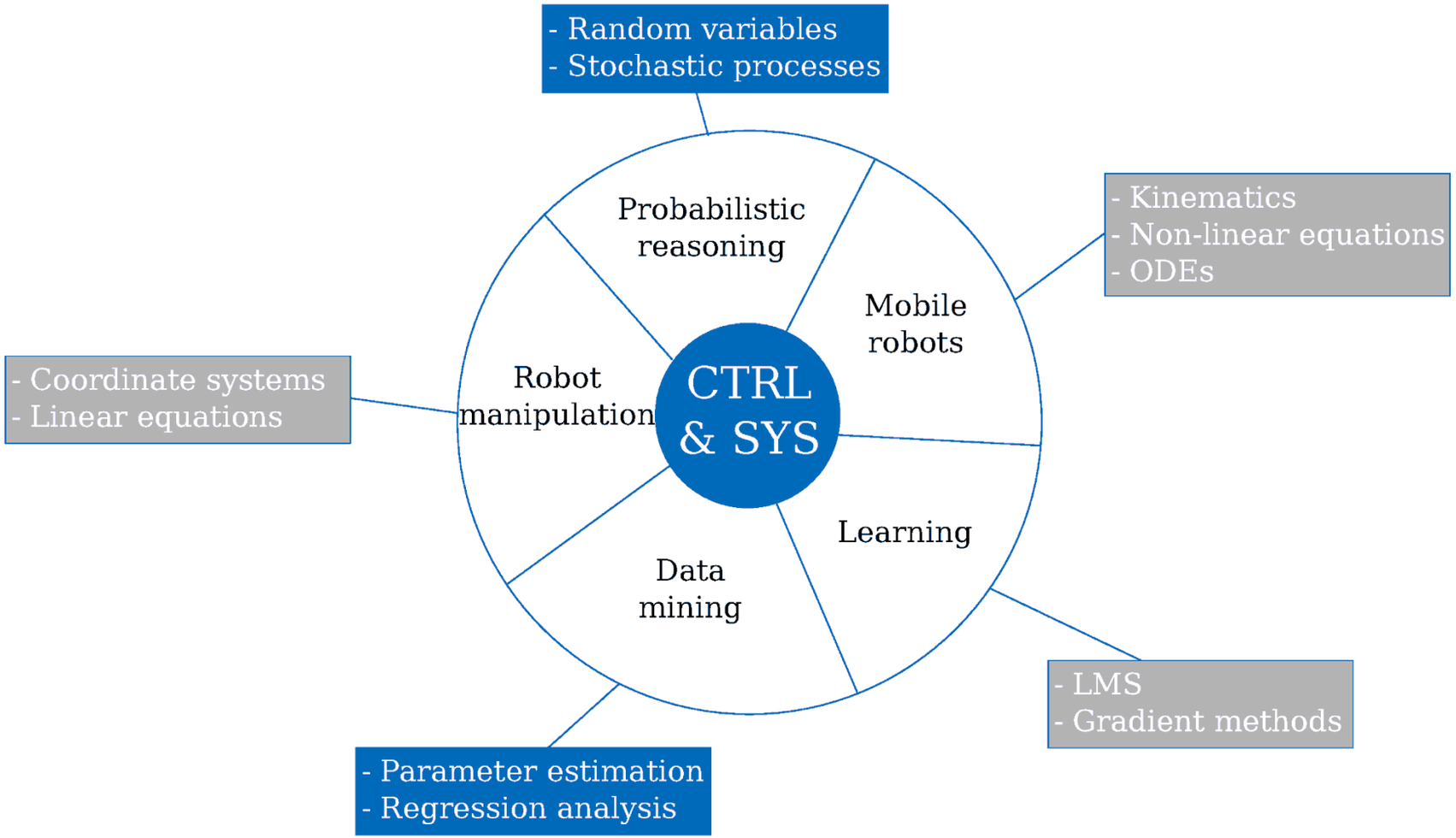
Pre-Lecture: What to know

- You need to be aware of the definition of probability and what it entails
  - Kolmogorov's axioms
    - Probabilities are functions on event sets
  - Conditional probabilities
  - Joint probabilities
  - Independence and conditional independence
- If you have heard it already, review
  - what a random variable is
  - what the expected value and the empirical mean (or average) are
  - what the variance and empirical variance are

Sensor Models



Bayes' theorem and sensor models

→ Random variables
- → Discrete and continuous random variables
- → Probability distributions and densities
- → Expected value, variance
- → Relation to measurement processes

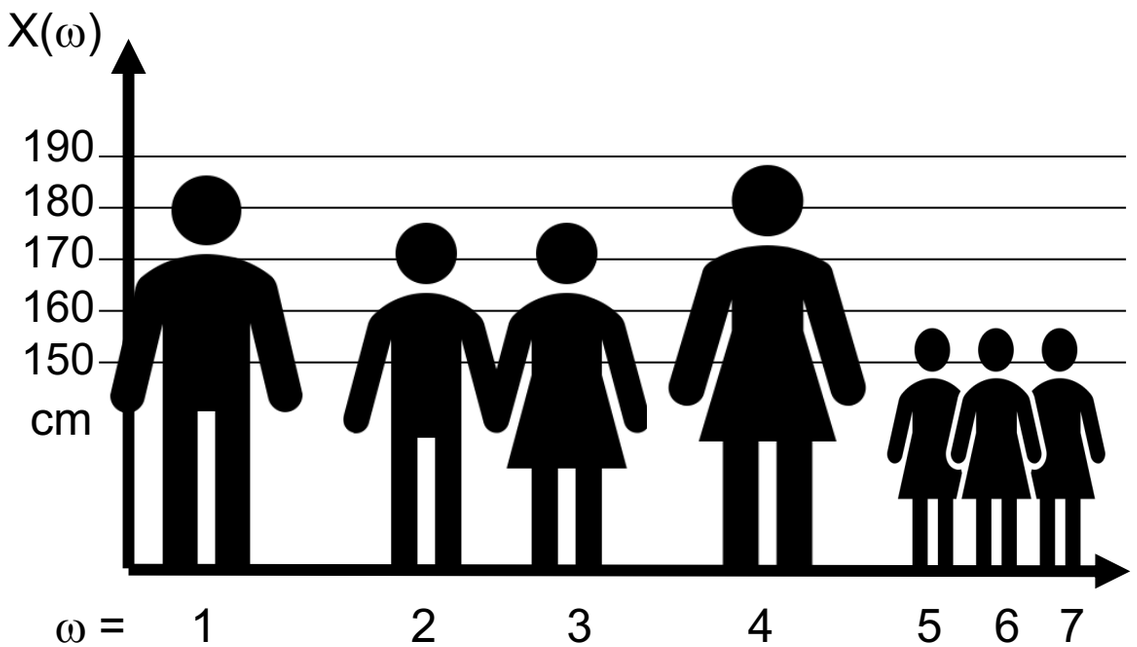
Statistical independence
- Joint distributions
- Correlation and covariance
- Examples leading towards Kalman filter

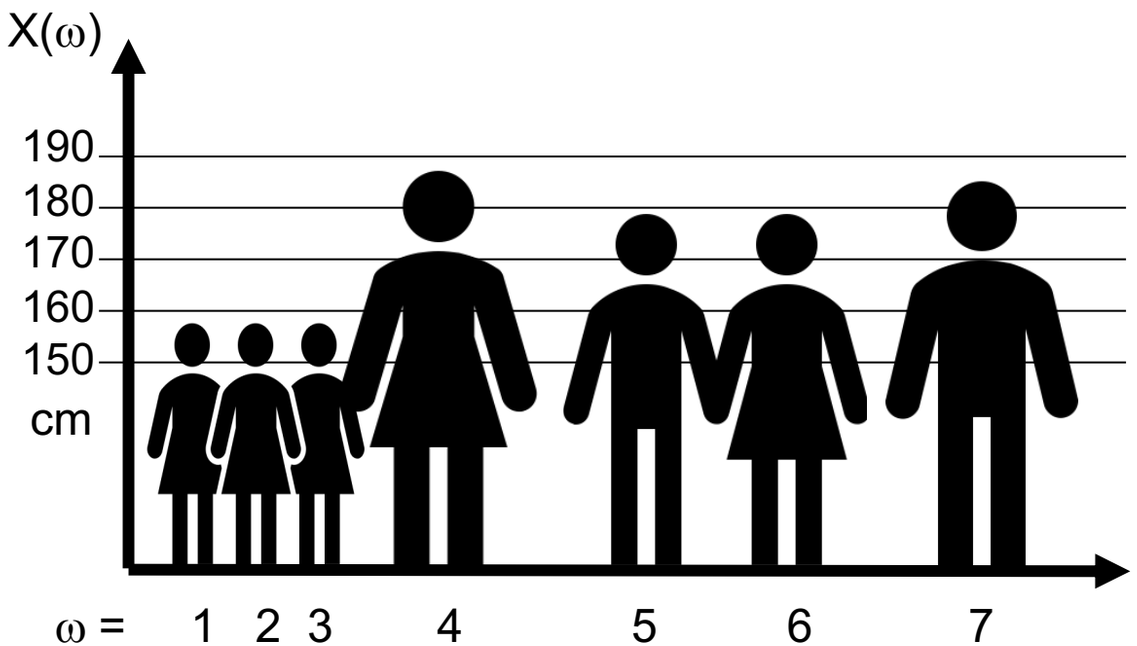
Random Variables

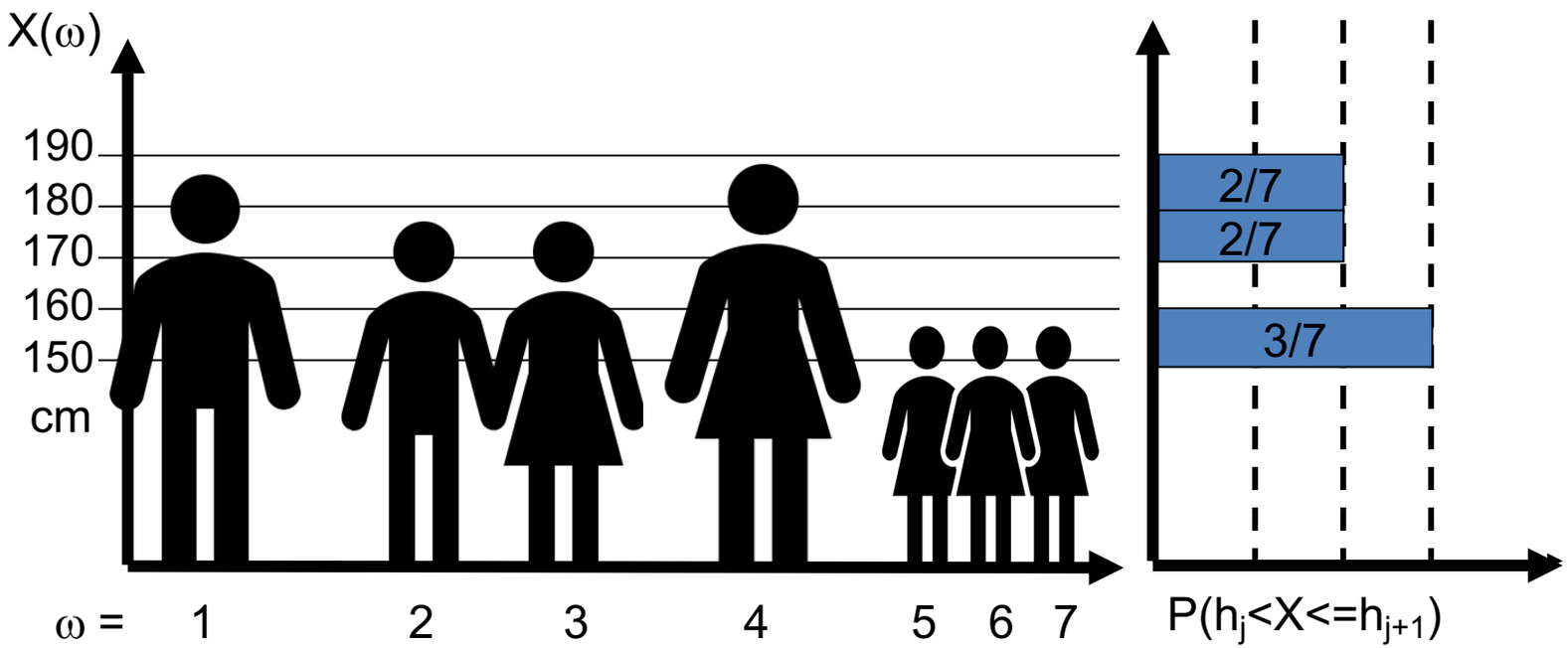
Random variables

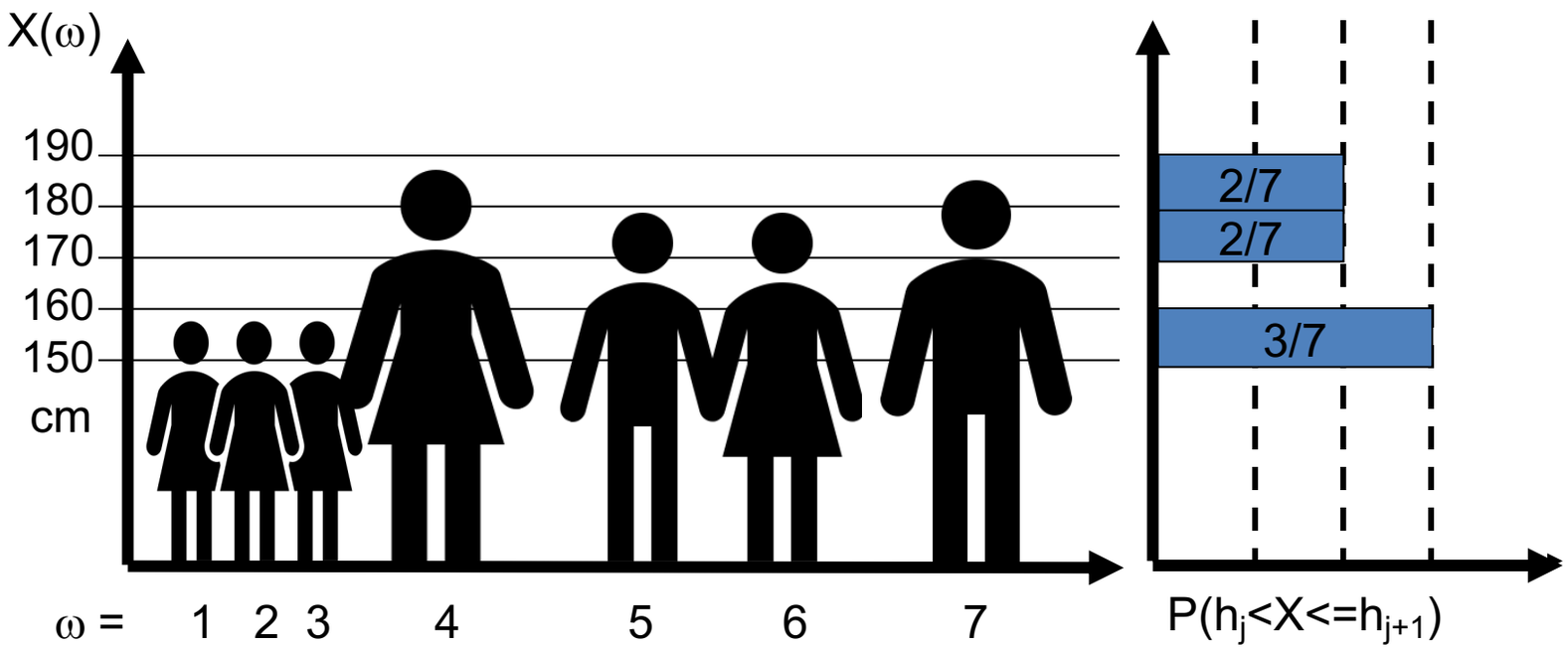
A **random variable** $X(\omega)$ is a function that takes on real-numbered values at random

Random variables can be seen as functions $\omega \to X(\omega) \subseteq \mathbb{R}$, namely as functions of ω, the outcomes of some random experiment, to a subset of the real numbers

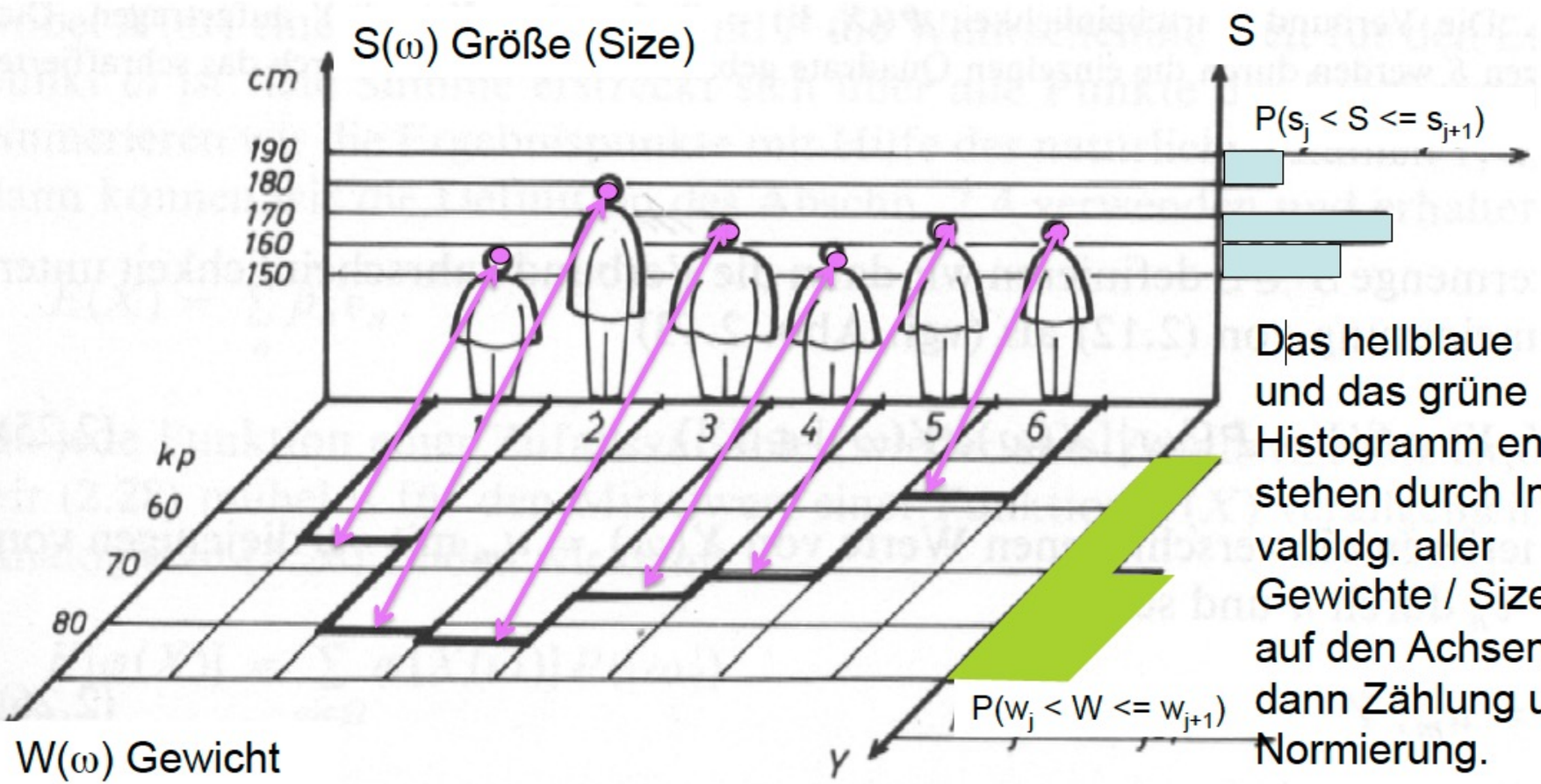

X(ω)
190
180
170
160
150
cm
ω =
1
2
3
4
5
6
7


X(ω)
190
180
170
160
150
cm
ω =
1
2
3
4
5
6
7


X(ω)
190
180
170
160
150
cm
ω = 1 2 3 4 5 6 7
2/7
2/7
3/7
P(h_j<X<=h_j+1)


X(ω)
190
180
170
160
150
cm
ω = 1 2 3 4 5 6 7
2/7
2/7
3/7
P(h_j<X<=h_j+1)

Two Random Variables
(on same probability space)



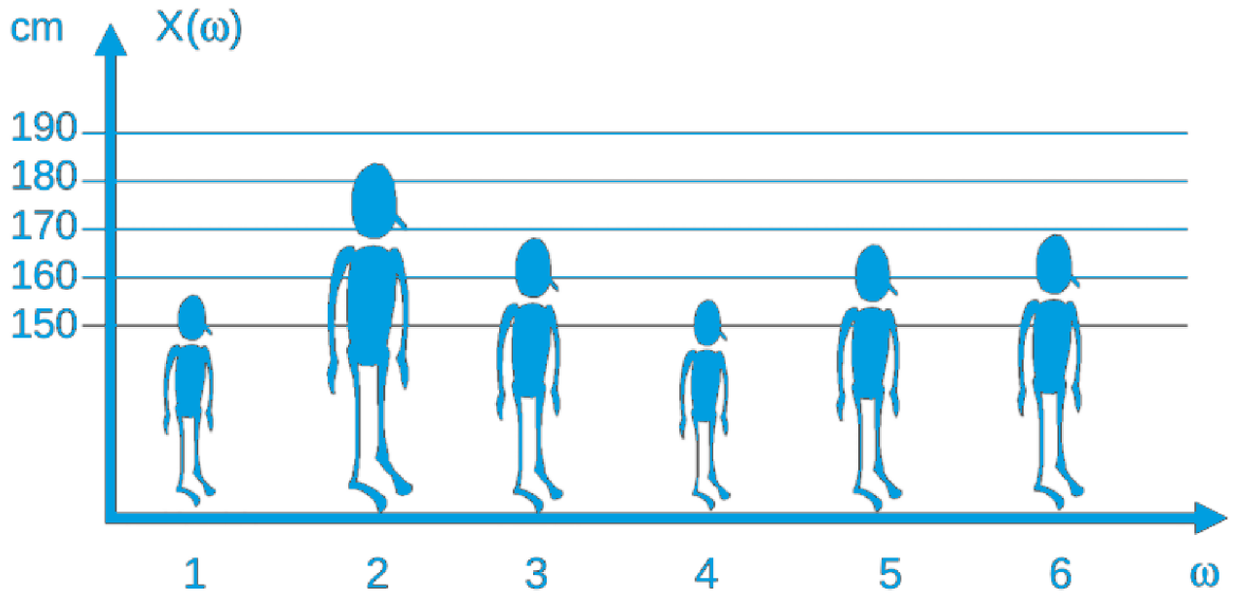
Das hellblaue und das grüne Histogramm entstehen durch Intervalbldg. aller Gewichte / Sizes auf den Achsen, dann Zählung und Normierung.

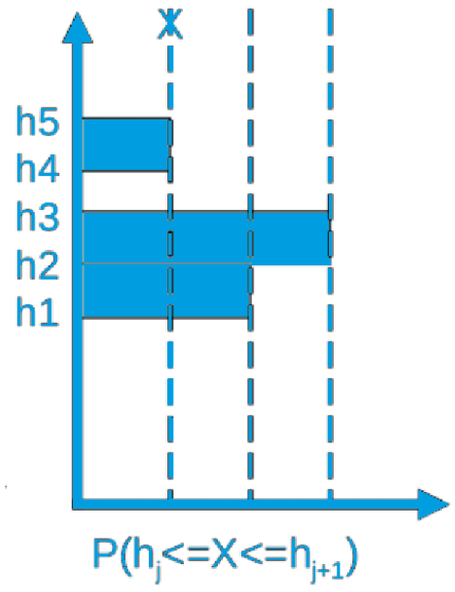
Probability Distributions $P_X(A) := P(X^{-1}(A))$

Invert the random variable mapping: look for sets of people such that

$$\{\omega \mid a < X(\omega) < b\}$$

Ask: $P(\{\omega \mid a < X(\omega) < b\}) = ?$ (abbreviated $P(a < X < b) = ?$)





For practical reasons, one usually does not record all observed sizes; instead, only classes of sizes (eg. between $h = 180$ and $h = 190$) are recorded

Histogram

Special Cases (1/2)

$P(X = x) = 0$ is possible for a size X that is not observed (e.g. $X = 177$)

If X is explicitly known (e.g. $X(\omega) = x$ for some ω), then $P(a < X < b) = P(X = x) = P\left(X \epsilon \{x\}\right)$ when a and b are close enough to x

When working with a class, one can replace the set of all class values by the mean value of the class

Special Cases (2/2)

We are given that X lies in the class $[h_i, h_{i+1}]$

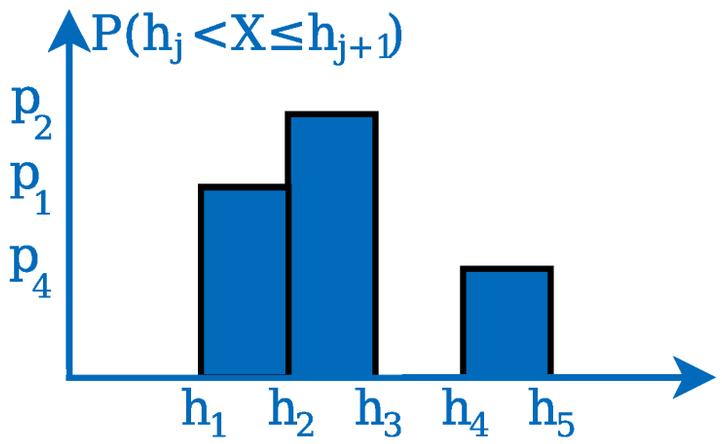
Each class can be replaced by its mean value v_i, whose probability is now concentrated at one point

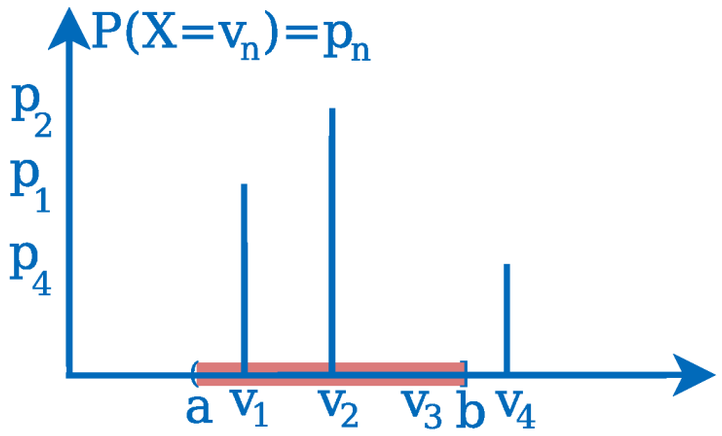
If $(a, b]$ is an arbitrary interval, then

$$P(a < X \leq b) = \sum_{a < v_n \leq b} p_n$$

More generally,

$$P(x \in A) = \sum_{v_n \in A} p_n$$



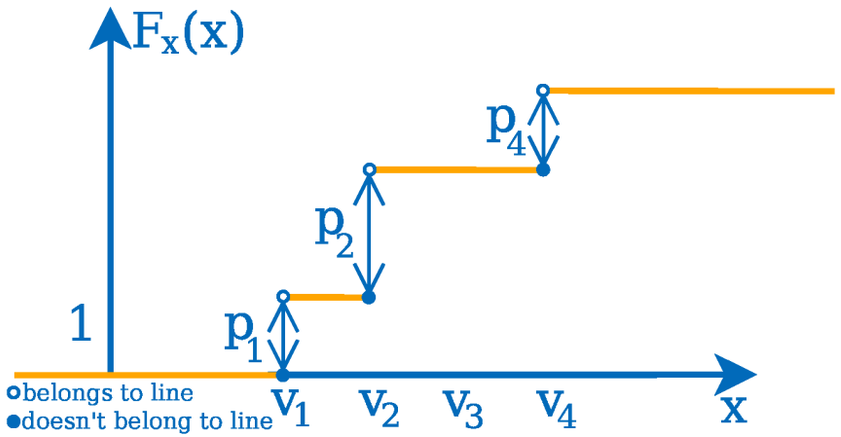


Cumulative Distribution Function (CDF)

Choose an interval $(-\infty, x]$

Run x from left to right

From the random variable X and measure P, get a **distribution function** $F_X(x)$



Cumulative distribution function

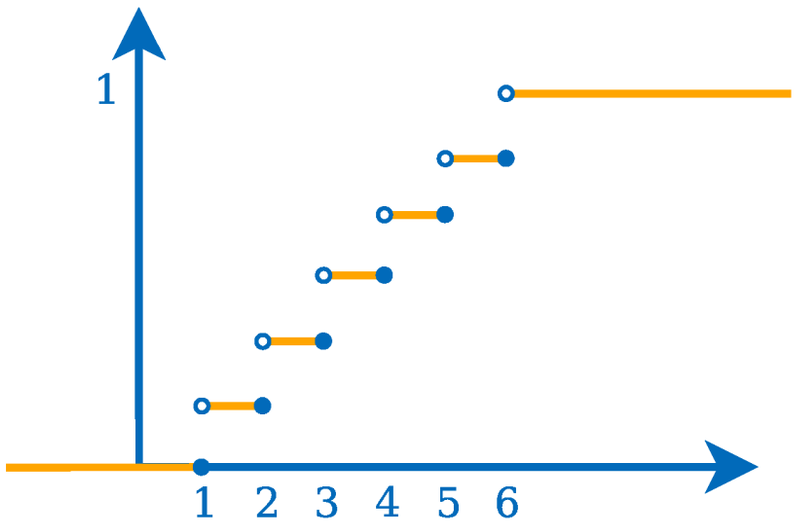
$$F_X(x) = P(X \leq x) = \sum_{v_n \leq x} p_n$$

CDF Examples

Throwing a die

$$P(x) = \frac{1}{6}, x \in \{1, 2, 3, 4, 5, 6\}$$

$$F_{\text{die}}(x) : p_n = \frac{1}{6}$$
$$v_n = 1, ..., 6$$
$$n = 1, ..., 6$$

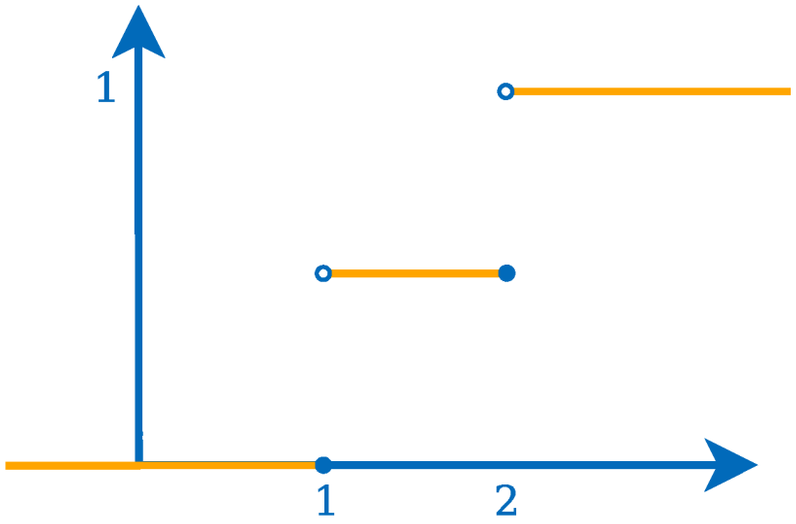


Throwing a coin

$$P(x) = \frac{1}{2}, x \in \{1, 2\}$$

$$F_{\text{coin}}(x) : X(\omega) = 1 \text{ for } \omega = \text{ heads}$$
$$= 2 \text{ for } \omega = \text{ tails}$$

$$p_n = \frac{1}{2}, \quad v_n = \frac{1}{2}$$



Exa. (discrete case): *chain* of Bernoulli experiments

- first: take a single *Bernoulli distribution*:
  a coin with tossing outcomes **`T,H`** where **`H`** has probability q and **`T`** has $(1-q)$.
- then: use n coins and count the number of **`H`**`'s`
  (note: no particular order!)
- the event space Ω is based on tuples, e.g. (n=10)
  - **`(T,T,H,T,T,T,T,H,H,T)`**

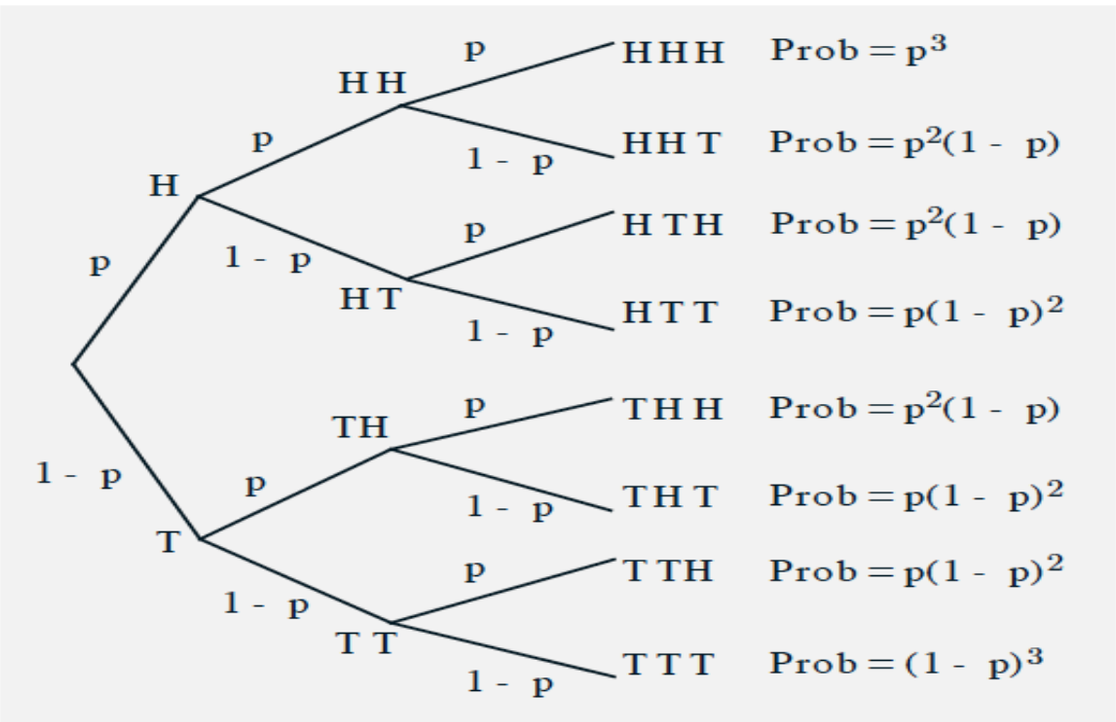
⇒ How many different events are there?

⇒ How probable is each of them?

⇒ math model: ***Binomial*** distribution:
`binom.pmf(n,p,k)`

Sol: Bernoulli chain experiment

Illustration by tree



e.g. here : n=3 levels

- sequential experiment, resu as string:
- $(a_1, a_2, \ldots, a_n)$ => ordered as tree
- a_i = << i-th letter == **H** >>
- **=> independent events (WHY?)**

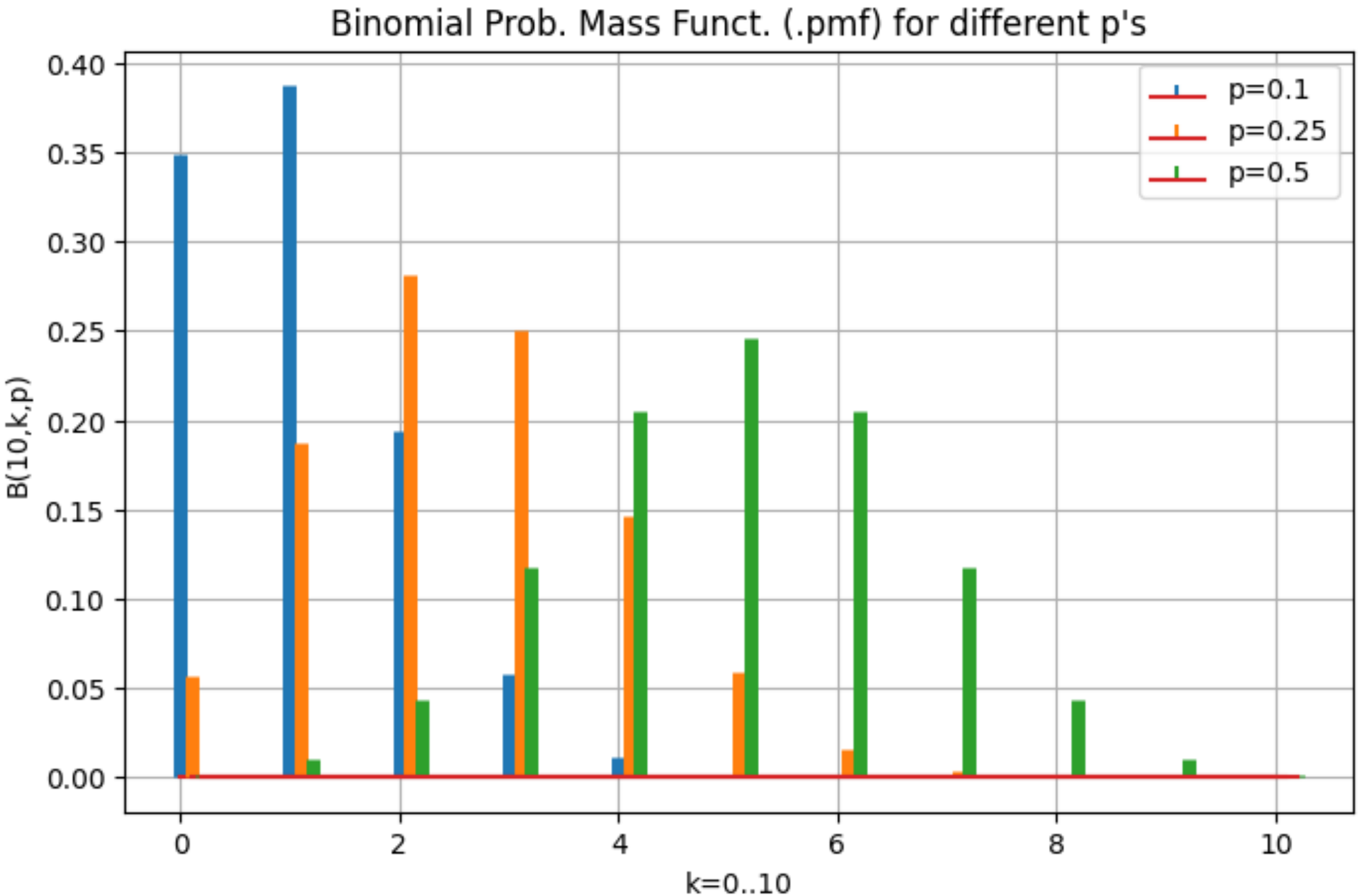
- p(k):=P({pick k times **H** in n trials})

$$p(k) = \binom{n}{k} q^k (1-q)^{n-k}$$

Definition:

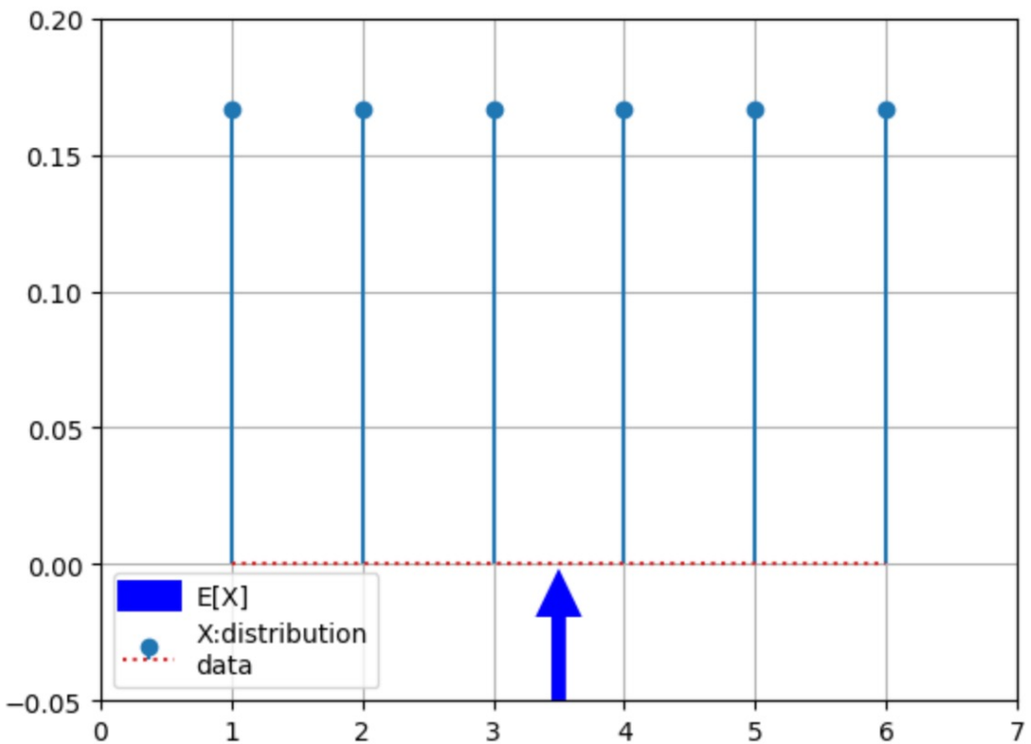
when the probabilities of the RV follows aboves ***p(k)*** for the parameters k , q and n then the distribution is called ***Binomial distribution***

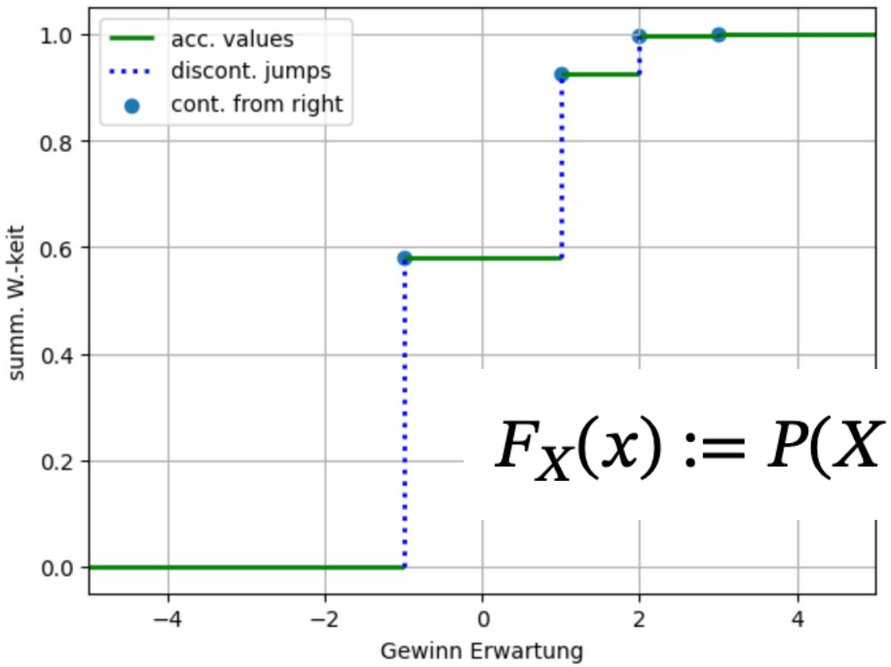
=> first example of a non-trivial RV


Binomial Prob. Mass Funct. (.pmf) for different p's
p=0.1
p=0.25
p=0.5
B(10,k,p)
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
2
4
6
8
10
k=0..10

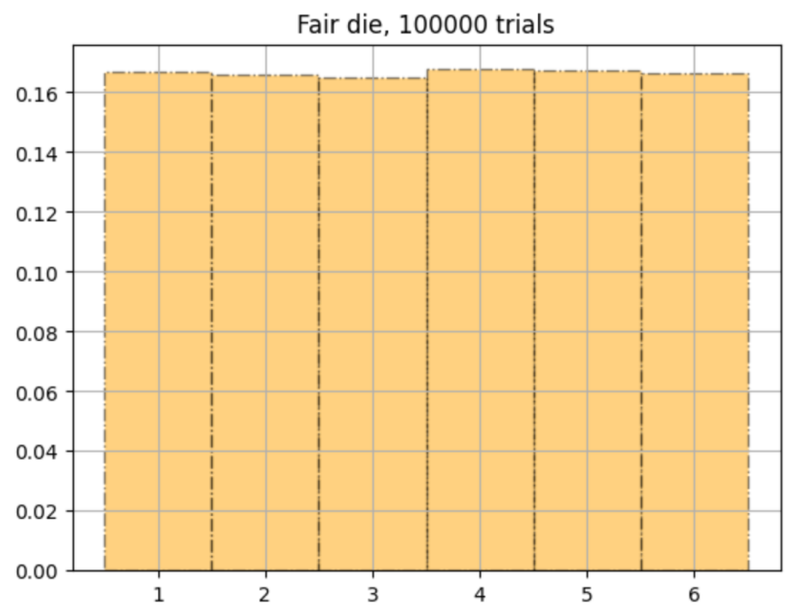
How to plot / represent a RV

ω	-1	1	2	3
$P(\omega)$	$\frac{125}{216}$	$\frac{75}{216}$	$\frac{15}{216}$	$\frac{1}{216}$



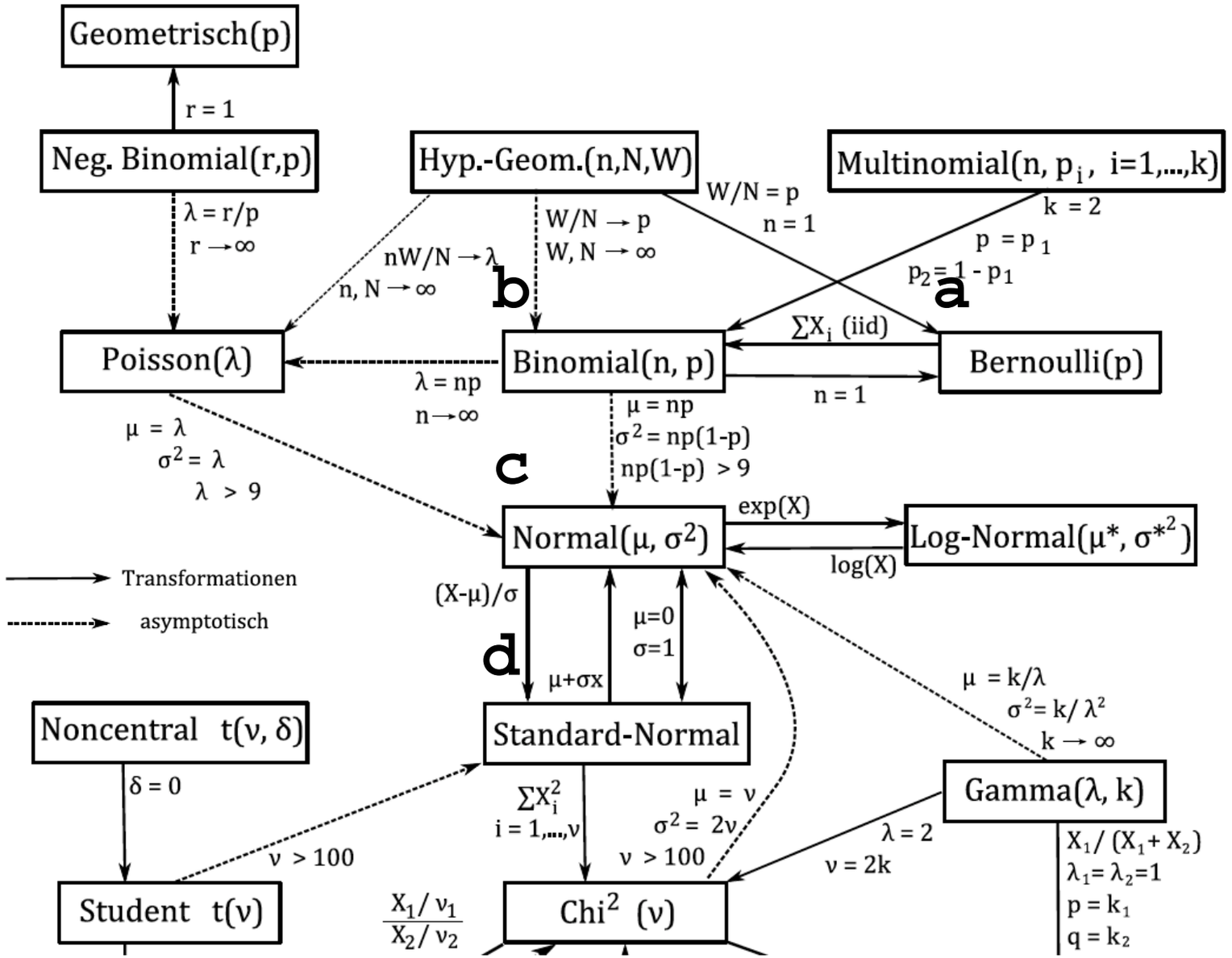


$$F_X(x) := P(X \leq x)$$




```
myy = [B_(1,6),B_(1,6),B_(1,6),B_(1,6),B_(1,6),B_(1,6)]
```

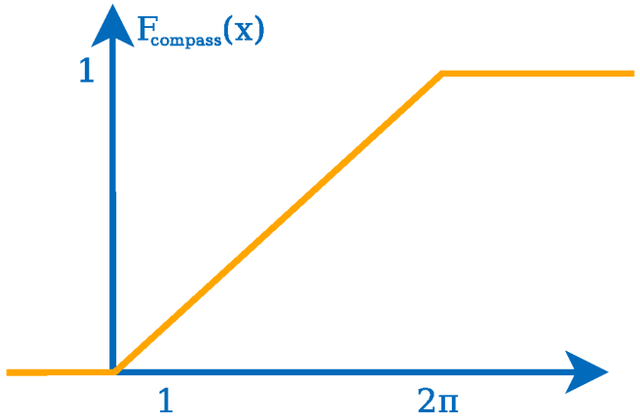
```
die = {"one":1/6,"two":1/6,"three":1/6,"four":1/6,"five
```

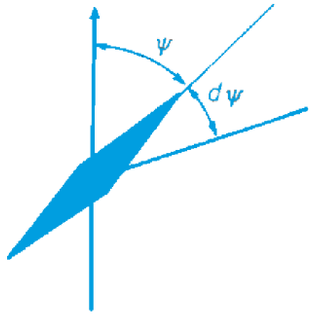


Interrelationship

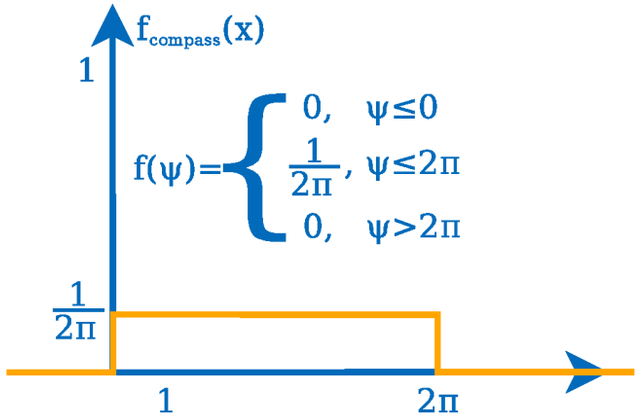
Jürgen Hedderich, Lothar Sachs, Angewandte Statistik, 15. Aufl, Springer 2016

Continuous Case: Compass Example





Taking the derivative of $F_X(x)$, we get



Probability density function (PDF)

$f(z)$ is called a **probability density function** (PDF)

$$P(\psi_1 < \psi \leq \psi_2) = \int_{\psi_1}^{\psi_2} \frac{1}{2\pi} d\psi = \frac{1}{2\pi}(\psi_2 - \psi_1)$$

Continuous Case

$$P(a < X \leq b) = \int_a^b f(z)dz$$

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(z)dz$$

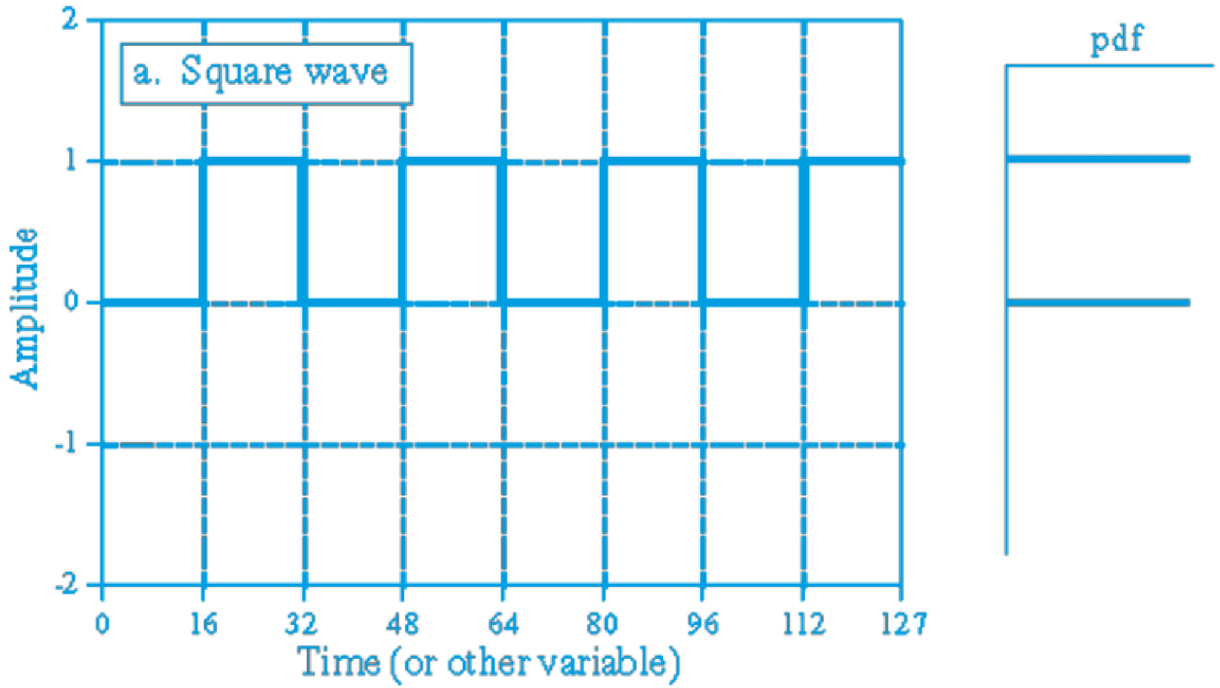
For all probability **density functions** $f(z)$, it holds that

1) $\forall z : f(z) \geq 0$

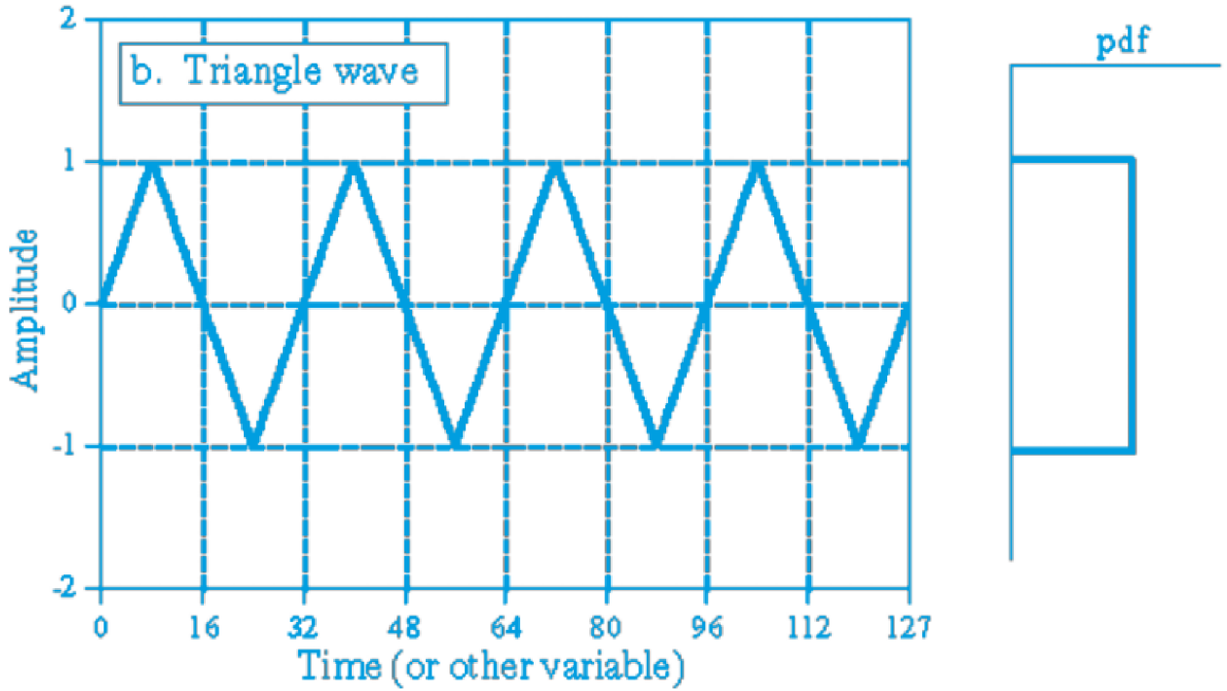
$$f(x) = \lim_{dx \to 0} \frac{F(x + dx) - F(x)}{dx} = \lim_{dx \to 0} \frac{P\left(x < X \leq x + dx\right)}{dx}$$

2) $\int_{-\infty}^{\infty} f(z)dz = 1$

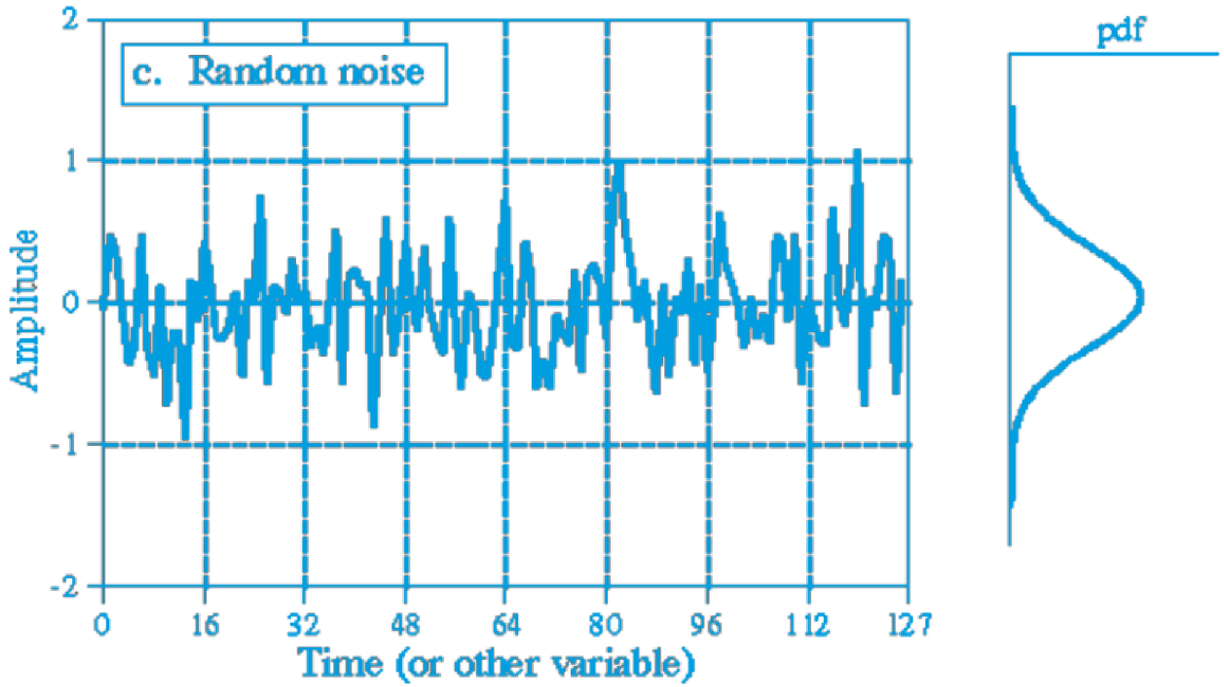
Signals and PDFs (1/3)



Signals and PDFs (2/3)



Signals and PDFs (3/3)



Normal (Gaussian) Probability Density Function

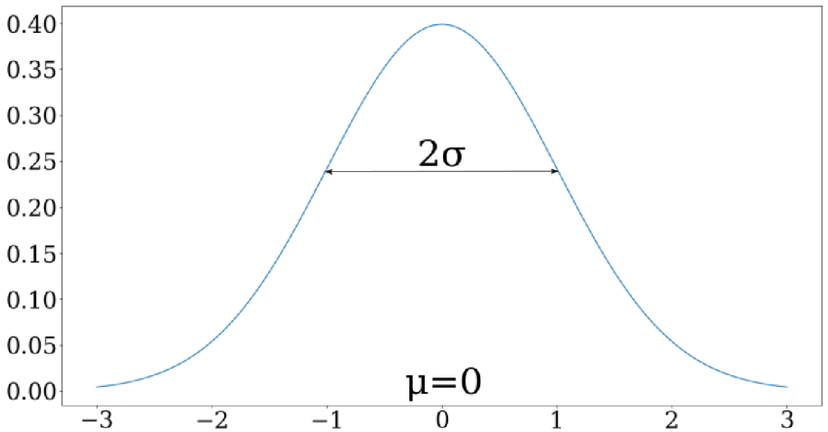
Normal distribution

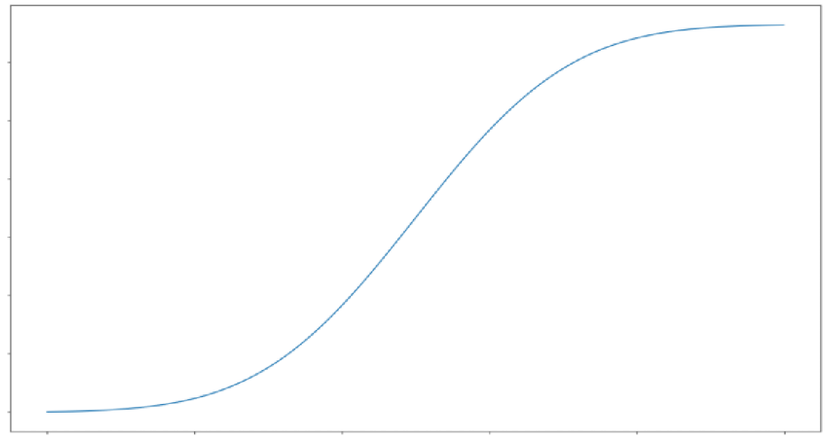
$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(t-\mu)^2}{\sigma^2}}$$

We often abbreviate this to $t \sim \mathcal{N}(\mu, \sigma^2)$

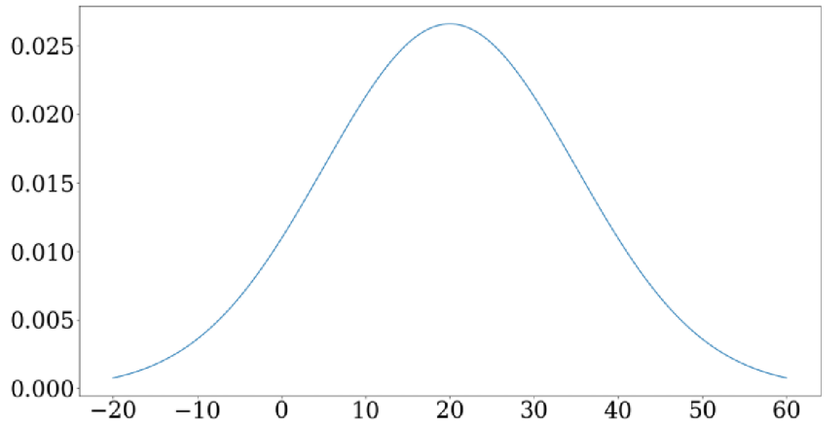
CDF

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

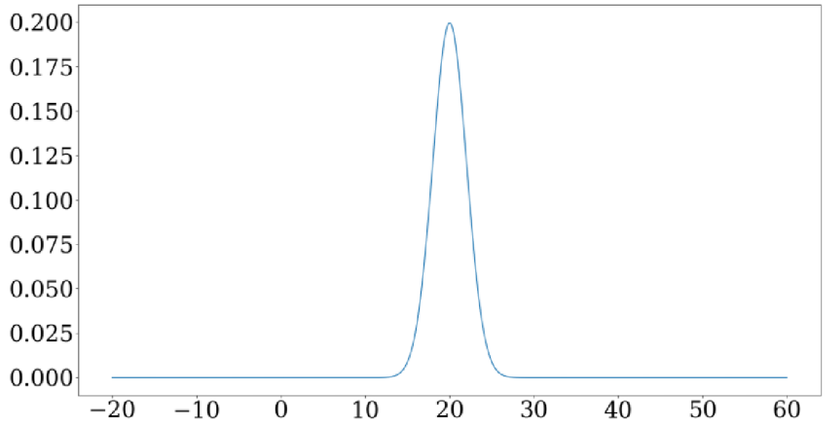




Gaussian Distributions Example



PDF of a Gaussian random variable X with mean $\mu = 20$ and variance $\sigma_x^2 = 225$



PDF of a Gaussian random variable X with mean $\mu = 20$ and variance $\sigma_x^2 = 4$

PDFs and CDFs in the Discrete Case

Recall that a PDF is defined as follows:

$$f(x) = \lim_{dx \to 0} \frac{F(x + dx) - F(x)}{dx} = \lim_{dx \to 0} \frac{P(x < X \leq x + dx)}{dx}$$

In the discrete case (e.g. rolling a die), the limit might become undefined (i.e. we may have steep steps at discrete points x_i); we therefore use

$$f(x) = \sum_i p_i \delta(x - x_i)$$

where

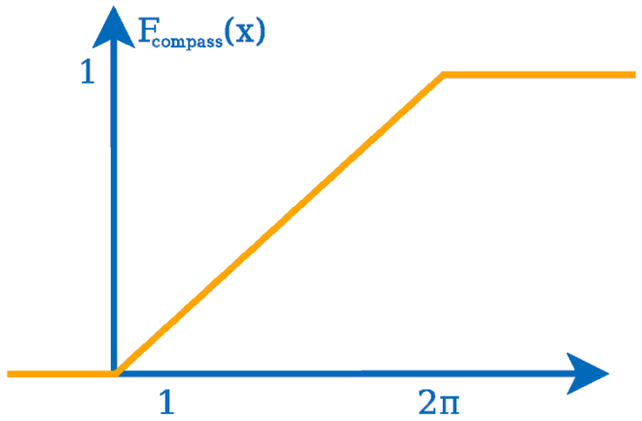
$$\delta(x) = 0 \;\; \forall x \neq 0, \quad \int \delta(x) dx = 1$$

We also have that

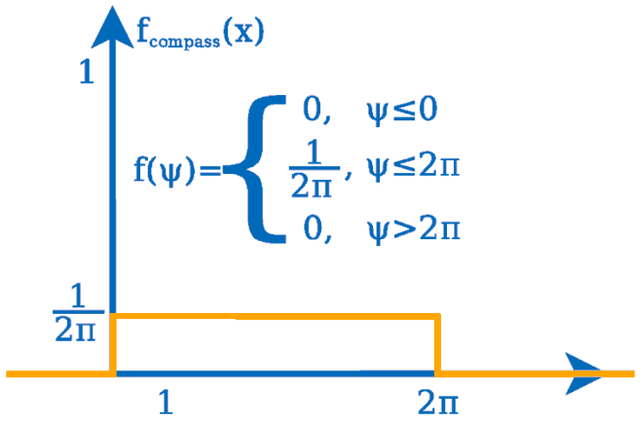
$$\int f(x) \delta(x - x_i) dx = f(x_i)$$

Warning: PDFs vs CDFs

Make sure not to confuse probability density functions (PDFs) and cumulative distribution functions (CDFs)



CDF $= F(x)$



PDF $= f(x)$

Expected Value and Variance

Expected value of a random variable X

Defined as the sum of all values that X can take weighted by the probability with which each value is taken

$$\mu = E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

Variance of X

$$\sigma^2 = \int_{-\infty}^{\infty} (x - E[X])^2 f(x) dx$$
$$= E[X^2] - E[X]^2$$

Expected value of a function of X

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Mean squared value

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx$$

μ Examples

Die

$$\mu = \int_{-\infty}^{\infty} x f(x) dx = \int x \sum_{i=1,\ldots,6} p_i \delta(x - x_i) dx$$
$$= \sum_{i=1,\ldots,6} p_i \int x \delta(x - x_i) dx$$
$$= \sum_{i=1,\ldots,6} p_i x_i = \frac{1}{6}(1 + \ldots + 6) = 3.5$$

Coin

$$\mu = \sum_{i=1,2} p_i x_i = \frac{1}{2}(1 + 2) = 1.5$$

Compass

$$\mu = \int_0^{2\pi} z \frac{1}{2\pi} dz = \frac{1}{2\pi} \frac{z^2}{2} \Big|_0^{2\pi} = \pi$$

σ^2 Examples

Die

$$\sigma^2 = \int (x-\mu)^2 f(x)dx = \sum_{i=1,\ldots,6} p_i \int (x-\mu)^2 \delta(x-x_i)dx$$

$$\frac{1}{6}\left((1-3.5)^2 + \ldots + (6-3.5)^2\right) = 2.917$$

Coin and compass

Find the variances in these examples as an exercise

Empirical Measurements

The exact analytical distribution is **usually** unknown; we therefore use samples to approximate its parameters

Empirical mean

The empirical mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

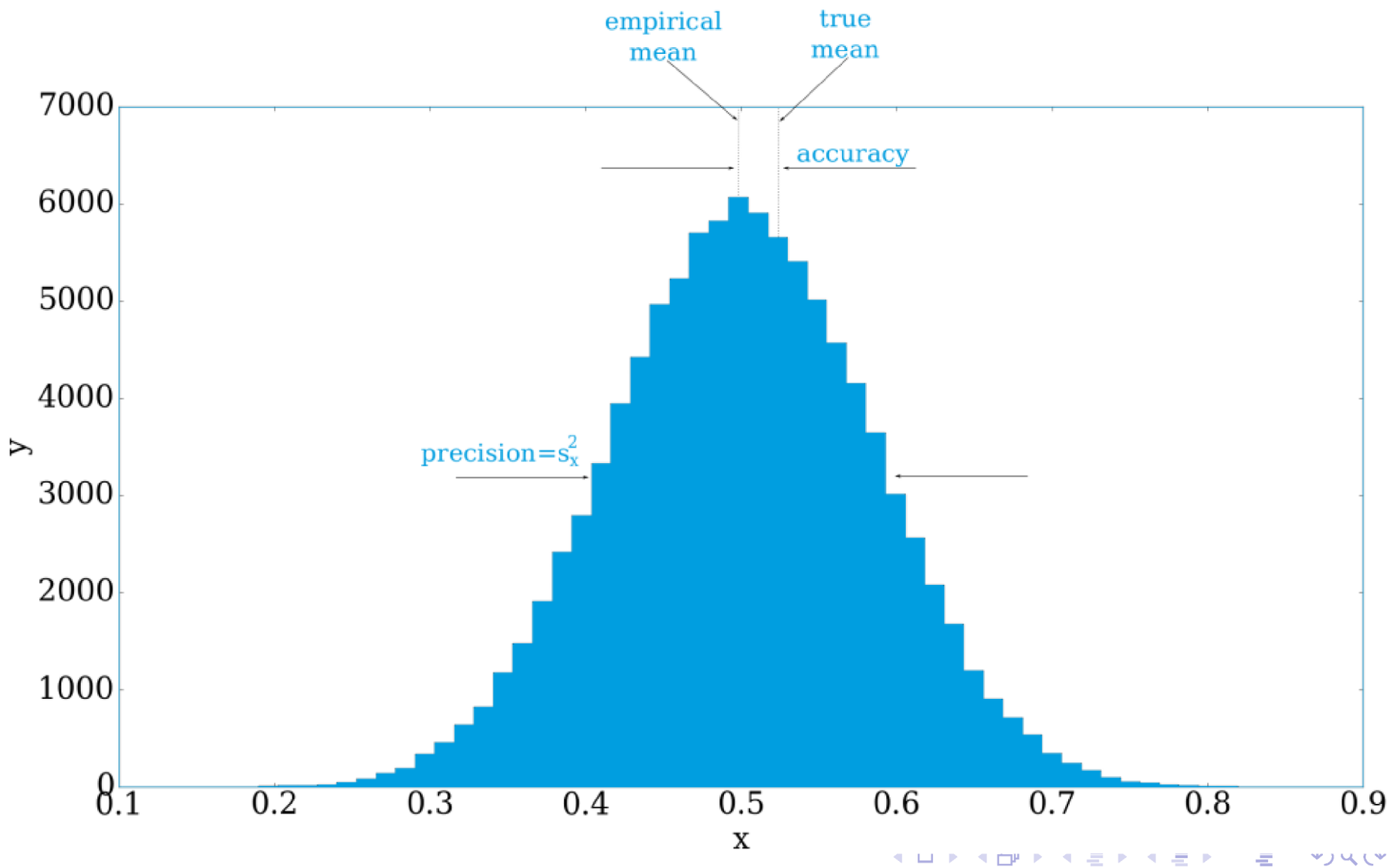
is used as a "best bet" approximation to the true expectation (i.e. the one of the actual distribution)

Empirical variance

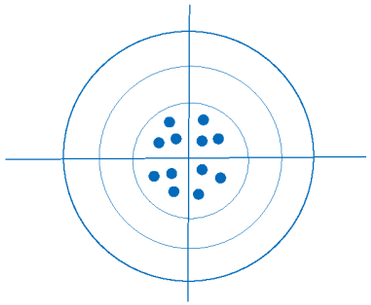
$$s_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Why are these numbers good/the best approximators?

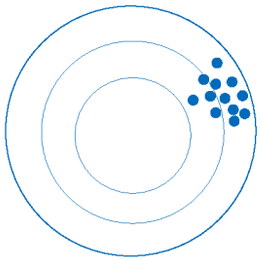
Empirical Measurements



Errors: Systematic vs. Random



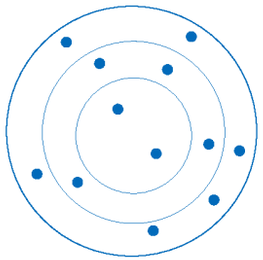
(a) Random: small
Systematic: small



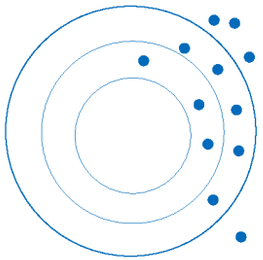
(b) Random: small
Systematic: large

(a) Random: small
Systematic: ?

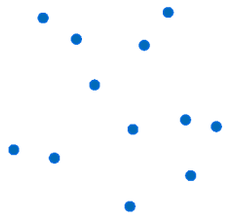
(b) Random: small
Systematic: ?



(c) Random: large
Systematic: small



(d) Random: large
Systematic: large



(c) Random: large
Systematic: ?

(d) Random: large
Systematic: ?

Target known: we can evaluate systematic errors

Target unknown: only random errors can be evaluated

Empirical Mean

$$\begin{aligned}\sum (x_i - \bar{x}) &= \sum x_i - \sum \bar{x} \\ &= \sum x_i - n\bar{x} \\ &= \sum x_i - n\frac{1}{n}\sum x_i = 0\end{aligned}$$

The empirical mean equilibrates the differences to all elements

The empirical mean is also minimal for the sum of squared differences, i.e.

$$\min_c \sum (x_i - c)^2 = \bar{x}$$

Proof That the Empirical Mean Minimises the Sum of Squared Differences

$$
\begin{aligned}
f(c) &= \sum_{i=1}^{n} (x_i - c)^2 = \sum_{i=1}^{n} \left(x_i^2 - 2cx_i + c^2\right) \\
&= nc^2 - 2c\left(\sum_{i=1}^{n} x_i\right) + \sum_{i=1}^{n} x_i^2 \\
&= n\left(c - \frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 + \left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right)
\end{aligned}
$$

The above expression is minimised for

$$
c = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}
$$

Mean Square Deviation

How far are all shoots astray? Measure the any-to-any distance by either

$$D = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|$$

or

$$S^2 = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - x_j)^2 = 2s_x^2$$

Proof That the Mean Square Deviation Equals Twice the Variance

$$
\begin{aligned}
S^2 &= \frac{1}{n^2} \sum_i \sum_j (x_i - x_j)^2 \\
&= \frac{1}{n^2} \sum_i \sum_j (x_i - \bar{x} + \bar{x} - x_j)^2 \\
&= \frac{1}{n^2} \sum_i \sum_j \left((x_i - \bar{x})^2 - 2 (x_i - \bar{x}) (x_j - \bar{x}) + (x_j - \bar{x})^2 \right) \\
&= \frac{1}{n^2} n \sum_i (x_i - \bar{x})^2 + \frac{1}{n^2} n \sum_j (x_j - \bar{x})^2 - 2 \frac{1}{n^2} \sum_i \sum_j (x_i - \bar{x}) (x_j - \bar{x}) \\
&= \frac{2}{n} \sum_i (x_i - \bar{x})^2 - 2 \frac{1}{n^2} \left(\sum_i (x_i - \bar{x}) \right) \left(\sum_j (x_j - \bar{x}) \right) \\
&= \frac{2}{n} \sum_i (x_i - \bar{x})^2 = 2 s_x^2
\end{aligned}
$$

Partial Wrap Up

Random variables are real valued results of random experiments (i.e. everything that can be "measured")

The probability measure induces a cumulative distribution function F

Most of the times, F induces a probability density function f

Using f, the calculation of probabilities can be done via integration

The mean value μ and the variance σ^2 can be easily calculated (via integration) if all of p or F is known

We can build empirical approximations to μ and σ^2 using the empirical mean $\bar{x}$ and the empirical variance s_x^2

Post-Lecture: What to do

Work over Lab Class 09

`09_probability.ipynb`

- Problems named:
  - Discrete random variables example 1
  - Discrete random variables example 2 (next time, not yet)
  - Common discrete random variables (maybe important for assignment)
  - Continuous random variables (maybe important for assignment)
  - Sampling (beyond this class , but doable)

Post-Lecture: What to do

You should be able to

- define what a random variable is
- explain the connection between a random variable, its probability density, and its cumulative distribution function
- explain what the central limit theorem states
- write down the density of the Gaussian / Normal distribution
- compute the expected value and variance for a given density
- compute the empirical mean and empirical variance for a sample
- know the difference between mean and median and what measure either of the minimizes