Combinations

Order Matters $P_k^n = \binom{n}{k} = \frac{n!}{k!(n-k)!}$

Not matters $P(n,k) = \frac{n!}{(n-k)!}$

Discrete description

- Mean $\mu_x = \Sigma x_i \cdot P(X=x_i)$
- Varionce $\sigma^2 = \Sigma x_i^2 \cdot P(X=x_i) - \mu_x^2$
- Standard deviation: $\sigma_x = \sqrt{\sigma_x^2}$

Example ⟹

X	P(X)	$x \cdot P(x)$	$x^2 P(x)$
0	1/8	0	0
1	3/8	3/8	3/8
2	3/8	3/4	3/2
3	1/8	3/8	9/8
		$\mu = 3/2$	3

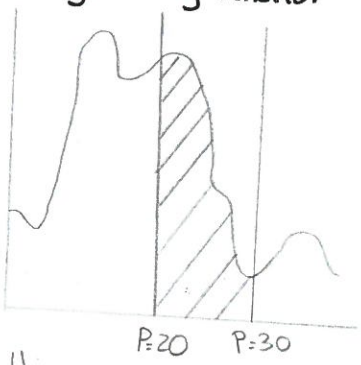
$\sigma = \sqrt{\sigma^2} = \sqrt{\Sigma x_i^2 \cdot P(X=x_i) - \mu^2} =$

$= \sqrt{3 - (3/2)^2} = 0'866$

Continuous random variables

Can take any value from the real numbers

Probability Density Function



$P(a<X<b) = \int_a^b f(x)\,dx$

Mean: $\mu_x = \int_{-\infty}^{\infty} x \cdot f(x)\,dx$

Variance: $\sigma^2 = \int_{-\infty}^{\infty} x^2 \cdot f(x)\,dx - \mu_x^2$

Standard Deviation: $\sigma_x = \sqrt{\sigma_x^2}$

$P(x=20)? = 0$

$P(20<x\leq 30)? = 0'31 \rightarrow 31\%$ to fall in the interval

↳ Example

Radioactive mass emits particles periodically. Time between two emissions is random.

Density func. is:

$$f(t) = \begin{cases} 0'12e^{-0'12t} & t>0 \\ 0 & t\leq 0 \end{cases}$$

a) Average time between emissions

$\mu_x = \int_0^{\infty} x \cdot f(x)\,dx = \int_0^{\infty} x \cdot 0'12e^{-0'12x} = 5$ with abs error of $3'8e{-}05$

"integrate(function(x){x*0'12*exp(-0'12*x)}, 0, Inf)"

b) Calculate the standard deviation between emitions

$\sigma^2 = \int_0^{\infty} x^2 \cdot 0'12 \cdot e^{-0'12x}\,dx - \mu_x^2 = 50 - 25 = 25 \rightarrow$ resul_integral ↘ " $\sqrt{\text{resul\_integral\$value} - 5\text{^}2}$

c) Determine the distribution function of time between emissions

$F(t) = \int_0^t 0'12e^{-0'12t}\,dt = [-e^{-0'12t}]_0^t = -e^{-0'12t} + 1$

$$F(t) \begin{cases} -e^{0'12t} + 1 & t>0 \\ 0 & t\leq 0 \end{cases}$$

d) Prob of the time when we have an emission in less than 10 secs

$F(t) = \int_0^{10} 0'12\,e^{-0'12t}\,dt = [-e^{-0'12t}]_0^{10} = -e^{-2} + 1 = 0'8647$

d) Calculate median time (t where $P(T<t) = 0'5$

$F(t) = -e^{-0'12t} + 1 = 0'5;\ 0'5 = e^{-0'12t};$

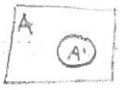
$0'5 = \frac{1}{e^{0'12t}};\ e^{0'12t}0'5 = 1;\ e^{0'12t} = 2$

$0'12t = \ln 2;\ t = \ln 2 / 0'12 = 3'43$

e) Calculate the 90th percentile

↳ The same but $= 0'9$

Properties



- Complement $P(A') = 1 - P(A)$
- Sure $E = \Omega$
- Impossible $P(\emptyset) = 0$
- Intersection $P(A \cap B) = P(A) \cdot P(B)$
- Union $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Difference $P(A \cap \overline{B}) = P(A) - P(A \cap B)$

Conditional probability

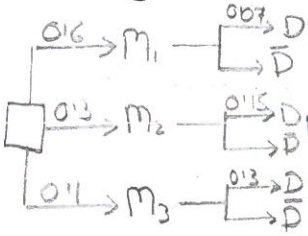
?? Given events A and B from the same probabilistic space (Ω, β, P) and $P(B) > 0$

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

Mutually exclusive and echaustive events

Prob of B? $\Rightarrow P(B) = \sum P(B \cap E_i) = \sum P(B|E_i) \cdot P(E_i)$

Probability tree



$P(M_1 \cap D) = P(D|M_1) \cdot P(M_1) = 0'07 \cdot 0'6 = 0'042$

$P(M_2 \cap D) = P(D|M_2) \cdot P(M_2) = 0'15 \cdot 0'3 = 0'045$

$P(M_3 \cap D) = P(D|M_3) \cdot P(M_3) = 0'13 \cdot 0'1 = 0'013$

$P(D) = \sum P(D|M_i) \cdot P(E_i) = 0'042 + 0'045 + 0'013 = 0'117$

Baye's rule

$$P(E_i|B) = \frac{P(E_i) \cdot P(B|E_i)}{P(B)}$$

Independent events

They are independent if knowing the outcome of one doesn't provide useful information for the other

Independence

If two events are not independent the conditional probability will be either greater or less than both unconditional probabilities

Unit 4. Random variables

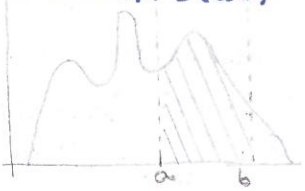
Random variables

Is a function from a sample space Ω into a set of real numbers (the result)

Always uppercase, and the values are lowercase ($X = x$). Random var X takes the value x

Types

- Discrete: Finite possible outcomes. Profit of a 1'5$ bet in roulette. Variable can be -1'5 (lose), 1'5 (win)
- Continuous: Random variable with possible outcomes in real number interval



Probability density function - Discrete (pdf)

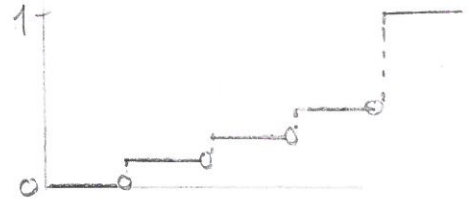
Function for which each outcome has a probability

$P(x) = P(X = x) \; \forall x \Rightarrow$ 1. $P(x) \geq 0$, 2. $\sum p(x) = 1$

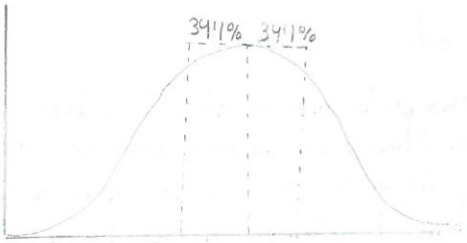
$P(a < X < b) = \int_a^b f(x)\,dx$

Cumulative distribution function (cdf)

$0 \leq F(x) \leq 1$ if $x \leq y$; $F(x) \leq F(y)$



Normal distribution:



- Most values are around center
- Mean and median are equal
- It's unimodal
- Symmetric (no skew)

Test for normality

- Shapiro-Wilk's test
  - If $n > 3$ and < 5000
- Kolmogorov-Smirnoff (not a normality test at all)
  - Lilliefor's correction

A p value < 0.05 means that the distribution is not normal

~ Standarization ~

2 students → 8.2

Engineering: mean 6.9, sd: 2.1

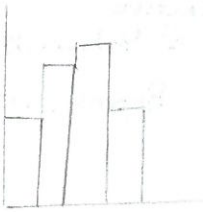
Law-school: mean 8.5, sd: 1.5

$$\text{Engineer} = \frac{8.2 - 6.9}{2.1} = 0.619 \quad \text{Law} = \frac{8.2 - 8.5}{1.5} = -0.2$$

→ Better Student

Graphs

- Pie chart 
- Histogram

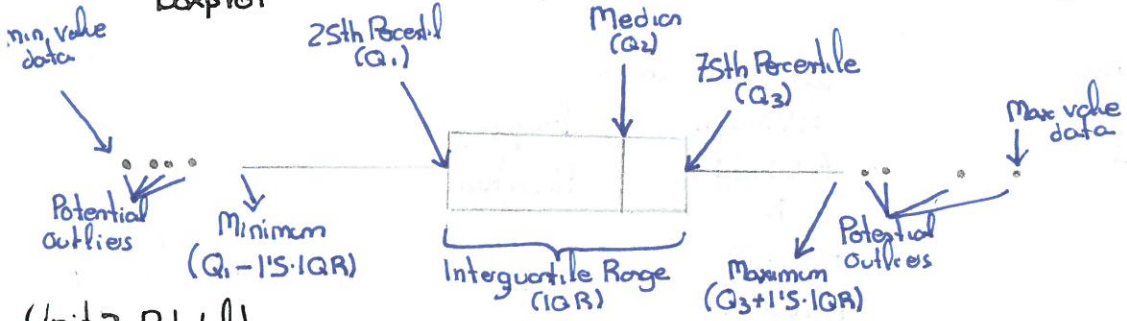


  - Used for quantitative continuous variables
  - Values included in intervals (bars)
  - Bars are called bins and can be uniforms or not

$$\text{Rectangle height} = \frac{\text{bin freq.}}{\text{bin width}}$$

- Boxplot

??



Unit 3. Probability

Probability

Helps quantify otcomes that cannot be predicted (random events)

Prob. of an event is the favorable cases / possible cases (Laplace's law)

Experiment

Action or process that generates observations

Random experiment

All possible results are known beforehand, the outcome is unknown

It can be repeated in identical conditions

Concepts

- Sample Space (Ω) Set of all possible outcomes
  - Ω discrete: Finite set of outcomes
  - Ω continuous: Interval of real numbers
- Event β: Any subset of Ω
- Probability P: Gives an event a value between 0 and 1

Three of them constitute the probability space (Ω, β, P)

Statistics

Unit 1. Introduction to data

Statistics: Transforms data into useful information
- Descriptive Statistics: Collect, summarize and analyze data
- Inferential Statistics: Generalize conclusions based on a "small" sample

Population: Set of items that share one property that is the set of the statistical analysis

Sample: Its a fragment of the population
- Representative Sample:
  - Simple random sample: Has an equal chance of being included in the final sample
  - Stratified Sampling: Strata are chosen so that subjects have any number of characteristics in common

Selection within strata is done using a second method (generally simple random sample)
- Cluster sampling: Population divided into cluster. Divided groups for many samples and some objective

Random variables: Characteristic being studied in a statistical problem

Must be: atomic (one thing measured), exhaustive (all possible values), unique options (one option per category)

Types: Qualitative
- Yes/No → Nominal
- Non-ordered → Nominal
- Ordered → Ordinal (Tumor class. T1, T2, T3)

Quantitative
- Height and weight → Continuous
- Discrete: N° of children

Types of studies: Representative Sampling // Observational Studies (No intervention) // Experimental Studies (Intervention)

Unit 2. Descriptive analysis

Steps of descriptive analysis:
1. Define the variables
2. Collect appropriate data
3. Organize data in tables
4. Visualize with charts
5. Reach conclusions

Categorical data

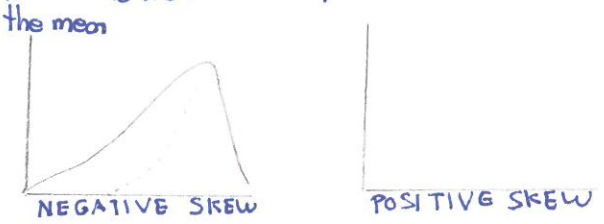
f_i Value(1)	Frequency
$\frac{f_i \text{ Value(1)}}{N}$	Proportion
$\frac{f_i \text{ Value(1)}}{f_i \text{ Value(2)}}$	Odds

Numeric data

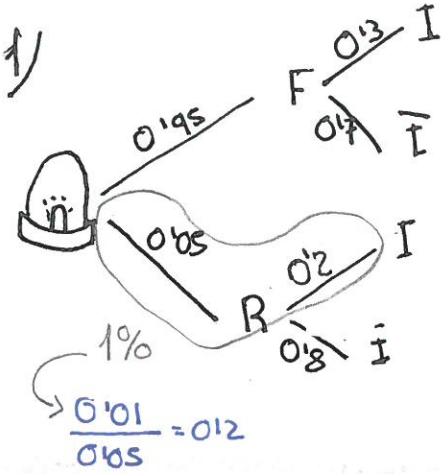
Central tendency
- Mean: Average Value
  - Affected by extremes
- Median: At the 50/50
  - Not affected by extremes
- Mode: Most repeated value
  - From 0 to ∞

- Trimmed mean
  - Resistant to extreme values
  1°. % to trim (p)
  2°. Remove (p/2)% of extremes
  3° Mean

Dispersion
- Range = X largest − X smallest
- Variance: Mean value of the distance of each element to the mean squared
  ↳ $\sigma^2 = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 \cdot n}{n}$
- Standard Deviation = $\sqrt{\sigma^2}$
- Pearson's Coefficient of Variation $V_p = \frac{\sigma}{|\bar{x}|}$
- Quantiles: What divides in 4 blocks an observation 1st → 25%, 2nd → 50%, 3rd → 75%
- Percentile: Value that leaves below no more than p % of the sample and no more than 100 − p% above
- Kurtosis: Describe the distribution of observed data around the mean
- ?? Skewness: Pull the mean and median, but the median will be more affected
  - Negative: lower mean
  - Positive: bigger mean



1/



a) Prob de que sea de tipo Falsa, no Inusual

0'7

b) Inusual y falsa alarma

$0'195 \cdot 0'3 = 0'285$

c) 10 falsas alarmas e independientes. Prob de que 4 menos sean inusuales

pbinom(3,10,0'3) = 0'6496

d) Prob de que sea una actividad inusual

$$\frac{0'195 \cdot 0'3 + 0'05 \cdot 0'2}{0'195 \cdot 0'3 + 0'05 \cdot 0'2 + 0'195 \cdot 0'7 + 0'05 \cdot 0'8} = 0'295$$

4/ IA genera texto

Dist. normal media 85 desv. típica 1'67

a) Prob. precisión obtenida entre 88 y 92

$Z_1 = \frac{(x-\mu)}{\sigma} = \frac{88-85}{1'67} = 1'79$ $\quad P(1'79 < Z < 4'19) =$

$Z_2 = \frac{(x-\mu)}{\sigma} = \frac{92-85}{1'67} = 4'191616$ $\quad$ pnorm(4'191916)-pnorm(1'79) = 0'036720

b) Calcular media de precisión obtenida en 10 textos indep. sea mayor de 85'6

Nos dan una muestra, cambiamos con

$\frac{\text{dev\_tip}}{\sqrt{10}} = \frac{1'67}{\sqrt{10}} = 0'52$

pbinom(85'6, 85, 0'52) => 1 − ...

Da menor que

d) Calcula media de precisión obtenida en 30 textos, prob de que sea menor de 85'5

$\frac{\text{dev\_tip}}{\sqrt{30}} = \frac{1'67}{\sqrt{30}} = 0'3048$

pbinom(85'5, 85, 0'3048)

c) 20 textos al azar e indep. prob de que menos de 7 tengan una precisión menor de 83'33

pnorm(83'33, 85, 1'67) = 0'1586553 # Prob for one text having accuracy less than 83'33

pbinom(6, 20, 0'15865) = 0'97087 # For fewer than 7 out of 20

3) Copilot usa poisson con una media de 53'33

a) Prob de más de 21 trabajos en una hora

Mean(landba) = 53'33

We want P(X > 21) = P(X ≤ 21)

→ 1 - ppois(21, 53'33) = 0'999

b) Nº de trabajos cuando el percentil es del 95%?

qpois(0'95, 53'33) = 66

c) En una hora ha procesado menos de 3 trabajos ¿Probabilidad de que haya procesado más de un trabajo?

Probability Unit 3

Ex. 1.) 36 possible results for 2 fair dice.

The first dice is a 4, chances of both dice equals 8?

1/6

Sum equals 8 ⇒ E $P(E \cap F) = 1/36$

Num 4 in first roll ⇒ F $P(F) = 1/6$

$$P(E|F) = \frac{P(F|E) \cdot P(E)}{P(F)} = \frac{1/36}{6/36} = 1/6$$

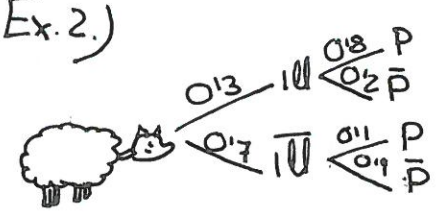
Are these events independent?

$$P(F|E) = \frac{P(E|E) \cdot P(F)}{P(E)} = \frac{1/6 \cdot 1/6}{5/36} = 1/5$$

$P(E|F) \neq P(F|E)$

↳ NOT independent

Ex. 2.)



a) $P(\text{Not. Ill}) = 0'7$

b) $P(\text{P not. Ill})$ Conditional $= 0'1$

c) $P(\text{P not. Ill}) = 0'7 \cdot 0'1 = 0'07$

d) Prop of sheeps with past. $= P(\text{Ill} \cap P) + P(\bar{\text{I}} \cap P) = 0'13 \cdot 0'8 + 0'7 \cdot 0'1 = 0'31$

e) Prob being Ill, knowing it's possit. $= \frac{P(\text{Ill} \cap P)}{d)} = \frac{0'13 \cdot 0'8}{0'31} = 0'745$

f) Prob not. Ill, knowing $\bar{P}$: $\frac{P(\bar{\text{I}} \cap \bar{P})}{P(\bar{\text{I}} \cap \bar{P}) + P(\text{Ill} \cap \bar{P})} = \frac{0'7 \cdot 0'9}{0'7 \cdot 0'9 + 0'3 \cdot 0'2} = 0'914$

Binomial distribution → Replacement!, Always same number, Independent too

$X \sim Bi$

20 questions with 4 answers

$X \sim Bi(20, 0'8)$ →

3 heads in 12 flips of coin

En R → pbinom(3, 12, 0'5)

$P(x \leq 3)$

↳ so not 3, 2!!

Random var. . Unit 4

$$f(x)=\begin{cases}\frac{x^2}{3} & -1<x<2\\ 0 & o.w\end{cases}$$

$P(0<x\leq 1)$?

$$\int_{-1}^{2}\frac{x^2}{3}dx=\left[\frac{x^3}{9}\right]_1^2=\frac{8}{9}-\frac{-1}{9}=1$$

$$P(0<x\leq 1)=P(0<x<1)+P(x=1)=\int_0^1\frac{x^2}{3}+0=\left[\frac{x^3}{9}\right]_0^1=1/9$$

PDF

60% babies develop illness. we sample?. Prob dist of x

DD DN DN NN

0'6·0'6 0'6·0'4 0'4·0'6 0'4·0'4

$P(X=x)=\binom{2}{x}\cdot 0'6^x\,0'4^{2-x}$

CDF

X	P(X)	F(X)
0	1/8	1/8
1	3/8	4/8
2	3/8	7/8
3	1/8	

$F(0)=P(X\leq 0)=P(X=0)=1/8$

$F(1)=P(X\leq 1)=P(X=1)+P(X=0)=3/8+1/8=4/8$

$F(2)=P(X\leq 2)=P(X=2)+P(X\leq 1)=3/8+4/8=7/8$

PDF VS. CDF

Particle function with a waffo with a loge particle, with a prob of 0'1 containing it

PDF?

X=x	P(X=x)
1	0'1 =
2	0'1·0'9 = 0'09
3	0'1·0'9² = 0'081
4	0'1·0'9³ = 0'0729
...	...

CDF

x	F(x)
x<1	0
1≤x2	0'1
2≤x<3	0'1+0'09 = 0'19
3≤x<4	0'1+0'09+0'081 = 0'271

Combinations and PDF & CDF

1/8 parts contains 3 defective. We take 3 random without replacement ↳5 good

Order does not matter

$P(8,2)=\frac{8}{(8-2)!}=8\cdot 7=56$ possible cases

PDF

X	0	1	2
P(x)	20/56	30/56	6/56

5·4 3·5+5·3 3·2

1bad 1good

CDF

X	0	1	2
F(X)	20/56	50/56	56/56

Discrete description

X	0	1	2	3	4	
$P_1(x)$	0'2	0'2	0'3	0'1	0'1	= 0'9
$P_2(x)$	0'1	0'3	0'3	0'2	0'2	= 1'1
$P_3(x)$	0'4	0'2	0'1	0'2	0'1	= 1 → ✓

Correct PDF

$\Sigma P(x)=1$

μ? σ?

X	P(x)	x·P(x)	x²P(x)
0	0'1	0	0
1	0'2	0'2	0'2
2	0'4	0'8	1'6
3	0'2	0'6	1'8
4	0'1	0'4	1'6
		Σ $\mu=2$	Σ $5'2=\sigma^2$

$\sigma=\sqrt{\sigma^2-\mu^2}$;

$\sigma=\sqrt{5'2^2-2^2}=1'095$

Unit 5. Univariate probability distributions

Types of discrete distribution

- Uniform distribution — Rolling a dice
- Binomial distribution — N° of broken chips
- Poisson distribution — N° of machines working

Probability distribution

Discrete	Continuous
- Uniform	- Uniform
- Binomial	- Exponential
- Poisson	- Normal

Uniform

→ Every possible value has the same probability

$P_x = P(X=k) = 1/k \qquad \mu = \frac{1}{k}\cdot\sum x_i \qquad \sigma^2 = \frac{1}{k}\sum x_i^2 - \mu^2$

Example ⇓

Light bulb 40W, 60W, 75W and 100W. Have same chances of being picked. Mean and variance

$\mu = \frac{40+60+75+100}{4} = 68'75$

$\sigma^2 = \frac{40^2+60^2+75^2+100^2}{4} - 68'72^2 = 479'6875$

Binomial

→ Trial with 2 possible outcomes

Trials independents

$X \sim B(n,p) \qquad P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad \mu = n\cdot p \qquad \sigma^2 = n\cdot p\cdot(1-p)$

⇓ Example

Chip with random pattern. 16 bits of 0s and 1. 0'6 prob for 1s, 0'4 for 0's

a) Prob for every bit = 1?

$P(X=10) = \binom{10}{10} p^{10}(1-p)^{n-x} = P(X=10) = \binom{10}{10} 0'6^{10}\cdot 0'4^0 = \frac{10!}{10!(10-10)!}\cdot 0'6^{10} = 0'006$

b) Prob for every bit = 0?

. . .

c) For 5 zeros and 5 ones

. . . $= P(X=10) = \binom{10}{10} 0'6^0\cdot 0'4^{10} = 0'000410\ldots$

$P(X=5) = \binom{10}{5} 0'6^5\cdot 0'4^5 = \frac{10!}{5!\,5!} 0'6^5\cdot 0'4^5 = 0'2006\ldots$

d) Prob. having less than 2 bits with a 1?

$P(X\leq 1) = \sum \binom{10}{i} 0'6^i\cdot 0'4^{10-i} = 0'00167 + 0'000104 = 0'001782$

e) Prob having more than 2 bits with a 1?

$1 - P(X\leq 2) = 1 - \sum_{i=0}^{2} \binom{10}{i} 0'6^i\cdot 0'4^{10-i} = 1 - 0'01124455 = 0'98877$

Poisson

X represents the n° of successes in an interval time or geographical region

Independent to other region or interval

Example ⇓

$P(X=x) = \frac{e^{-\lambda}\cdot\lambda^x}{x!}; \quad x = 1,2,3\ldots \qquad \mu = \lambda \qquad \sigma^2 = \lambda$

N° of messages into a server follows a poisson distribution of $\lambda = 0'3t$

a) Three messages arriving in exactly 10 seconds

$P(X=3) = \frac{e^{-0'3\cdot 10}\cdot(0'3\cdot 10)^x}{x!} = \frac{e^{-3}\cdot 3^3}{3!} = 0'224$

b) At most one message in 20 secs

$P(X\leq 1) = \sum \frac{e^{-0'3\cdot 20}\cdot(0'3\cdot 20)^i}{i!} = e^{-0'3\cdot 20} + e^{-0'3\cdot 20}\cdot 6 = 0'01735$

Poisson to binomial

Only if $p<0.1$ and $n \cdot p < 5$

↳ Example

prob of wrong bit = 0.001

a) Prob of at least happening once in a package of 400 bits

$X \sim B(n;p) \rightarrow X \sim B(400, 0.001)$

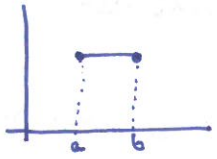
Binomial $\rightarrow P(X \geq 1) = 1 - P(X=0) = 0.32\,948$ } so close

Poisson $\rightarrow P(X \geq 1) = 1 - P(X=0) = 0.32\,97$

Continuous probability distribution

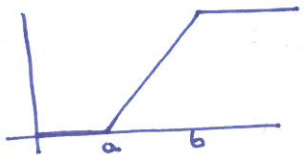
Density function

$$f(x) = \begin{cases} 0 & x<a \\ \frac{1}{b-a} & a \leq x \leq b \\ 0 & x>b \end{cases}$$



Distribution function

$$f(x) \begin{cases} 0 & x<a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x>b \end{cases}$$



Example ↳ $\mu = \frac{a+b}{2}$; $\sigma^2 = \frac{(b-a)^2}{12}$; $\sigma = \frac{b-a}{\sqrt{12}}$

Follows a cont. uniform. distrib with $a=2, b=7$

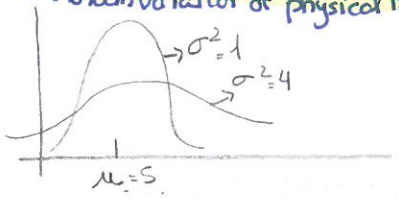
1) Calculate $P(X \geq 4)$

$P(X \geq 4) = 1 - p(X<4) = 1 - F(4)$

Normal

→ Many numerical populations have distributions that can be approximated with the normal distribution

→ Random variation of physical measurements



$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \qquad -\infty < x < \infty$$

Denoted as $N(\mu, \sigma)$

When $\mu=0$, $\sigma=1$

$$Z = \frac{X-\mu}{\sigma}$$

Soft drink discharges 200 ml per cup. Normally distributed with sd = 15 ml

1) fraction of cups with more than 224 ml

$$Z = \frac{X-\mu}{\sigma} = \frac{224-200}{15} = 1.6$$