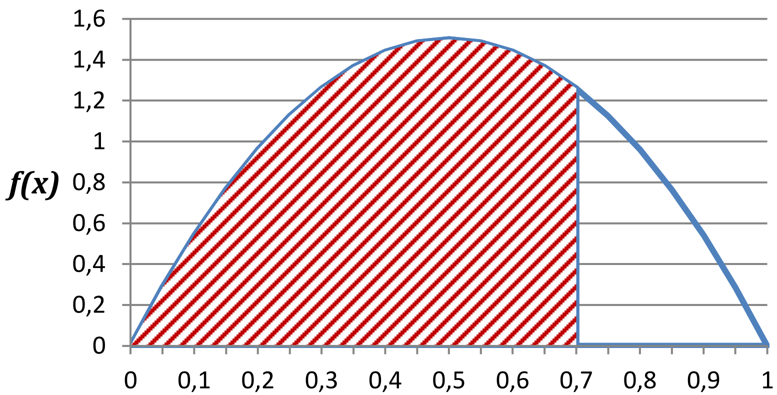
$\boldsymbol{P(0,25 \leq X \leq 0,70)}$



Stockholm University

The probability of a single point x

- Suppose you want to calculate $\boldsymbol{P(X = x)}$.
- We have that $\boldsymbol{P(X = x) = P(X < x) - P(X \leq x)}$. *Draw!*
- The probability is defined as the area under the curve $f_X(x)$.



- Remove all area except at $x = 0{,}70$
- What is the area of a line?
- **Zero. (0)**

- It does NOT matter if it says $<$ or $\leq$, it is the same!
- $\boldsymbol{P(X < x) = P(X \leq x) = F_X(x)}$

Summary measures

- The density function $\boldsymbol{f_X(x)}$ alternatively the cumulative density function $\boldsymbol{F_X(x)}$ includes all necessary information about $\boldsymbol{X}$.
  - *Assumed but not stated: we know what $\boldsymbol{S_X}$ is*

- **Summary measures** describe some important attributes of a probability function.

- ***"Location" – expected value***
  - theoretical average, what we can expect the average to be

- ***"Spread" – variance***
  - How spread out the values are

The expected value of a continuous r.v.

Expected value is defined as: $\mu_X = E(X) = \int_{x \in S_X} x f_X(x) dx$

- The expected value is a **weighted average** of the elements in $\boldsymbol{S_X}$ weighted with respect to the density function
- **You do not need to know calculus!**

Excepted value of a function $g(x)$: $E[g(X)] = \int_{x \in S_X} g(x) f_X(x) dx$

Variance for a continuous r.v.

Variance is defined as: $$\sigma_X^2 = Var(X) = \int_{x \in S_X} (x - \mu_X)^2 f_X(x) dx$$

- Weighted average of the squared distance to the mean expected value

- Again, **you do not need to know calculus!**

Stockholm University

Linear combinations

- En function the type $\boldsymbol{Y = g(X) = a + bX}$ is called a **linear combination**
- Suppose $E(X) = \mu_X$ and $Var(X) = \sigma_X^2$

  Then:

$$\boldsymbol{\mu_Y} = E(Y) = E(a + bX) = \boldsymbol{a + b\mu_X}$$

$$\boldsymbol{\sigma_Y^2} = Var(Y) = Var(a + bX) = \boldsymbol{b^2\sigma_X^2}$$

Notice! $\boldsymbol{b}$ squared, $\boldsymbol{b^2}$

- If you have a linear function $Y = g(X)$ you may calculate expected value and variance of Y directly, **this is also true for continuous r.v.**

Stockholm University

Standardization

Standardization **IMPORTANT!**

Suppose $\boldsymbol{Z} = \frac{\boldsymbol{X} - \boldsymbol{\mu_X}}{\boldsymbol{\sigma_X}} = -\frac{\mu_X}{\sqrt{\sigma_X^2}} + \frac{1}{\sqrt{\sigma_X^2}} \cdot X$

$$\boldsymbol{\mu_Z} = -\frac{\mu_X}{\sqrt{\sigma_X^2}} + \frac{1}{\sqrt{\sigma_X^2}} \cdot \mu_X = \mathbf{0} \qquad \boldsymbol{\sigma_Z^2} = \boldsymbol{b^2}\sigma_X^2 = \left(\frac{1}{\sqrt{\sigma_X^2}}\right)^2 \cdot \sigma_X^2 = \mathbf{1}$$

The normal distribution

- One of the most important, most famous, and most used probability functions of all time, for better or for worse
  - ”magnificent” mathematical properties
  - Occurs in nature (or as good approximation)
  - Particularly useful when one has many observations of X
    (We will return to this in L8)

- Alternative names:

  Bell curve

  Gaussian distribution

  after J.C.F. Gauss (1777-1855)

Stockholm University

The normal distribution

A **normally distributed** r.v. X has the density function

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} exp\left(-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right)$$

where μ_X and σ_X^2 are the function's **parameters** which fully determine what the distribution looks like and how you write $X \sim N(\mu_X, \sigma_X^2)$.

The sample space: $S_X = (-\infty, \infty)$ = all real numbers

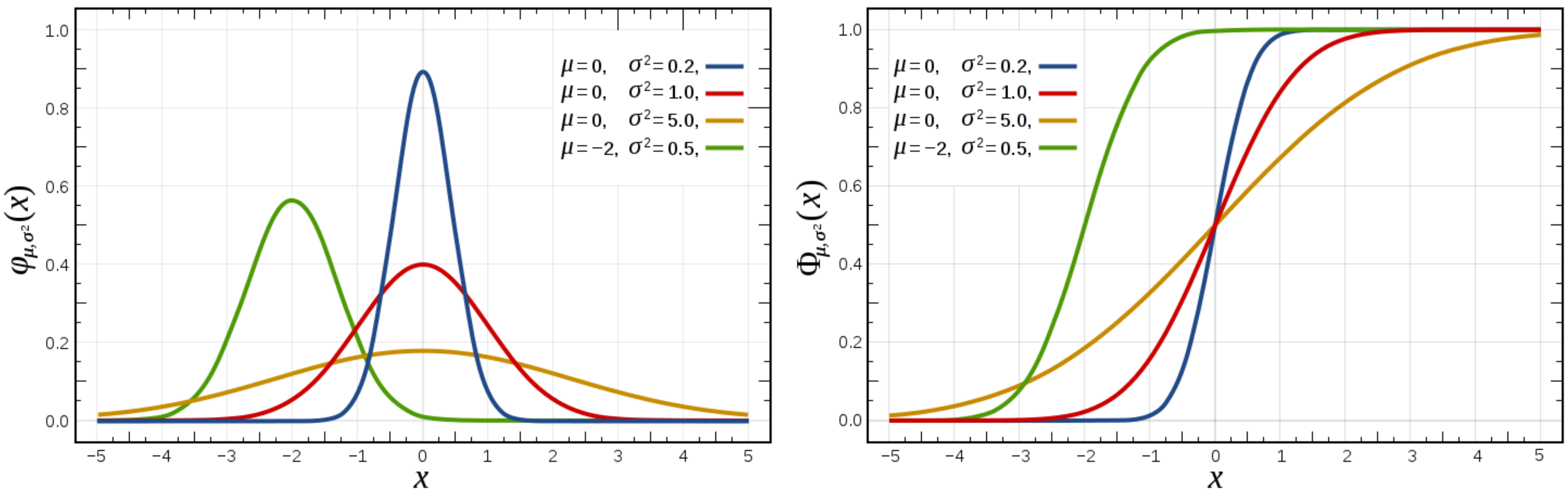
Expected value: $E[X] = \mu_X$ Variance: $Var(X) = \sigma_X^2$

The expected value and variance are equal to the parameters μ_X and σ_X^2 respectively.

Stockholm University

Different normal distributions

- Density functions and cumulative probability functions for different values of μ_X and σ_X^2



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

Stockholm University

Normal distribution, cont.

- The maximum of the curve

  = expected value $\boldsymbol{E(X) = \mu_X}$

- Symmetrical around $\boldsymbol{\mu_X}$

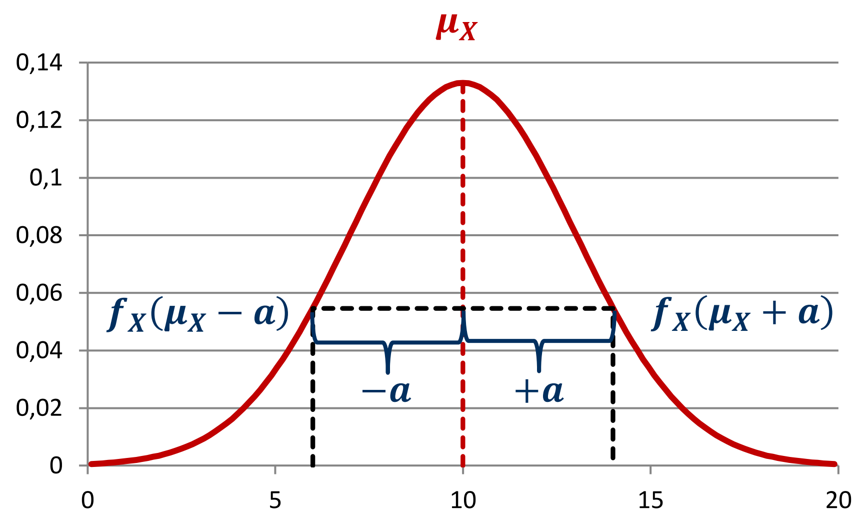
  $f_X(\mu_X - a) = f_X(\mu_X + a)$

  $P(X > \mu_X + a) = P(X < \mu_X - a)$

  $P(X < \mu_X + a) = P(X > \mu_X - a)$

  $P(X < \mu_X) = P(X > \mu_X) = 0{,}5$

  Median: $\boldsymbol{Md(X) = \mu_X} =$ the point where the left side has probability 0,5 and the right side has probability 0,5



Stockholm University

Linear combination of normal dist.

- If $\boldsymbol{X}$ is normally dist., then $\boldsymbol{Y} = \boldsymbol{a} + \boldsymbol{bX}$ is **normally distributed**

$$\boldsymbol{\mu_Y} = E(Y) = E(a + bX) = \boldsymbol{a} + \boldsymbol{b\mu_X}$$

$$\boldsymbol{\sigma_Y^2} = Var(Y) = Var(a + bX) = \boldsymbol{b^2\sigma_X^2}$$

$$\boxed{\begin{aligned}&\boldsymbol{X \sim N\left(\mu_X, \sigma_X^2\right)}\\&\Rightarrow \boldsymbol{Y \sim N(a + b\mu_X, b^2\sigma_X^2)}\end{aligned}}$$

Standardized normal distribution

- Special case of linear combination

$$\boxed{\boldsymbol{Z} = \frac{\boldsymbol{X - \mu_X}}{\boldsymbol{\sigma_X}}} = -\frac{\mu_X}{\sigma_X} + \frac{1}{\sigma_X} \cdot X$$

$$\boldsymbol{\mu_Z} = -\frac{\mu_X}{\sqrt{\sigma_X^2}} + \frac{1}{\sqrt{\sigma_X^2}} \cdot \mu_X = \boldsymbol{0}$$

$$\boldsymbol{\sigma_Z^2} = \boldsymbol{b^2}\sigma_X^2 = \frac{1}{\sigma_X^2} \cdot \sigma_X^2 = \boldsymbol{1}$$

$$\boxed{\boldsymbol{X \sim N\left(\mu_X, \sigma_X^2\right) \Rightarrow Z \sim N(0, 1)}}$$

Stockholm University

Calculations of probabilities

- **Cannot be done by hand!**
- There does not exist a simple formula for $F_X(x)$
- Advanced numerical methods are required.

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

- **Alternatives**: tables or computers
- Swedish Excel:

  $f_X(t)$ write ”=NORM.FÖRD(x;mu;sigma;0)”

  $F_X(x)$ write ”=NORM.FÖRD(x;mu;sigma;1)”

Substitute numbers in place of x, μ and sigma (NOTE! not sigma squared!)

English language Excel: NORMDIST(x,mu,sigma,0)

Stockholm University

Step 1: Standardize

- Standardize $\boldsymbol{X}$ to $\boldsymbol{Z}$
- If $\boldsymbol{X} \sim \boldsymbol{N}(\boldsymbol{\mu_X}, \boldsymbol{\sigma_X^2})$ then $\boldsymbol{Z} = \frac{\boldsymbol{X} - \boldsymbol{\mu_X}}{\boldsymbol{\sigma_X}} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{1})$
- If you want the probability $\boldsymbol{P}(\boldsymbol{X} \leq \boldsymbol{x})$, then use:

$$\boldsymbol{P}(\boldsymbol{X} \leq \boldsymbol{x}) = \boldsymbol{P}\left(\boldsymbol{Z} \leq \frac{\boldsymbol{x} - \boldsymbol{\mu_X}}{\boldsymbol{\sigma_X}}\right)$$

Write $\boldsymbol{Z}$ instead of $\boldsymbol{X}$ and exchange $\boldsymbol{x}$ for $\boldsymbol{z} = \frac{\boldsymbol{x} - \boldsymbol{\mu_X}}{\boldsymbol{\sigma_X}}$

Example

$X \sim N(10, 9)$ **NOTE!** $\sigma_X^2 = 9$ so $\sigma_X = 3$

- $P(X \le 13) = P\left(Z \le \frac{13-10}{3}\right) = P\left(Z \le \frac{3}{3}\right) = P(Z \le 1)$
- $P(X \le 7) = P\left(Z \le \frac{7-10}{3}\right) = P\left(Z \le -\frac{3}{3}\right) = P(Z \le -1)$
- $P(4 \le X \le 13) = P\left(\frac{4-10}{3} \le Z \le \frac{13-10}{3}\right) = P\left(-\frac{6}{3} Z \le \frac{3}{3}\right) = P(-2 \le Z \le 1)$

$$= P(Z \le 1) - P(Z \le -2)$$

$$= F_x(1) - F_X(-2)$$

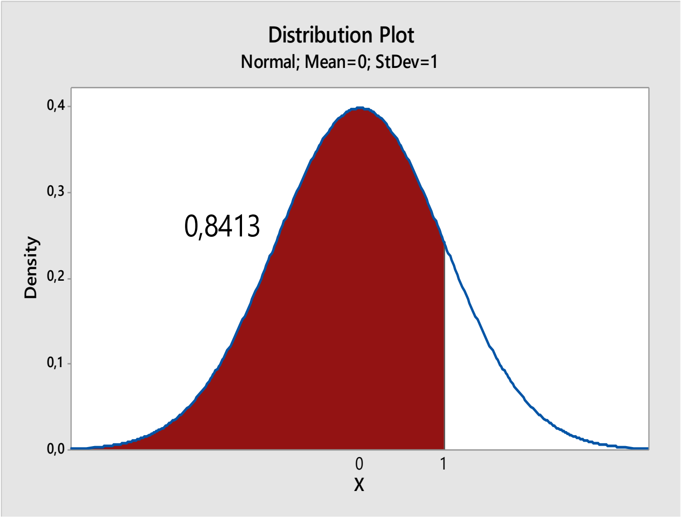
Stockholm University

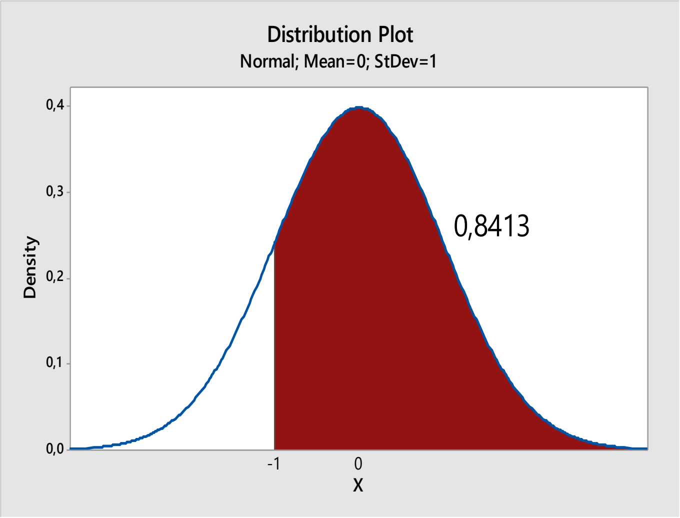
Step 2: Use the symmetry around $\mu_Z = 0$

- Of what form is the probability we seek?

$$P(Z \leq 1{,}00) = [\text{table}] = 0{,}8413$$

$$P(Z \geq -1{,}00) = P(Z \leq 1{,}00)$$





- You want it on the form: $\boxed{P(Z \leq z) = F_Z(z)}$ where z is positive

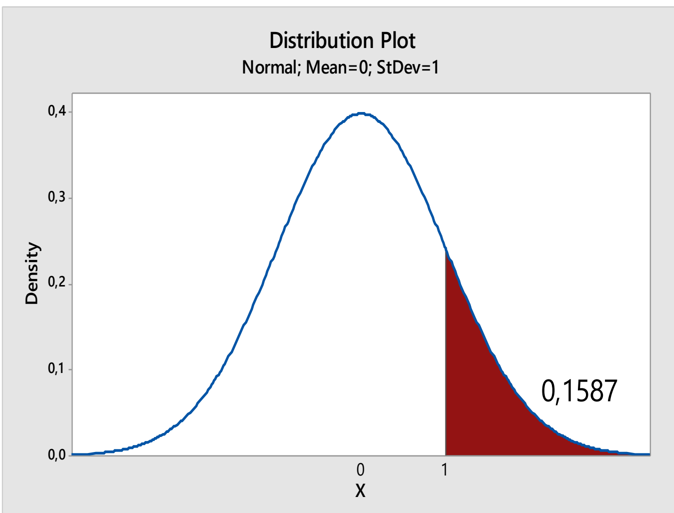
Stockholm University

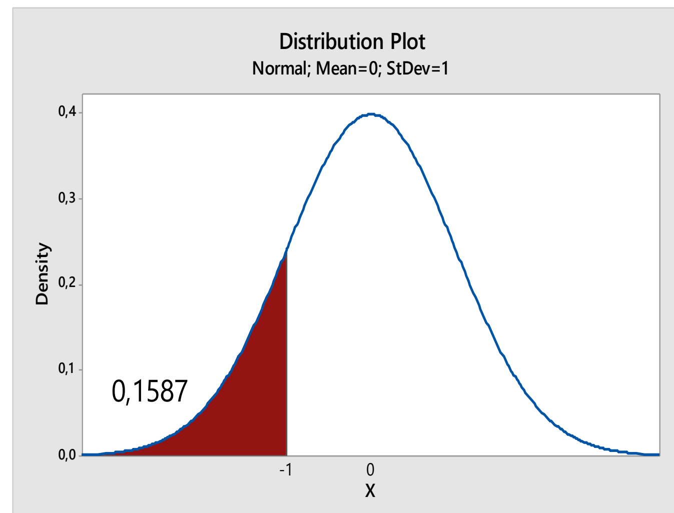
Steg 2: Use the symmetry around μ_Z, cont.

- Of what form is the probability we seek?

$$P(Z > 1,00) = 1 - P(Z \leq 1,00)$$

$$P(Z < -1,00) = P(Z > 1,00) = 1 - P(Z \leq 1,00)$$





- You want the form to be: $\boxed{P(Z \leq z) = F_Z(z)}$ where z is positive

Stockholm University

Example

$X \sim N(10, 9)$ **Notice!** $\sigma_X^2 = 9$ so $\sigma_X = 3$

positive number, 1

- $P(X \leq 13) = P(Z \leq 1) = F_x(1)$

positive number, 1

- $P(X \leq 7) = P(Z \leq -1) = P(Z \geq 1) = 1 - P(Z \leq 1) = 1 - F_x(1)$

- $P(4 \leq X \leq 13) = P(-2 \leq Z \leq 1) = P(Z \leq 1) - P(Z \leq -2)$

$$= P(Z \leq 1) - [1 - P(Z \leq 2)] = P(Z \leq 1) + P(Z \leq 2) - 1$$

$$= F_x(1) + F_X(2) - 1$$

positive numbers, 1 and 2

Stockholm University

Step 3: Use the table

- The table shows values for the **standardized** normal distribution with expected value = 0 och variance = 1;

$$\boldsymbol{Z \sim N(0, 1)}$$

  - **Standardize**

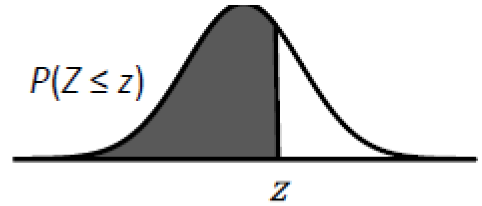
- The table gives cumulative probability $\boldsymbol{P(Z \le z) = F_Z(z)}$
  - **Use symmetry and reformulate the probabilities so that they are in the form $\boldsymbol{P(Z \le z)}$**

- And only positive values of $\boldsymbol{z}$, i.e. for $\boldsymbol{z \ge 0}$
  - **Positive values of $\boldsymbol{z}$**

Stockholm University

TABELL 1. Normalfördelningen, standardiserad

$\Phi(z) = P(Z \leq z)$ där $Z \in N(0, 1)$.



För negativa värden, utnyttja att $\Phi(-z) = 1 - \Phi(z)$.

The single digit place + the first decimal of z

The second decimal of z

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	[illegible]	[illegible]	[illegible]	[illegible]
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	[illegible]	[illegible]	[illegible]	[illegible]
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490
0,7	0,75804	0,76115	0,76424	0,76730	0,77035	0,77337	0,77637	0,77935	0,78230	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1,0	0,84134	0,84375	0,84614	0,84849	[illegible]	[illegible]	[illegible]	0,85769	0,85993	0,86214
1,1	0,86433	0,86650	0,86864	0,87076	[illegible]	[illegible]	[illegible]	0,87900	0,88100	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	[illegible]	[illegible]	[illegible]	0,89796	0,89973	0,90147
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574

i.e. $P(Z \leq 0,93)$

Exercise

The employees of an airline that traffics New York (JFK) note that the planes sometimes arrive a little early and sometimes a little late. Let $X =$ "difference between actual arrival time and planned arrival time" and suppose that X is normally distributed with expectation 12 minutes and standard deviation 8,6 minutes.

1. What is the probability that a flight arrives more than 30 minutes late, i.e. what is $P(X > 30)$?
2. What is the probability that a flights arrives early, i.e. $P(X < 0)$?
3. What is the probability that a flight is exactly 12 minutes late?

Stockholm University

Solution

1. Sought: $P(X > 30) =$

$$= [\textbf{standardize}] = P\left(Z > \frac{30-12}{8{,}6}\right) \approx P(Z > 2{,}09)$$

$$= [\textbf{substitute forms! draw!}] = 1 - P(Z < 2{,}09) = 1 - F_Z(2{,}09)$$

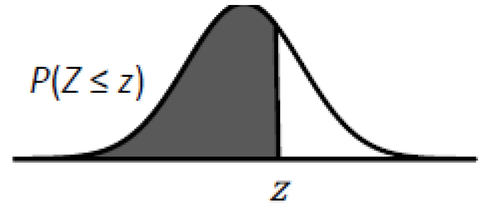
$$= [\textbf{use the table}] = 1 - 0{,}98169 = 0{,}01831$$

Answer: approximately 1,8 % probability

TABELL 1. Normalfördelningen, standardiserad

$\Phi(z) = P(Z \leq z)$ där $Z \in N(0, 1)$.

För negativa värden, utnyttja att $\Phi(-z) = 1 - \Phi(z)$.



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490
0,7	0,75804	0,76115	0,76424	0,76730	0,77035	0,77337	0,77637	0,77935	0,78230	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1,0	0,84134	0,84375	0,84614	0,84849	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92785	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408
1,6	0,94520	0,94630	0,94738	0,94845	0,94950	0,95053	0,95154	0,95254	0,95352	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,95994	0,96080	0,96164	0,96246	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,96784	0,96856	0,96926	0,96995	0,97062
1,9	0,97128	0,97193	0,97257	0,97320	0,97381	0,97441	0,97500	0,97558	0,97615	0,97670
2,0	0,97725	0,97778	0,97831	0,97882	0,97932	0,97982	0,98030	0,98077	0,98124	0,98169
2,1	0,98214	0,98257	0,98300	0,98341	0,98382	0,98422	0,98461	0,98500	0,98537	0,98574

Stockholm University

Solution

2. Sought: $P(X < 0) =$

$$= [\textbf{standardize}] = P\left(Z < \frac{0-12}{8{,}6}\right) \approx P(Z < -1{,}40)$$

$$= [\textbf{substitute forms! draw!}] = P(Z > 1{,}40) = 1 - P(Z < 1{,}40)$$

$$= 1 - F_Z(1{,}40)$$

$$= [\textbf{use the table}] = 1 - 0{,}91924 = 0{,}08076$$

Answer: approximately 8,1 % chance

3. Sought: $P(X = 12)$

$= 0$ (Remember: the area of a line is zero)

Summary

- Continuous random variables
  - can take any value in an interval S_X
  - S_X can be bounded or unbounded $(\pm\infty)$
  - cumulative distribution functions, $F_X(x) = P(X \leq x)$
  - events = subintervals of S_X
  - The values of the distribution function $f_X(x)$ are not probabilities.
  - Probability = the area under $f_X(x)$ within the subintervals
- Normal distribution, $\boldsymbol{X \sim N(\mu_X, \sigma_X^2)}$
  - standardized normal distribution $\boldsymbol{Z \sim N(0, 1)}$
  - calculation of probabilities using the table

Stockholm University

Next time

Two or more random variables at once

- Bivariate probability distributions
- Conditioning
- Covariance and correlation
- We save section 5.4 for L8
  - approximate a binomial with a normal distribution