Recap: Probability Theory

Mathematical Probability

- Ω (or $\mathcal{X}$) …sample space
- $\mathcal{A}$ …collection of events/subsets of Ω (σ-algebra)
- $P : \mathcal{A} \to [0, 1]$ …probability assignment

axioms of probability theory

ϕ = empty set

1. $P(\Omega) = 1$.
2. $P(\varnothing) = 0$.
3. If $A_i \in \mathcal{A}$, $i \in \mathbb{N}$, are mutually disjoint, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

We say the events $A, B \subseteq \Omega$ are independent if

$$P(A \cap B) = P(A)P(B).$$

Mathematical Probability: discrete (example)

Tossing a regular six sided die

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$\begin{aligned}
\mathcal{A} &= \text{ all subsets of } \Omega \\
&= \{\varnothing, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{1, 2\}, \{1, 3\}, \dots \{1, 2, 3, 4, 5\}, \Omega\}
\end{aligned}$$

$$P(\{i\}) := \frac{1}{6}, \quad i = 1, \dots, 6$$

compute probability of an odd number:

$$\begin{aligned}
P(\{1, 3, 5\}) &= P(\{1\} \cup \{3\} \cup \{5\} \cup \varnothing \cup \varnothing \dots) \\
&\overset{3.}{=} P(\{1\}) + P(\{3\}) + P(\{5\}) + \overset{=0}{P(\varnothing)} + \overset{=0}{P(\varnothing)} + \dots \\
&\overset{2.}{=} \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + 0 + 0 + \dots = \frac{3}{6} = \frac{1}{2}
\end{aligned}$$

Mathematical Probability: discrete (example)

Tossing two fair coins

$$\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$$
$$\mathcal{A} = \text{ all subsets of } \Omega$$
$$= \{\varnothing, \{(H,H)\}, \{(H,T)\}, \{(T,H)\}, \{(T,T)\}, \{(H,H),(H,T)\}, \ldots$$
$$\{(H,H),(H,T),(T,H)\}, \Omega\}$$
$$P(\{(a,b)\}) := \frac{1}{4}, \quad a,b \in \{H,T\}$$

The events $A = \{(H,H),(H,T)\}$ and $B = \{(H,T),(T,T)\}$ are independent:

$$P(A) = P(\{(H,H)\}) + P(\{(H,T)\}) = \frac{1}{2} = P(B)$$
$$P(A \cap B) = P(\{(H,T)\}) = \frac{1}{4} = P(A)P(B)$$

Mathematical Probability: discrete

- $\mathcal{X}$ = a finite or countably infinite set ($\{x_1, x_2, \dots\}$)
- $\mathcal{A}$ = the collection of all subsets of $\mathcal{X}$
- $f : \mathcal{X} \to [0,1]$ a **probability mass function (pmf)**, i.e.,

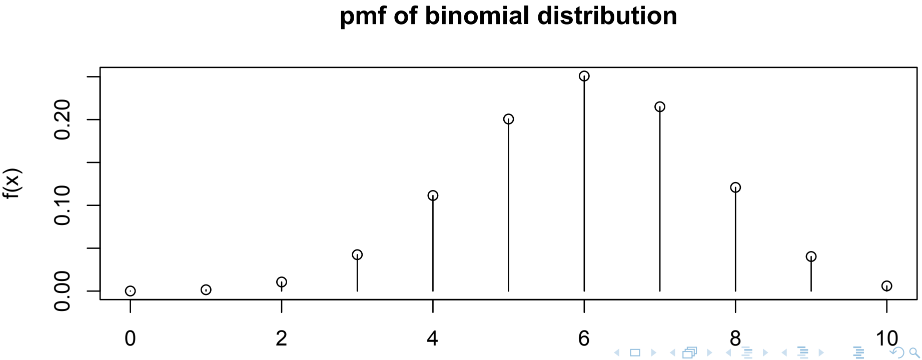
$$\boxed{f(x) := P(\{x\}),} \quad x \in \mathcal{X}.$$

$$P(A) := \sum_{x \in A} f(x), \qquad \text{for } A \in \mathcal{A}.$$

Mathematical Probability: discrete (example)

- ~~$\mathcal{X} = \mathbb{N}_0 = \{0, 1, 2, \dots\}$~~ $\mathcal{X} = \{0, 1, 2, \dots\, n\}$
- parameters $n \in \mathbb{N}$, $\theta \in [0, 1]$.

$$f(x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$



Mathematical Probability: continuous

$P(\{x\}) = ?$ $\quad x \in \mathbb{R}$

- $\mathcal{X} \subseteq \mathbb{R}^d$
- $\mathcal{A}$ = a collection of (measurable) subsets of $\cancel{\mathbb{R}}$ $\mathcal{X}$
- $f : \mathcal{X} \to [0, \infty)$ a **probability density function (pdf)**, i.e.,

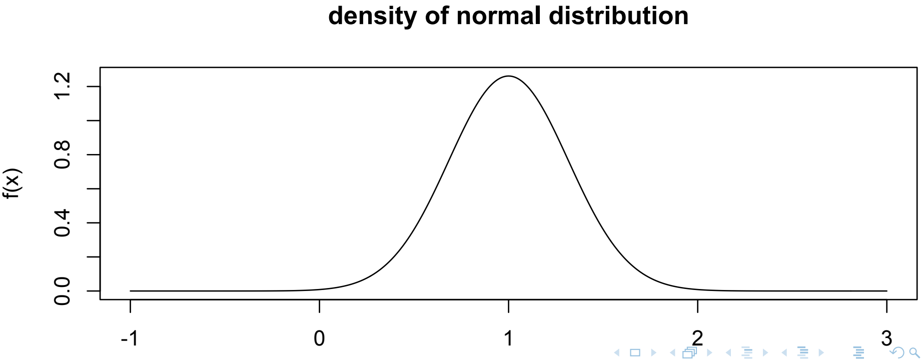
$$f(x) \geq 0, \quad \int_{\mathcal{X}} f(x)dx = 1.$$

$$P(A) := \int_A f(x)dx, \quad \text{for } A \in \mathcal{A}.$$

Mathematical Probability: continuous (example)

- $\mathcal{X} = \mathbb{R}$
- Parameters $\mu \in \mathbb{R}$ and $\sigma^2 \in (0, \infty)$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$



Random Variables

Informally: A random variable X represents **all** potential realizations of a random experiment.

Formally: A random variable X is a (measurable) function $X : \Omega \to \mathbb{R}$ and $P(X \in A) := P(\{\omega \in \Omega : X(\omega) \in A\})$.

We say that X has pmf or pdf f (notation: $X \sim f$), if

$$P(X \in A) = \begin{cases} \sum_{x \in A} f(x), & \text{in the discrete case} \\ \int_A f(x)dx, & \text{in the continuous case.} \end{cases}$$

$$\mathbb{E}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = ?$$

For $g : \mathbb{R} \to \mathbb{R}$, e.g. $g(x) = x$

$$:= \begin{pmatrix} \mathbb{E}(X_1) \\ \mathbb{E}(X_2) \end{pmatrix}$$

$$\mathbb{E}[g(X)] = \begin{cases} \sum_{x \in \mathcal{X}} g(x)f(x), & \text{discrete} \\ \int_{\mathcal{X}} g(x)f(x)dx, & \text{continuous.} \end{cases}$$

$$\mathrm{Var}[g(X)] := \mathbb{E}[(g(X) - \mathbb{E}[g(X)])^2] = \mathbb{E}[g(X)^2] - \mathbb{E}[g(X)]^2$$

We say that the random variables X and Y are **independent** if

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B),$$

for all events A and B.

$$P(\{\omega : X(\omega) \in A, Y(\omega) \in B\})$$
$$= P(\{\omega : X(\omega) \in A\} \cap \{\omega : Y(\omega) \in B\})$$

Random Variables

Can do abstract computations without fixing any particular number/realization/sample.

For X, Y real RVs and $a, b \in \mathbb{R}$:

$$\begin{aligned}
\mathbb{E}[aX + bY] &= a\mathbb{E}[X] + b\mathbb{E}[Y] \\
\operatorname{Var}[aX] &= a^2 \operatorname{Var}[X] \\
\mathbb{E}[XY] &= \mathbb{E}[X]\mathbb{E}[Y] \quad \text{if } X \text{ and } Y \text{ are independent} \\
\operatorname{Var}[X + Y] &= \operatorname{Var}[X] + \operatorname{Var}[Y] \quad \text{if } X \text{ and } Y \text{ are independent}
\end{aligned}$$

Random Variables (example)

- data: $X_1, \ldots, X_n$ i.i.d. (= **i**ndependent and **i**dentically **d**istributed)
- represents all potential samples of size n
- sample mean $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$

$$\mathbb{E}(X_1) = \mu = \mathbb{E}(X_2) = \ldots = \mathbb{E}(X_n)$$

$$\mathbb{E}[\bar{X}_n] = \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \mathbb{E}\left(\sum_{i=1}^n X_i\right)$$

$$= \frac{1}{n} \sum_{i=1}^n \underbrace{\mathbb{E}(X_i)}_{=\mu} = \frac{1}{n} n \cdot \mu = \mu$$

$$\mathrm{Var}[\bar{X}_n] = \mathrm{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \mathrm{Var}\left(\sum_{i=1}^n X_i\right)$$

$$= \frac{1}{n^2} \sum_{i=1}^n \underbrace{\mathrm{Var}(X_i)}_{=\sigma^2} = \frac{1}{n^2} n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

Cumulative Distribution Function (cdf)

Unifying discrete and continuous distributions:

$$\underline{F_X(x) := P(X \leq x)}, \quad x \in \mathbb{R}.$$

Properties:

- $F : \mathbb{R} \to [0, 1]$
- F is non-decreasing
- $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to +\infty} F(x) = 1$
- $F(x_0^+) := \lim_{x \downarrow x_0} F(x) = F(x_0)$ (right-continuity)

Cumulative Distribution Function (CDF)

$F(x_n)$

$\downarrow$

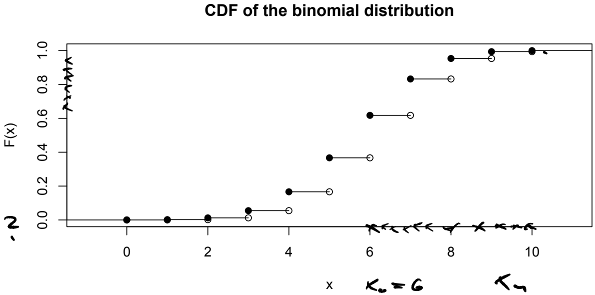
$F(x_0)$

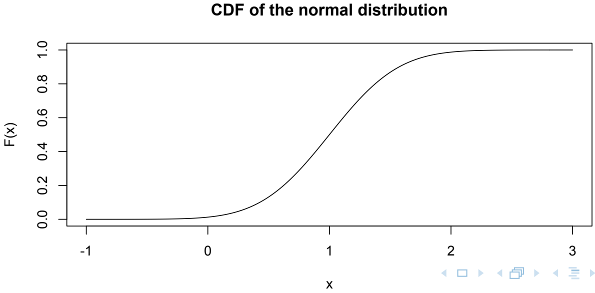
$F(6) \approx$

$\approx 0{,}6$

$F(6^-) = ?$

≈ 0.4





Cumulative Distribution Function (CDF)

The CDF encodes all information of a univariate distribution.

continuous:

pdf   cdf

$$F(x) = \int_{-\infty}^{x} f(u)\,du, \quad f(x) = \frac{d}{dx}F(x).$$

discrete:

$$F(x) = \sum_{y \le x} f(y), \quad f(x) = \overbrace{F(x) - F(x^-)}^{\text{jump size}}.$$

pmf   cdf

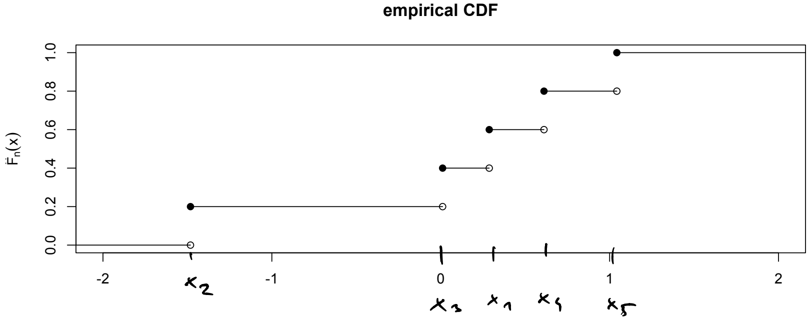
$$F(x-) := \lim_{y \uparrow x} F(y)$$

Empirical CDF

Given data $x_1, \dots, x_n \in \mathbb{R}$, the corresponding empirical cdf is given by

$$\hat{F}_n(x) := \frac{\#\{i : x_i \le x\}}{n} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(x_i)$$

$$\mathbb{1}_A(x) := \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

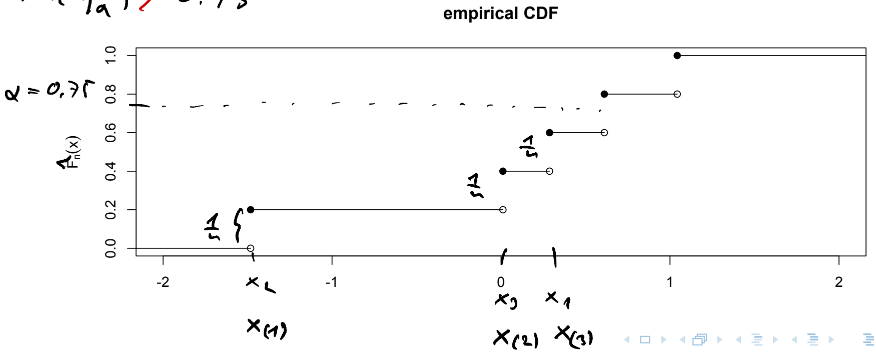


Empirical CDF

Given data $x_1, \ldots, x_n \in \mathbb{R}$, let $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ be the ordered values. Then (assuming no ties)

$$\hat{F}_n(x) = \begin{cases} 0, & \text{if } x < x_{(1)}, \\ \frac{i}{n}, & \text{if } x_{(i)} \leq x < x_{(i+1)}, \qquad i = 1, \ldots, n-1, \\ 1, & \text{if } x_{(n)} \leq x. \end{cases}$$

$\hat{F}_n(q_\alpha) \overset{!}{\neq} 0.75$



Quantiles

Informally: For a given probability $\alpha \in [0, 1]$, the α-quantile of a distribution F is the number q_α such that exactly α of the probability mass lies at or below q_α, i.e.,

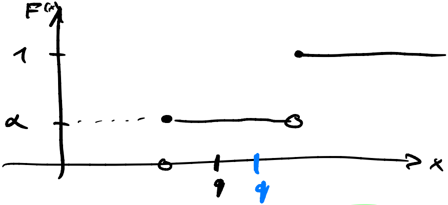
$$F(q_\alpha) = P(X \leq q_\alpha) = \alpha.$$

Note: q_α may not exist!

QUANTILES

• $F(q) = \alpha \geq \alpha$

• $F(q^-) = F(q) = \alpha \leq \alpha$



Formally: For a given probability $\alpha \in [0, 1]$ and a CDF $F : \mathbb{R} \to [0, 1]$, an α-quantile of F is an extended real number $q_\alpha \in \bar{\mathbb{R}} := [-\infty, \infty]$, such that

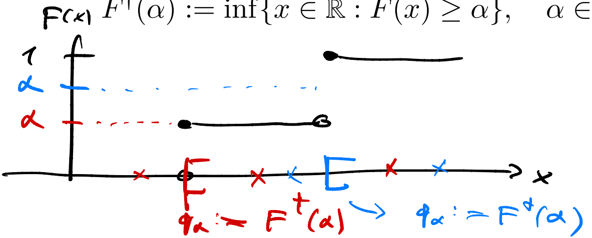
$$\boxed{F(q_\alpha) \geq \alpha \quad \text{and} \quad F(q_\alpha^-) \leq \alpha.}$$

Note: q_α may not be unique! How do we pick one?

QUANTILES

Quantile function aka. generalized inverse

$$F^{\dagger}(\alpha) := \inf\{x \in \mathbb{R} : F(x) \geq \alpha\}, \quad \alpha \in [0, 1]$$



- Is the smallest of all α-quantiles of F.
- If F is invertible, then $F^{\dagger}(\alpha) = F^{-1}(\alpha)$.
- For $i = 1, \dots, n$, we have $\hat{F}_n^{\dagger}(\alpha) = x_{(i)}$ if, and only if, $\frac{i-1}{n} < \alpha \leq \frac{i}{n}$. (empirical quantiles = order statistics)

Convention: $\inf \varnothing := +\infty$. HW

THE LAW OF LARGE NUMBERS (WEAK)

$X_i \sim \text{Bernoulli}(\theta)$

$\mathbb{E} X_i = \theta$

Let $X_1, \ldots, X_n$ be i.i.d. random variables with $\mathbb{E}[X_1] = \mu \in \mathbb{R}$ and $\operatorname{Var}[X_1] = \sigma^2 < \infty$. Then

"in probability"

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{i.p.} \mu.$$

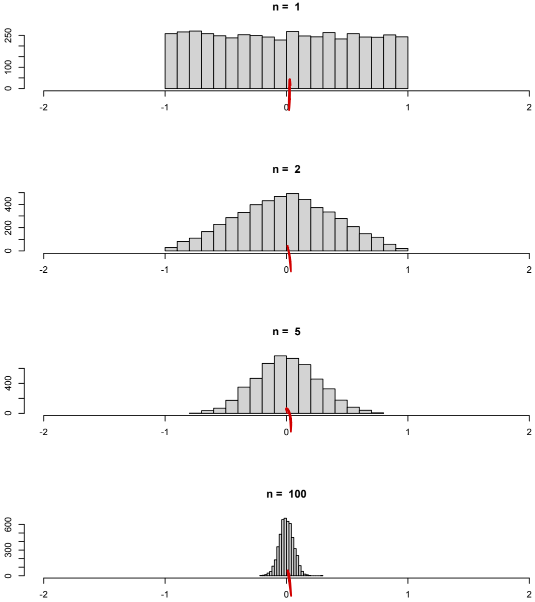
More precisely, for every $\varepsilon > 0$,

$$P\left(|\bar{X}_n - \mu| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0.$$

Proof: on black board

THE LAW OF LARGE NUMBERS

histograms of 1000 $\overline{X}_n$



$\mu = 0$

The Central Limit Theorem

Let $X_1, \ldots, X_n$ be i.i.d. random variables with $\mathbb{E}[X_1] = \mu \in \mathbb{R}$ and $\mathrm{Var}[X_1] = \sigma^2 < \infty$. Then

in distribution

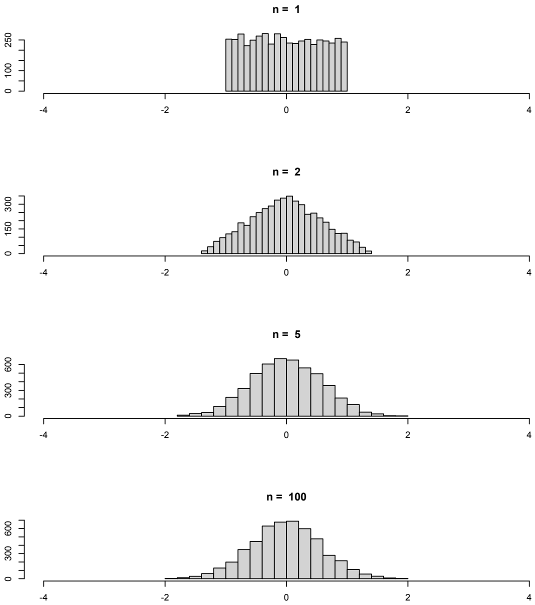
$$\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \xrightarrow[n\to\infty]{d.} N(0,1).$$

More precisely, for every $x \in \mathbb{R}$,

$$P\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \leq x\right) \xrightarrow[n\to\infty]{} \Phi(x) := P(N(0,1) \leq x).$$

Note: $\mathrm{Var}(\sqrt{n}\frac{\bar{X}_n-\mu}{\sigma}) = \frac{n}{\sigma^2}\mathrm{Var}(\bar{X}_n) = 1$

The Central Limit Theorem



Multivariate distributions

The notions of pmf, pdf, and cdf naturally extend to more than one dimension:

Let X and Y be random variables and $A \subseteq \mathcal{X}^2$:

joint pmf ($\mathcal{X}$ discrete): $f_{X,Y} : \mathcal{X}^2 \to [0, 1]$

$f_{X,Y}(x,y) =$

$$P((X, Y) \in A) = \sum_{(x,y) \in A} f_{X,Y}(x, y)$$

$= P(X = x, Y = y)$

joint pdf ($\mathcal{X} = \mathbb{R}$): $f_{X,Y} : \mathbb{R}^2 \to [0, \infty)$

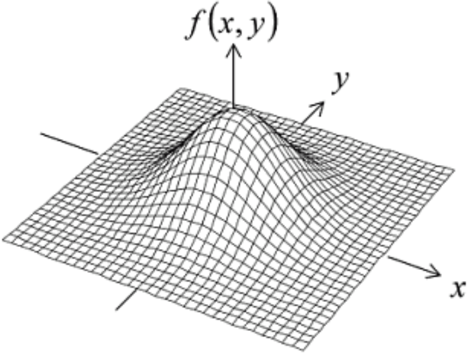
$$P((X, Y) \in A) = \int_A f_{X,Y}(x, y) dx dy$$

joint cdf: $F_{X,Y} : \mathbb{R}^2 \to [0, 1]$, $F_{X,Y}(x, y) = P(X \le x, Y \le y)$

Multivariate normal distribution

- Parameters $\mu \in \mathbb{R}^p$ and $\Sigma \in \mathbb{R}^{p \times p}$ positive definite.

$$f_{\mu,\Sigma}(x) = (2\pi)^{-p/2} \det(\Sigma)^{-1/2} \exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right), \quad x \in \mathbb{R}^p.$$



$p = 2$

$\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$\Sigma = I_2$

Multivariate distributions

Let X and Y be random variables with joint pdf/pmf $f_{X,Y}$.

- The **marginal pmf/pdf** of X is given by

$$f_X(x) = \begin{cases} \int_{\mathbb{R}} f_{X,Y}(x,y)dy, \\ \sum_{y \in \mathcal{X}} f_{X,Y}(x,y). \end{cases}$$

- X and Y are **independent** if, and only if,

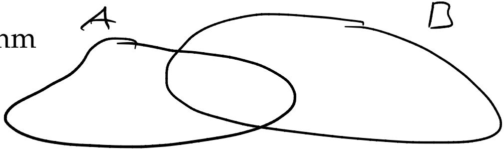
$$f_{X,Y}(x,y) = f_X(x) f_Y(y), \qquad \forall x, y.$$

Conditional probability

For events $A, B \subseteq \Omega$ with $P(B) > 0$, the conditional probability of A given B is defined as

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

Intuition: Venn-Diagramm



Note: $A \mapsto P(A|B)$ is a probability assignment, i.e.,

- $P(\Omega | B) = 1$
- $P(\emptyset | B) = 0$
- $P\left(\bigcup_{i=1}^{\infty} A_i \,\middle|\, B\right) = \sum_{i=1}^{\infty} P(A_i | B)$ if pairwise disjoint

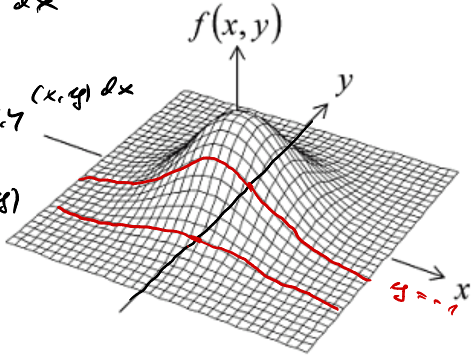
Conditional density

Let X and Y be random variables with joint pdf/pmf $f_{X,Y}$.
Then the function

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

is called the **conditional pdf/pmf** of X given $Y = y$.

$$\int_{\mathbb{R}} f_{X|Y=y}(x)\, dx = \frac{1}{f_Y(y)} \int_{\mathbb{R}} f_{X,Y}(x,y)\, dx = \frac{1}{f_Y(y)} f_Y(y) = 1$$



BAYES THEOREM

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

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

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