Probability theory lecture 1 ~~and 2~~

Sample space

Def: a collection of all elementary results, or outcomes of an experiment

Events

Def: a set of outcomes
- an event is a subspace of the sample space

Probability

- the probability that the ~~ve~~ event A occurs is:

$$P(A) = \frac{\text{nr. of favorable outcomes for the occurence of } A}{\text{nr. of all possible outcomes of the experiment}}$$

- if the experiment is repeated n times and A occurs k_n times, the relative frequency of A is:

$$h_n(A) = \frac{k_n}{n} \text{ ; } k_n \text{ is the absolute freq. of } A$$

$$h_n(A) \sim P(A) \text{ if } n \to \infty$$

- the ~~&~~ events $A_1, A_2 \ldots$ are exhaustive if their union equals the whole sample space:

$$A_1 \cup A_2 \ldots = \Omega$$

Axiomatic def:

- sample space (Ω): the set of all possible outcomes
- events (F): let F be a σ-algebra over Ω; any subset $A \in F$ is an event
- probability measure (P): assigns a real nr. between 0 and 1 to each event in F
- outcome (w): a specific event in Ω

Probability space

$P : F \to [0;1]$

$P(\Omega) = 1$, $P(\emptyset) = 0$

~~$P \cup A$~~ $P\left(\bigcup_{i \geq 1} A_i\right) = \sum_{i \geq 1} P(A_i)$

- (Ω, F, P) is called a probability space

- $A^C \to$ complement of A

$\bar{A} = A^C = \Omega \setminus A$

$P(A^c) = \cancel{B} 1 - P(A)$

$2) P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$3) P(A \setminus B) = P(A) - P(A \cap B)$

4) if $A \subseteq B \Rightarrow P(A) \leq P(B)$

~~5) $P(A \cup B) = P(A) + P(B)$~~ where $A, B \in F$

$(A_n)_{n \geq 1}$ is: an increasing seq. of events from F, if

$A_n \subseteq A_{n+1}, \forall n \in \mathbb{N}^*$

a decr. seq. of events from F, if $A_{n+1} \subseteq A_n$

$\forall n \in \mathbb{N}^*$

- if $(A_n)_{n \geq 1}$ is:

- an incr. seq. of events from F, then:

$$\lim_{n \to \infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right)$$

- a decr. seq. -//- :

$$\lim_{n \to \infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right)$$

$\overline{A \cup B} = \bar{A} \cap \bar{B}$

$\overline{A \cap B} = \bar{A} \cup \bar{B}$

~~Arr~~ Seminar 1

- Arrangements of n objects taken k: selections of k distinct ordered items from n objects

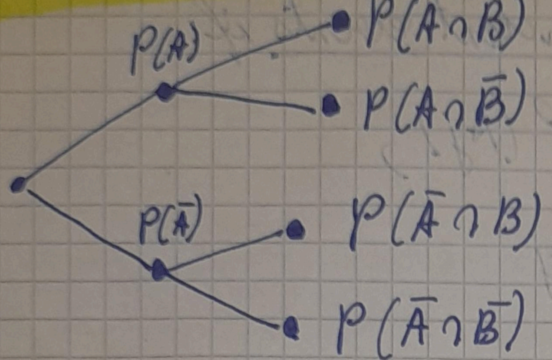
- Permutations with repetitions:
  - n objects:
    - ~~A~~ n_1 alike, n_2 alike .... n_m are alike
  - the nr. of distinct permutations is:

$$\frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_m!}$$

Probability lecture 2

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

probability of A given B (conditional prob.)



- the events A and B are <u>independent</u> if

$$P(A|B)=P(A) \Longleftrightarrow P(A\cap B)=P(A)\cdot P(B)$$

The following are equivalent:

1) A and B independent
2) $\bar{A}$ and B ind.
3) A and $\bar{B}$ ind.
4) $\bar{A}$ and $\bar{B}$ ind.

Def:

- $A_1, A_2 \ldots A_n \in F$ are ind. if:

$$P(A_{i_1} \cap \ldots \cap A_{i_m}) = P(A_{i_1}) \cdot \ldots \cdot P(A_{i_m})$$

for each subset $\{i_1 \ldots i_m\} \subset \{1, \ldots n\}$

- $A_1 \ldots A_n$ are pairwise independent events if:

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j),$$

$$\forall i, j \in \{1 \ldots n\}, \; i \neq j$$

Total probability

- consider a partition of Ω with mutually exclusive and exhaustive events $B_1, B_2 \ldots B_k$, that is

$$B_i \cap B_j = \emptyset, \forall i \neq j, \text{ and } B_1 \cup B_2 \cup .. \cup B_k = \Omega$$

- assume these events also partition the event A:

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_k)$$

and this is also a union of mutually exclusive events

- then $P(A) = \sum_{j=1}^{k} P(A \cap B_j)$ → Total Probability Law

Bayes' rule

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

~~Probability~~ RVs

Discrete RV

- prob. mass function (pmf):

$$X \sim \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}, \text{ where } p_i := P(X = x_i)$$

- cumulative distr. function ~~CD~~ (cdf):

$$F(x) = P(X \le x) = \sum_{\substack{i \in I \\ x_i \le x}} P(X \le x_i), \quad x \in \mathbb{R}$$

e.g.

$$X \sim \begin{pmatrix} -1 & 1 & 2 \\ 0.5 & 0.25 & 0.25 \end{pmatrix}$$

$$F(x) = \begin{cases} 0 & , \text{ if } X < -1 \\ 0.5 & , \text{ if } -1 \le X < 1 \\ 0.5 + 0.25 = 0.75 & , \text{ if } 1 \le X < 2 \\ 0.5 + 0.25 + 0.25 = 1 & , \text{ if } 2 \le X \end{cases}$$

$\lim_{X \to -\infty} F(X) = 0$ and $\lim_{X \to \infty} F(X) = 1$

$$P(a < X \le b) = F(b) - F(a)$$

- if X continuous $\Rightarrow P(X) = 0$, we must use $F(X)$ instead

- $P(X = b) = F(b) - F(b - 0)$, $b \in \mathbb{R}$
- F monotonically increasing
- $F(b + 0) = \lim_{X \searrow b} F(X) = F(b)$, $\forall b \in \mathbb{R}$

Continuous RV

- $F(x)$ cont.
- $f(x) = F'(x) \to$ probability density function (~~PDF~~, pdf, density)
- a distribution is cont. if it has a density

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} \underbrace{f(t)}_{\text{pdf}}\, dt \;, \; \forall x \in \mathbb{R}$$

$$P(x) = P(x \leq X \leq x) = \int_{x}^{x} f(t)\, dt = 0$$

if: $\int_{-\infty}^{\infty} f(x)\, dx = 1$
and $f(x) \geq 0 \; \forall x \in \mathbb{R}$ $\Bigg] \implies f$ is a density function!

~~Discrete RV:~~ Expectation

- if X discrete: $E[X] = \sum_{i \in I} x_i \cdot P(X = x_i) = \mu$
- if X cont: $E[X] = \int_{\mathbb{R}} x \cdot f(x)\, dx \cancel{= \frac{1}{b-a} \int_a^b x\, dx} = \frac{a+b}{2}$

- $E[a \cdot X + b] = a \cdot E[X] + b, \; \forall a, b \in \mathbb{R}$
- $E[X + Y] = E[X] + E[Y]$
- if $X(w) \geq 0, \forall w \in \Omega \implies E[X] \geq 0$
- if $X(w) \geq Y(w), \forall w \in \Omega \implies E[X] \geq E[Y]$
- discrete: $E[h(X)] = \sum_{i \in I} h(x_i) \cdot P(X = x_i)$
- cont.: $E[h(x)] = \int_{\mathbb{R}} h(x) \cdot f(x)\, dx$

...riance

$\sigma^2 = V(X) = E[(X-E[X])^2] \rightarrow$ variance (or dispersion) of X

$\sigma = \sqrt{V(X)} \rightarrow$ standard deviation of X

- discrete RV: $\boxed{\sum_{i=I} (x_i-\mu)^2 \cdot P(X=x_i)}$

- $V(X) = E[X^2] - (E[X])^2$
- $V(a\cdot X+b) = a^2 \cdot V(X)$, $\forall a, b \in \mathbb{R}$

Covariance

$cov(X,Y) = E[(X-E[X])\cdot(Y-E[Y])] = E[X\cdot Y] - E[X]\cdot E[Y]$

- $V(a\cdot X+b\cdot Y) = a^2\cdot V(X) + b^2 \cdot V(Y) + 2\cdot a\cdot b\cdot cov(X,Y)$, $\forall a, b \in \mathbb{R}$

Corellation coefficient

$$\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{V(X)\cdot V(Y)}}$$

- $\rho^2(x,y) \le 1$

Moments of a RV

X an RV, $k \in \mathbb{N}^*$

- $E[X^k]$ is the moment of order k of X
- $E[|X|^k]$ is the absolute moment of order k of X
- $E[(X-E[X])^k]$ is the central moment of order k of X

Moment generating function

- X an RV
- the moment generating function (MGF) of X is

$$M_X : D \subseteq \mathbb{R} \to \mathbb{R}$$

$$M_X(t) = E[e^{t\cdot x}]$$

Th. 27

- X, Y RVs
- if $Y = a\cdot X + b$ for fixed $a, b \in \mathbb{R}$, then $M_Y(t) = e^{t\cdot b} \cdot M_X(a\cdot t)$ $\forall b \in \mathbb{R}$, $\forall a, t \in \mathbb{R}$ for which $M_X(a\cdot t)$ is defined

Th. 29

- X, Y RVs for which $\exists\, \delta > 0$ s.t. $M_X(t) = M_Y(t)$, $\forall t \in (-\delta, \delta)$; then $F_X = F_Y$, i.e. X and Y have the same distribution

- ~~there~~ $\exists$ there is a one-to-one corespondence between the PDF of a RV and the MGF

Th. 30

- X a RV with $E[X], V[X]$; assume $a > 0$; then:

1) Markov's inequality

$$P(|X| \geq a) \leq \frac{1}{a} \cdot E[|X|]$$

2) Chebyshev's inequality

$$P(|X - E[X]| > a) \leq \frac{1}{a^2} \cdot V(X)$$

Distributions (discrete)

Bernoulli → Bernoulli(p), p is prob of success

- repeated independent trials for an experiment s.t. ∃ only success or failure

Binomial

- the nr. of successes in a series of n independent Bernoulli trials

$X \sim Bino(n, p)$, X denotes the nr. of successes

Hypergeometric

$X \sim Hyge(n, n_1, n_2)$

← prob. of obtaining a certain nr. of successes in a fixed-size sample, where each draw is made <u>without</u> replacement

Geometric

$X \sim Geo(p)$

← X denotes the nr. of failures before the first success in an infinite sequence of Bernoulli trials

Poisson

$X \sim Poiss(\lambda)$ ← X denotes the total nr. of occurences of some phenomenon ~~of~~ in a fixed period of time or region of space

Distributions (continuous)

Continuous & Uniform

$X \sim Unif([a, b])$ $\quad f(x) = \frac{1}{b-a}$

- used when we pick a value arbitrarily from a given interval, without any particular preference for higher, lower or medium values

The uniform property

$\forall h > 0, \forall t \in [a, b-h]$ we have

$$P(t < X < t+h) = \int_t^{t+h} \frac{1}{b-a}\, dx = \frac{h}{b-a}$$

$\cancel{f(x) = \frac{1}{b-a}}$

Exponential

$X \sim Exp(\lambda)$ $\quad f(x) = \lambda \cdot e^{-\lambda \cdot x}, \quad x > 0$

- models time: waiting time, lifetime of components, failure times, service time in a queue etc.
- if the (discrete) nr. of events has Poisson dist., the (cont.) time between events has ~~ex~~ exponential dist.

$\cancel{f(x)}$ - λ represents the frequency and is measured in min^{-1}

Memoryless

$P(T > t+x \mid T > t) = P(T > x); \quad t, x > 0$

- geometric distr. is also ~~memoryless~~ memoryless

Gamma distr.

- a specific process takes place in α independent steps, and each step takes Exp (λ), then the total time has Gamma distribution with params. α and λ

$X \sim \Gamma(\alpha, \lambda)$

$$f(x) = \frac{\lambda^a}{\Gamma(a)} \cdot x^{a-1} \cdot e^{-\lambda \cdot x}$$

where $\Gamma(t) = \int_0^\infty x^{t-1} \cdot e^{-x}\,dx$, $t>0$

$\alpha \to$ ~~fr~~ shape parameter

$\lambda \to$ freq. parameter

Normal

$X \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{\frac{-(x-\mu)^2}{2 \cdot \sigma^2}} \quad \text{for } x \in (-\infty; \infty)$$

where $\mu = E[X]$ and $\sigma^2 = V(X)$

- models the distribution of the sum (or average) of a large nr. of independent and identically distributed RVs

e.g height, weight

Sequences of random variables

Def. 10

$(X_n)_{n\geq 1}$ (a seq. of RVs) **converges almost surely** to a RV X if

$$P(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$$

not. $X_n \xrightarrow{a.s.} X$

Def. 32

$(X_n)_{n\geq 1}$ of RVs **converges in probability** to X if

$$\lim_{n\to\infty} P(|X_n - X| \leq \varepsilon) = 1, \ \forall \varepsilon > 0$$

not. $X_n \xrightarrow{P} X$

Obs: $X_n \xrightarrow{P} X \iff \lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0, \forall \varepsilon > 0$

Def. 33

$(X_n)_{n\geq 1}$ **converges in mean square** to X if

$$\lim_{n\to\infty} E[|X_n - X|^2] = 0$$

not. $X_n \xrightarrow{L2} X$

Def. 34

$(X_n)_{n\geq 1}$ **converges in distribution** to X if

$$\lim_{n\to\infty} \cancel{f}\, F_{X_n}(x) = F_X(x) \text{ in each continuity point } x \text{ of } F_X$$

not. $X_n \xrightarrow{d} X$

Th. 35

$$\boxed{X_n \xrightarrow{a.s.} X} \Longrightarrow \boxed{X_n \xrightarrow{P} X} \Longrightarrow \boxed{X_n \xrightarrow{d} X}$$

$$\Uparrow$$

$$\boxed{X_n \xrightarrow{L_2} X}$$

Law of large numbers (LLN)

- if the same experiment is repeated independently a large nr. of times, the average of the results must be close to the expected value

- Weak law of large numbers (WLLN)

- $(X_n)_{n\geq 1}$ s. t. $E[|X_n|] < \infty$, $\forall n \in \mathbb{N}$ obeys the WLLN if

$$\frac{1}{n} \cdot \sum_{k=1}^{n} (X_k - E[X_k]) \xrightarrow{P} 0$$

Th. 36

- $(X_n)_{n\geq 1}$ a seq. of pairwise independent RVs satisfying $V(X_n) \leq L$, $\forall n \in \mathbb{N}^*$, where L constant

- then $(X_n)_{n\geq 1}$ obeys the WLLN

Strong law of large nrs. (SLLN)

$(X_n)_{n\geq 1}$ a seq. of RVs. s. t. $E[|X_n|] < \infty$, $\forall n \in \mathbb{N}$ obeys SLLN if

$$\frac{1}{n} \cdot \sum_{k=1}^{n} (X_k - E[X_k]) \xrightarrow{a.s.} 0$$

- if $\sum_{n=1}^{\infty} \frac{1}{n^2} \cdot V(X_n) < \infty \Rightarrow (X_n)$ obeys ~~SA~~ SLLN

- if (X_n) obeys SLLN $\Rightarrow (X_n)$ obeys WLLN

- if $E[X_n] = m$ $\forall n \in \mathbb{N} \Rightarrow (X_n)$ obeys SLLN

Th. 40

$h_n(A) \xrightarrow{a.s.} P(a)$ if $n \to \infty$

Sums of RVs

Random walk

$S_n = S_{n-1} + X_n \leftarrow$ random walk in 1D

- used to model behavior of systems where each step is unpredictable (e.g. movement of particles)

$S_n \to$ position after n steps

$X_n \to$ step taken at step n

- for a random walk in 1D with step size $+1$ and -1 with prob. p and $1-p$:

$$E[S_n] = n \cdot p$$
$$V(S_n) = n \cdot p \cdot (1-p)$$

Central limit theorem

- $(X_n)_{n \geq 1}$ a seq. of independent identically distributed RVs s.t. $V(X_n) > 0, \forall n \geq 1$; then

$$\frac{(X_1 + \ldots + X_n) - n \cdot E[X_n]}{\sqrt{n \cdot V(X_n)}} \xrightarrow{P} Z, \text{ where } Z \sim N(0,1)$$

- the distr. of sample means approximates a normal distr. as the sample size increases, regardless of the sample's distr.