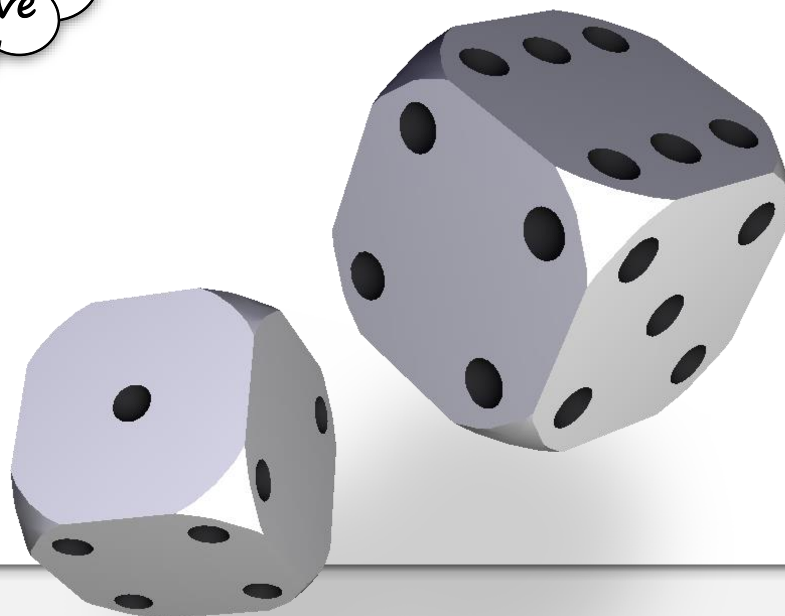
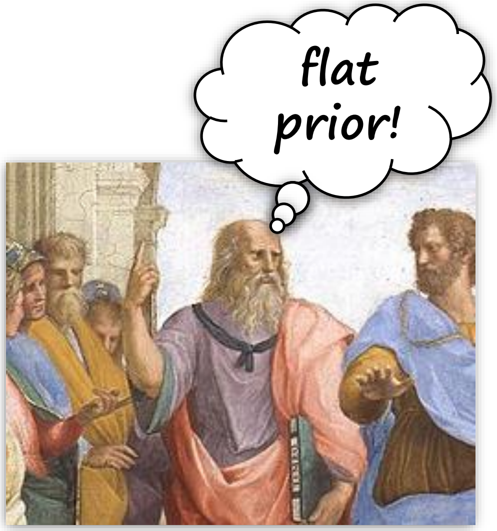
Modelling 2

STATISTICAL DATA MODELLING

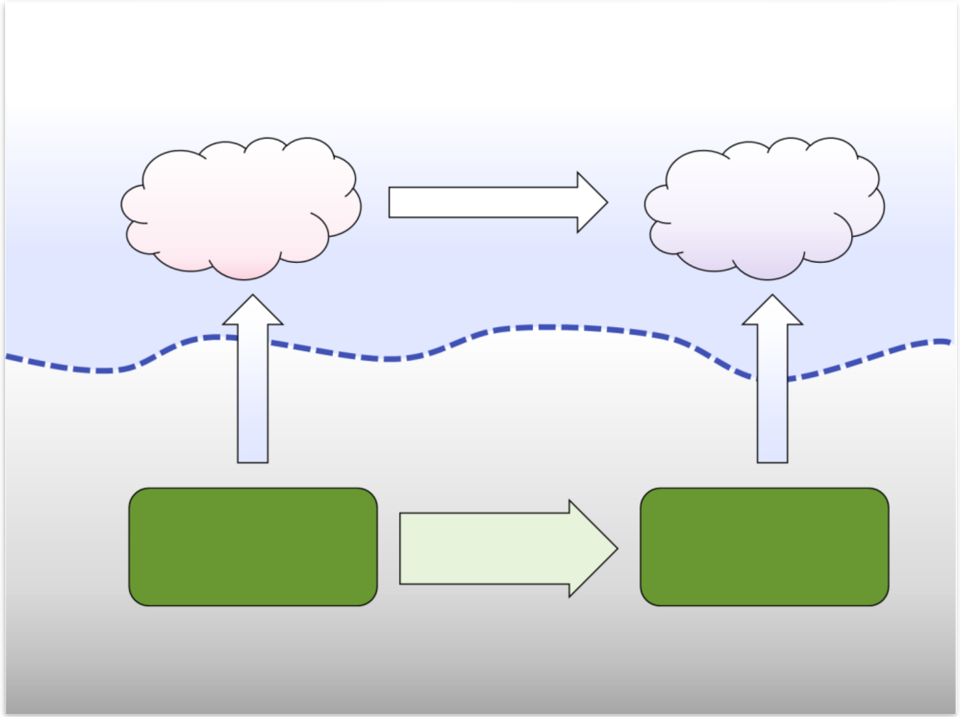




Chapter 2
Uncertainty

Michael Wand · Institut für Informatik, JGU Mainz · michael.wand@uni-mainz.de

Statistical Data Modeling



This lecture is about:

- ...understanding inductive reasoning
- ...done algorithmically / systematically

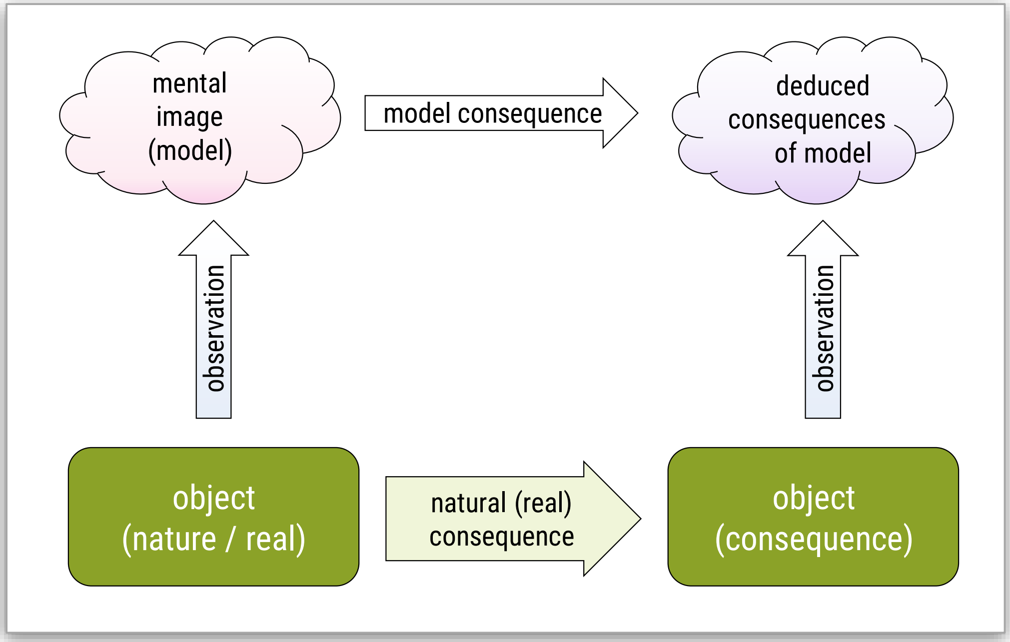
Our School of Thought

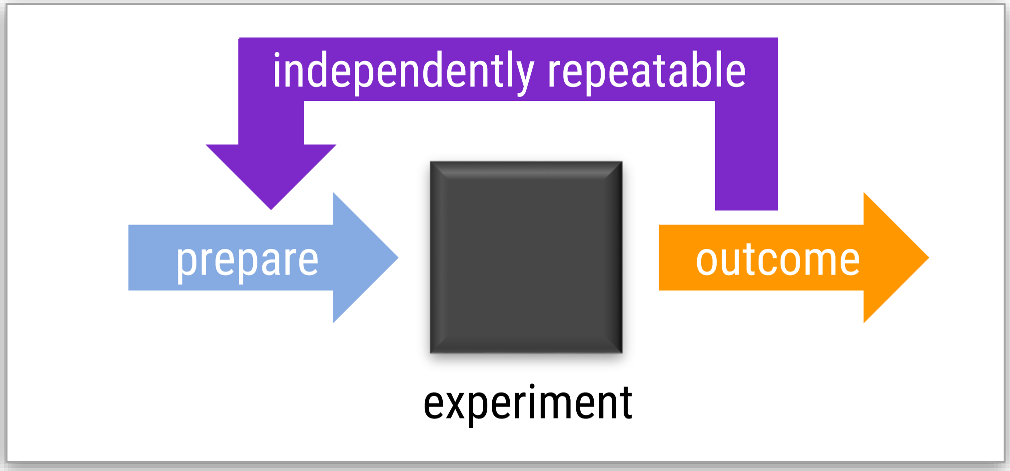
Empirical modeling

- Model for reality
  - Rely on observation
- Good models are
  - Predictive
  - Falsifiable

Learning from data

- Probabilistic
- Always comes with uncertainty





Probability Theory Recap

(skip ahead if familiar)

Modeling Uncertainty

Recap: Finite probability space (Ω, P)

- "*Sample space*" $\Omega = \{\omega_1, \ldots, \omega_n\}$
- "*Outcomes*" $\omega \in \Omega$
  - Exactly one $\omega \in \Omega$ will happen
- Probability $P(\omega) \in [0,1]$ for each $\omega \in \Omega$
  - The sum of all probabilities is 1.

Events

Event: Set of outcomes

- Sample space $\Omega = \{\omega_1, \dots, \omega_n\}$ (finite)
- Any subset $A \subseteq \Omega$ is called an "event"
- Rule: sum up

$$P(A) = \sum_{\omega \in A} P(\omega)$$

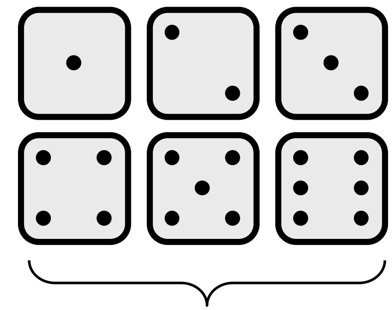
Example: Dice

- $P(\text{"odd"}) = P(\text{"1"}) + P(\text{"3"}) + P(\text{"5"})$
  $= 3 \times \frac{1}{6} = \frac{1}{2}$

Summary: Probability Measure

Basic Idea

- Every outcome has a likelihood
- Complex events: Sum up likelihoods



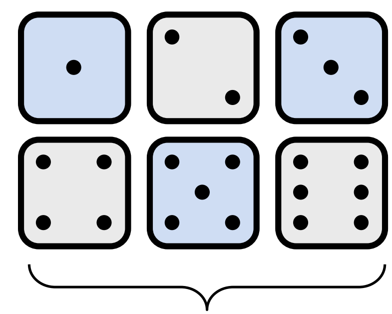
$P(\text{"1"}) = \ldots = P(\text{"6"}) = \frac{1}{6}$

"Learning" model from data

- Determine likelihood of outcomes

"Inferring" likelihood of events

- Sum up likelihoods of outcomes that lead to event



$P(\text{"odd"}) = \frac{1}{2}$

Formal Definition

Probability

Technical Complications

Basic stochastic lecture $\gg$ 5 slides

- Problems if Ω infinite
- Particularly relevant:
  - Real numbers as outcome
  - Real vectors as outcome
- Power set $\mathcal{P}(\mathbb{R})$ is not “measurable”
  - Cannot define consistent “sum” of probabilities

Technical Complications

Mathematical definition

- Replace set of all subset $\mathcal{P}(\Omega)$ by "set of reasonable subsets"
  - σ-Algebra of Ω
  - "Event space" $\mathcal{F}$
- Define $P(\text{event})$ as normed, non-negative, additive measure on that algebra

Intuition

- Same intuition: Summing up / integrating "*probability mass*" on domain

Kolmogorov's Axioms

Probability space

- Sample space: Ω
- Event space: $\mathcal{F}(\Omega) \subseteq \mathcal{P}(\Omega)$ ($\mathcal{F}$ is a σ-algebra)
- Events: $A \in S(\Omega)$
- Probability measure: $P: \mathcal{F} \to \mathbb{R}$

Axioms: Please behave like discrete case!

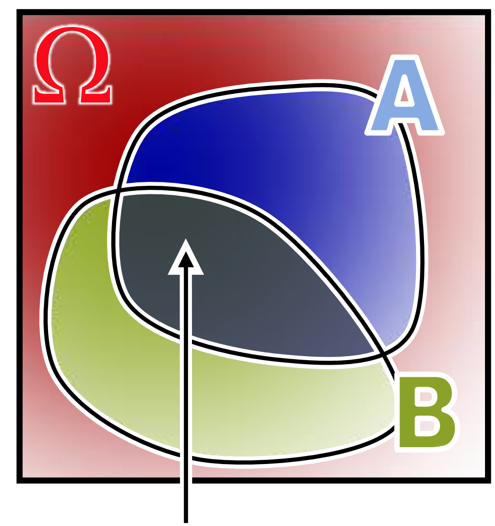
- Positive: $P(A) \geq 0$
- Additive: $[A \cap B = \emptyset] \Rightarrow [P(A) + P(B) = P(A \cup B)]$
- Normed: $P(\Omega) = 1$

Other Properties Follow

Derived from Kolmogorov's axioms

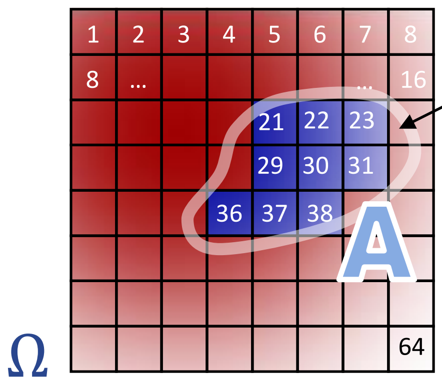
- $P(\bar{A}) \in [0..1]$
- $P(A) = P(\Omega \setminus A) = 1 - P(A)$
- $P(\varnothing) = 0$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- ...

We are still "summing up" density



counted twice

Discrete vs. General Model



p as "density" on Ω

A is an *event*

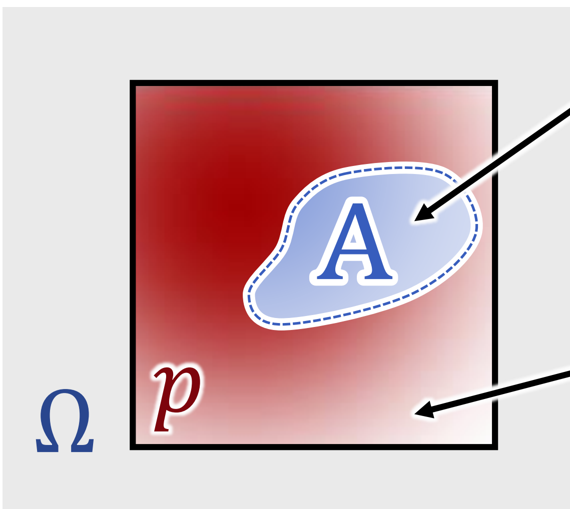
$$P(A) = \Sigma_{i \in A}\, p(w_i)$$
$$= p(w_{21}) + p(w_{22}) + p(w_{23})$$
$$+ p(w_{29}) + p(w_{30}) + p(w_{31})$$
$$+ p(w_{36}) + p(w_{37}) + p(w_{38})$$

Consistent with discrete model

Continuous Density

Major Motivation: Density model

- No elementary probabilities
- Instead: density $p: \mathbb{R}^d \to \mathbb{R}^{\geq 0}$



A is an *event*

$$P(A) = \int_A p(\mathbf{x})\, \mathrm{d}\mathbf{x}$$

Density $p(\mathbf{x})$ with

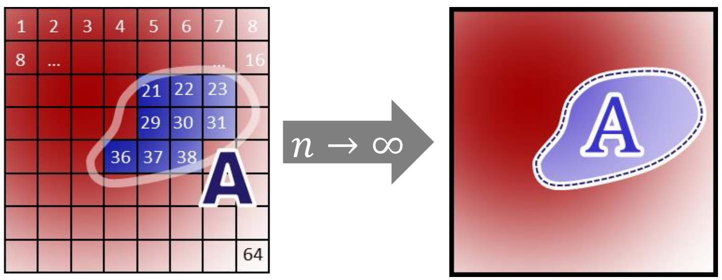
$$p(\mathbf{x}) \geq 0 \;\text{ and }\; \int_\Omega p(\mathbf{x})\, \mathrm{d}\mathbf{x} = 1$$

Probability Densities

Setup

- Domain $\Omega \subseteq \mathbb{R}^d$, outcomes $\mathbf{x} \in \mathbb{R}^d$
- Probability density
$$p: \Omega \to \mathbb{R} \quad \text{(integrable)}$$
- Properties
$$\forall \mathbf{x} \in \Omega: p(\mathbf{x}) \geq 0$$
$$\int_{\mathbf{x}\in\Omega} p(\mathbf{x}) d\mathbf{x} = 1$$
- Events
$$P(A) := \int_{\mathbf{x}\in A} p(\mathbf{x}) d\mathbf{x} \quad \text{(for } A \in \mathcal{B}(\Omega)\text{)}$$
($\mathcal{B}$ = Borel σ-algebra)

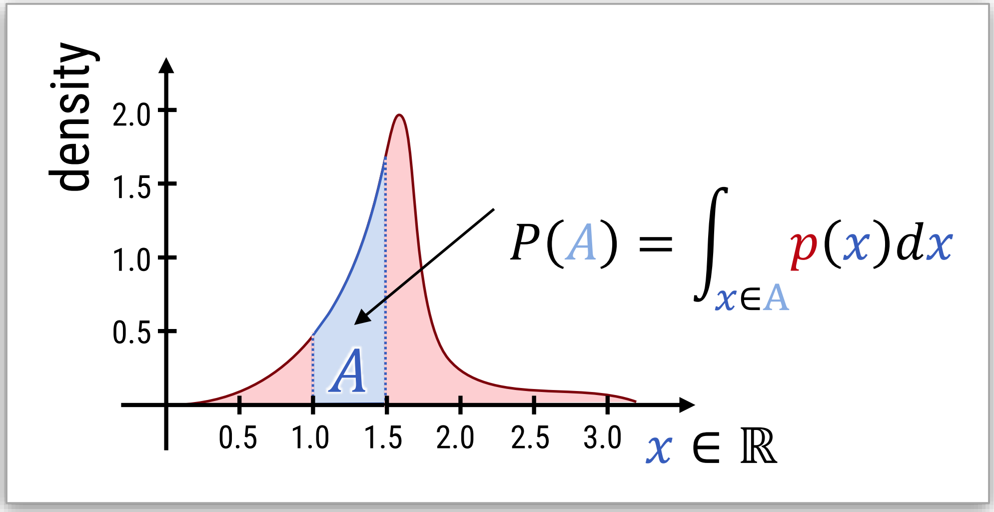
Continuous Density



Intuition

- Just "very small" outcome "buckets"

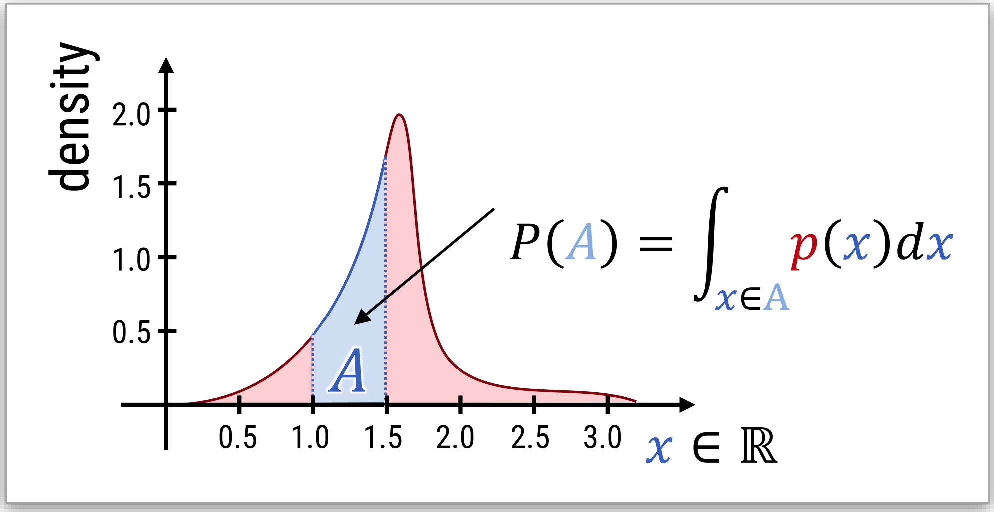
Probability Densities



Remarks

- Densities vs. probability
  - $P(A)$ to denote probability of events/outcomes
  - $p(\mathbf{x})$ to denote probability densities
- Only integrals of p are probabilities

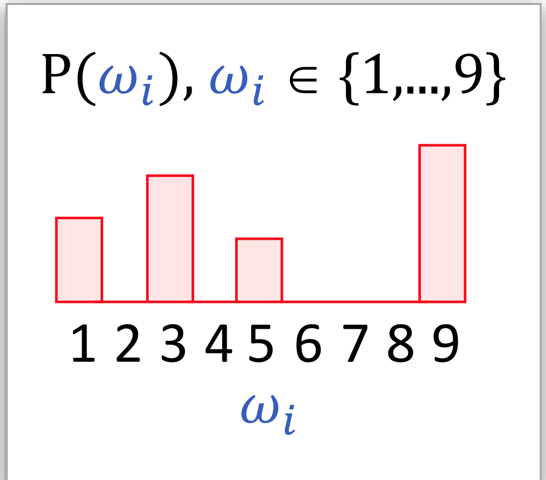
Probability Densities



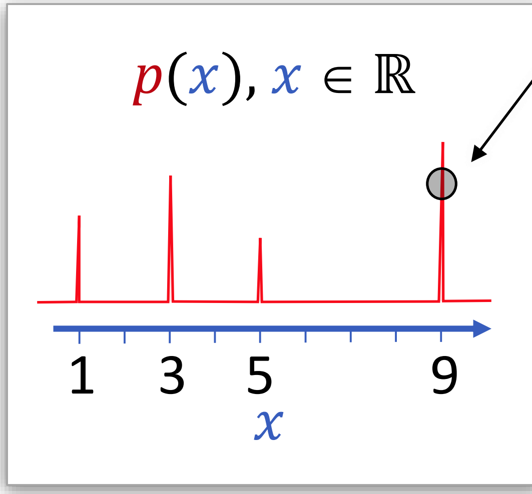
Remarks

- Remark: $p(\mathbf{x}) > 1$ is possible as long as $\int p = 1$
  - $p(\mathbf{x})$ are not probabilities, but densities

Probability Densities



discrete model



continuous model

Dirac-Delta pulses

$$p(x) = \Sigma_i \, \delta(x - \omega_i) \, P(\omega_i)$$

Intuition: (Modeling 1)

$$\int_{\mathbb{R}^d} \delta(x) \, dx = 1$$

$\delta(x)$ „very large close to x“

$\delta(x) = 0$ everywhere else

Remarks

- Discrete models through Dirac densities
- We will use this as much as possible to unify notation

Random Variables

Naming convention

- Sample space Ω with probability measure P
- Mapping $X \colon \Omega \to \mathbb{R}^d$ is called "random variable"
  - Often equivalent to $\Omega = \mathbb{R}^d$
  - $X = \mathbf{x}$ can be an "elementary" outcome, but does not have to

Description with densities

- We describe random variables with densities

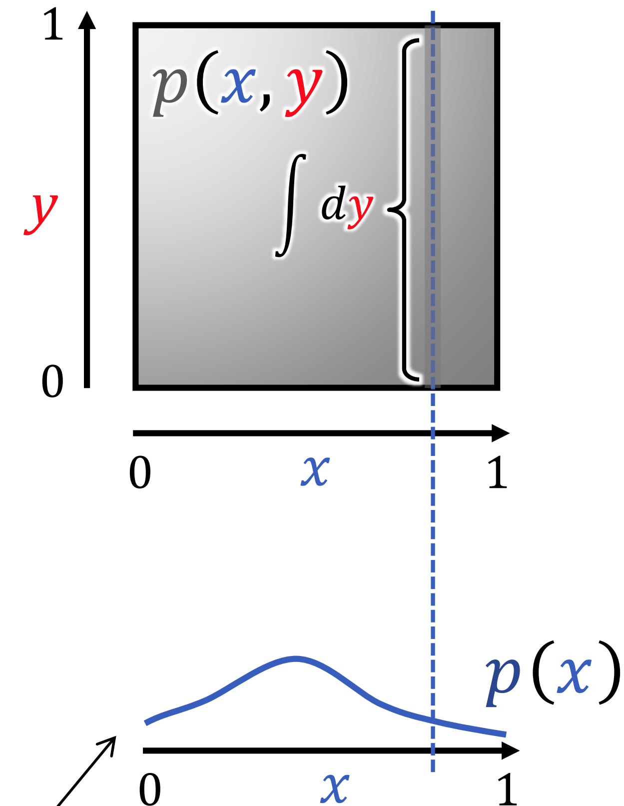
$$p(\mathbf{x}) = \text{probability density for “}X = \mathbf{x}\text{”}$$

Marginals

Example

- Random variables $X, Y \in [0,1]$
- Joint distribution $p(x, y)$
- We do not know y
  (could by anything)
- What is the distribution of x?

$$p(x) \coloneqq \int_0^1 p(x, y) dy$$

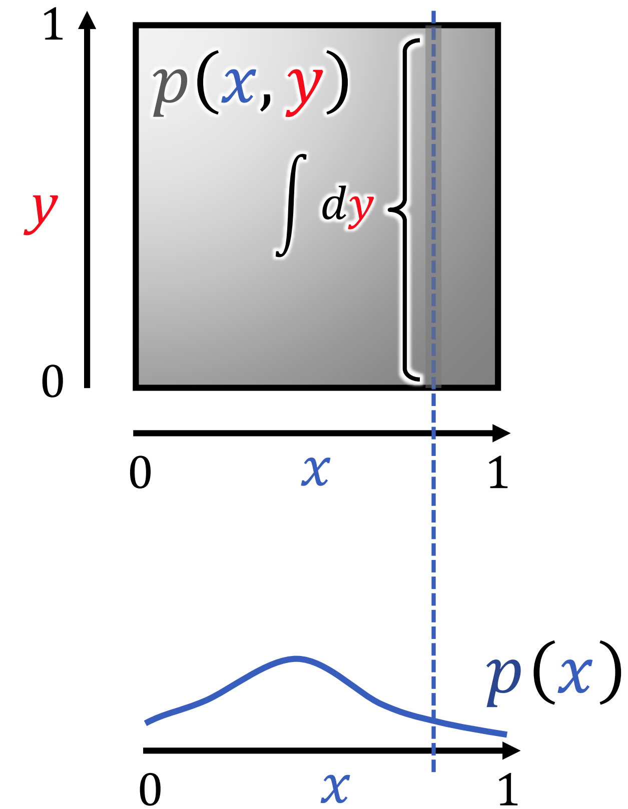


"Marginal Probability"

Marginals

General rule

- Marginal probability
  - Integrate / sum over all unspecified
- Specified variables
  - What we care about
  - Often: observed / measured
- Unspecified variables
  - Not relevant in this context
  - Might be "latent" (unobservable)
  - Might be model parameters (more later)



"Marginal Probability"

Summary

What we have seen so far...

Probability space

- Density on some domain, sums up to 100%

Probability densities

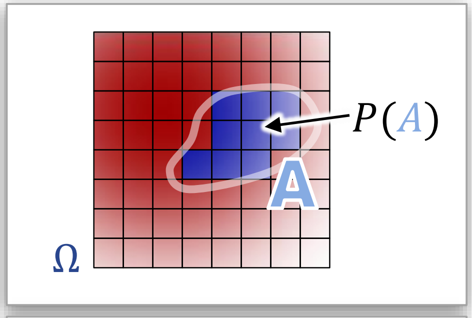
- Continuous elementary outcomes

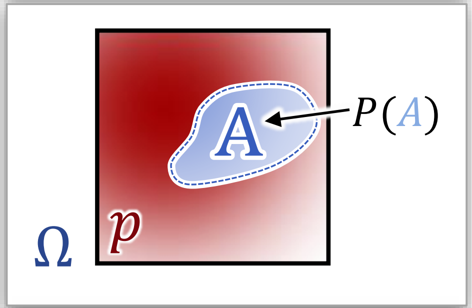
Events

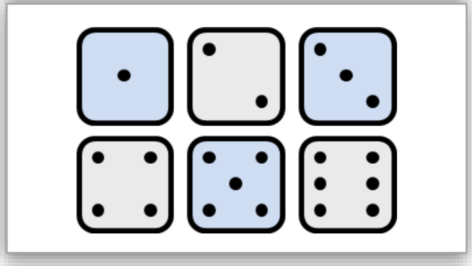
- Subsets (that can be measured)

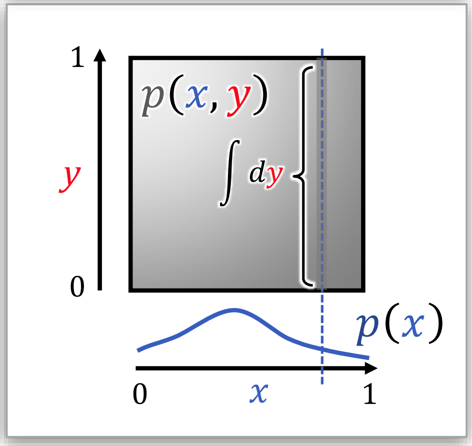
Marginal distributions

- Distribution for events (subsets) where we have only partial information:
  $$p(\mathbf{x}, \mathbf{y}) \to p(\mathbf{x})$$









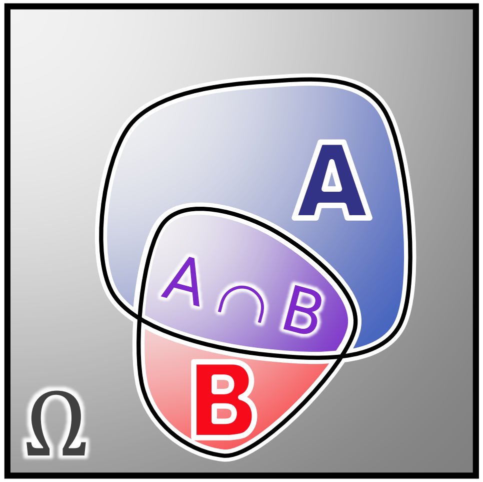
Statistical Dependency

Conditional Probability (Rnd-Var.)

Conditional Probability

- $\mathrm{P}(A\,|\,B)$ = Probability of A given B [is true]

- Definition
  $$\mathrm{P}(A \cap B) = \mathrm{P}(A|B) \cdot \mathrm{P}(B)$$



Corollary

- If $P(B) \neq 0$:

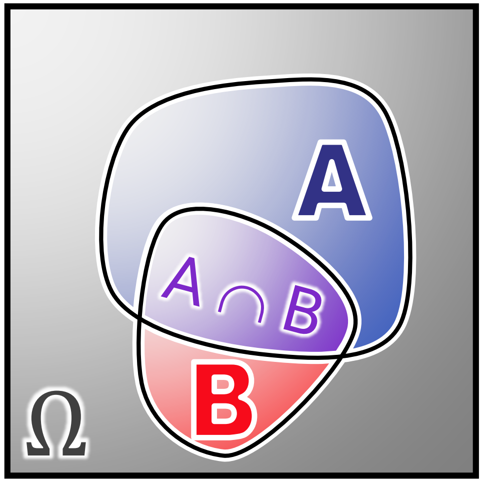
$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}$$

Conditional Probability

Statistical Independence

- **Definition**

$$A \text{ and } B \text{ independent}$$
$$\Leftrightarrow \; \mathrm{P}(A \cap B) = \mathrm{P}(A) \cdot \mathrm{P}(B)$$



- Knowing the value of A does not yield information about B
  - And vice versa
- Also: $\mathrm{P}(A \cap B) = \mathrm{P}(A) \cdot \mathrm{P}(B)\;(= \; \mathrm{P}(A|B) \cdot \mathrm{P}(B))$

  means that $\mathrm{P}(A|B) = \mathrm{P}(A)$, and $\mathrm{P}(B|A) = \mathrm{P}(B)$

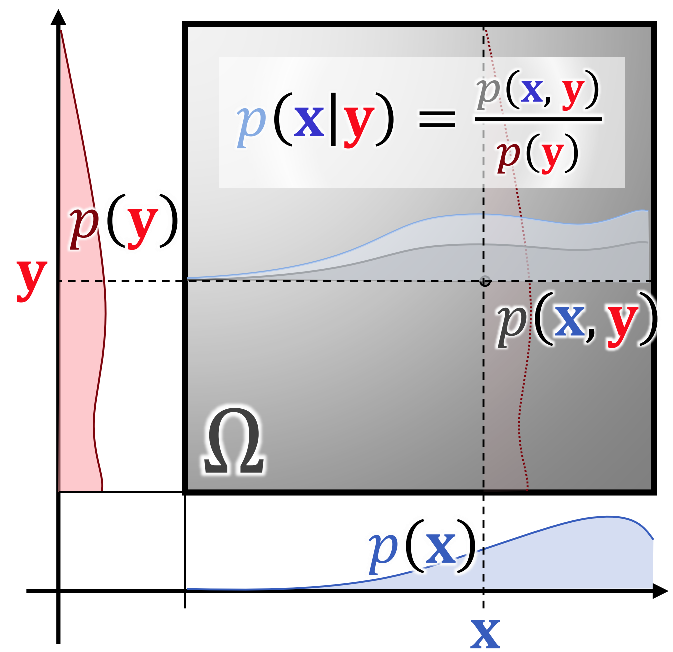
Random Variables

Conditional Probability

- $p(\mathbf{x}|\mathbf{y})$ = Probability density of $\mathbf{x}$ given $\mathbf{y}$ [has occured]

- Definition

$$p(\mathbf{x},\mathbf{y}) \;=\; p(\mathbf{x}|\mathbf{y}) \cdot p(\mathbf{y})$$



Corollary

- If $p(\mathbf{y}) \neq 0$:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{y})}$$

Conditional Probability

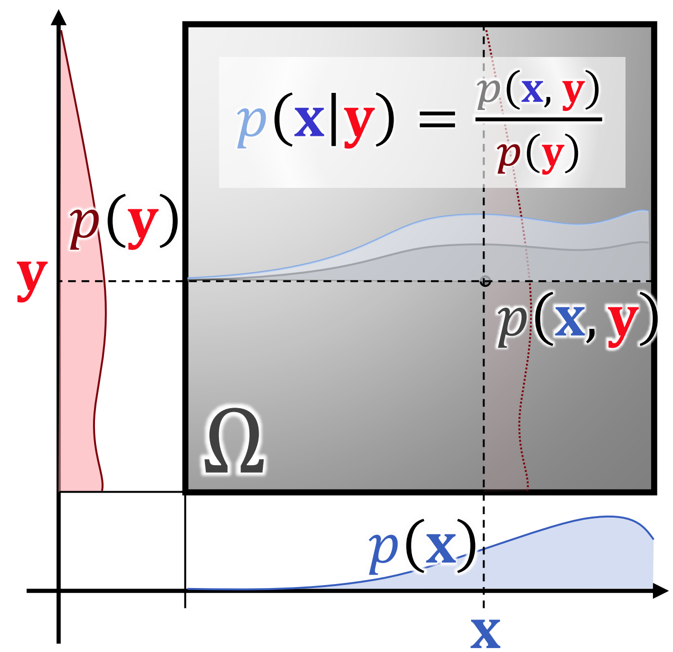
Statistical Independence

- **Definition:**

  $\mathbf{x}$ and $\mathbf{y}$ independent

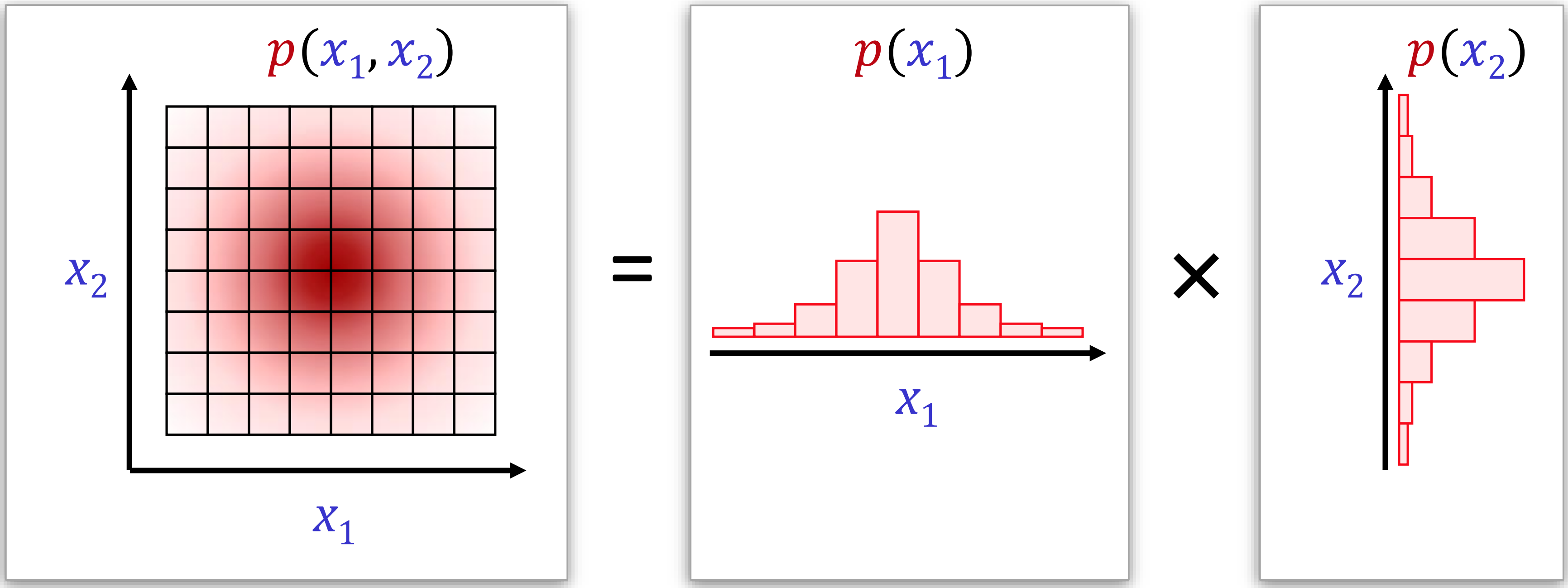
  $\Leftrightarrow p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}) \cdot p(\mathbf{y})$

- Knowing the value of $\mathbf{x}$ does not yield information about $\mathbf{y}$ (and vice versa)
  - $p(\mathbf{x}|\mathbf{y}) = p(\mathbf{x})$
  - $p(\mathbf{y}|\mathbf{x}) = p(\mathbf{y})$



Factorization

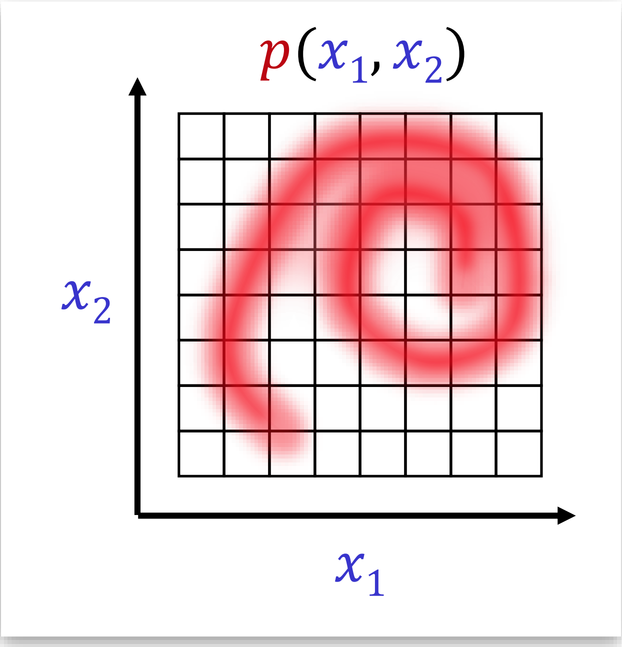
Independence = Density Factorization



$$p(x_1, x_2) = p(x_1) \times p(x_2)$$

Factorization

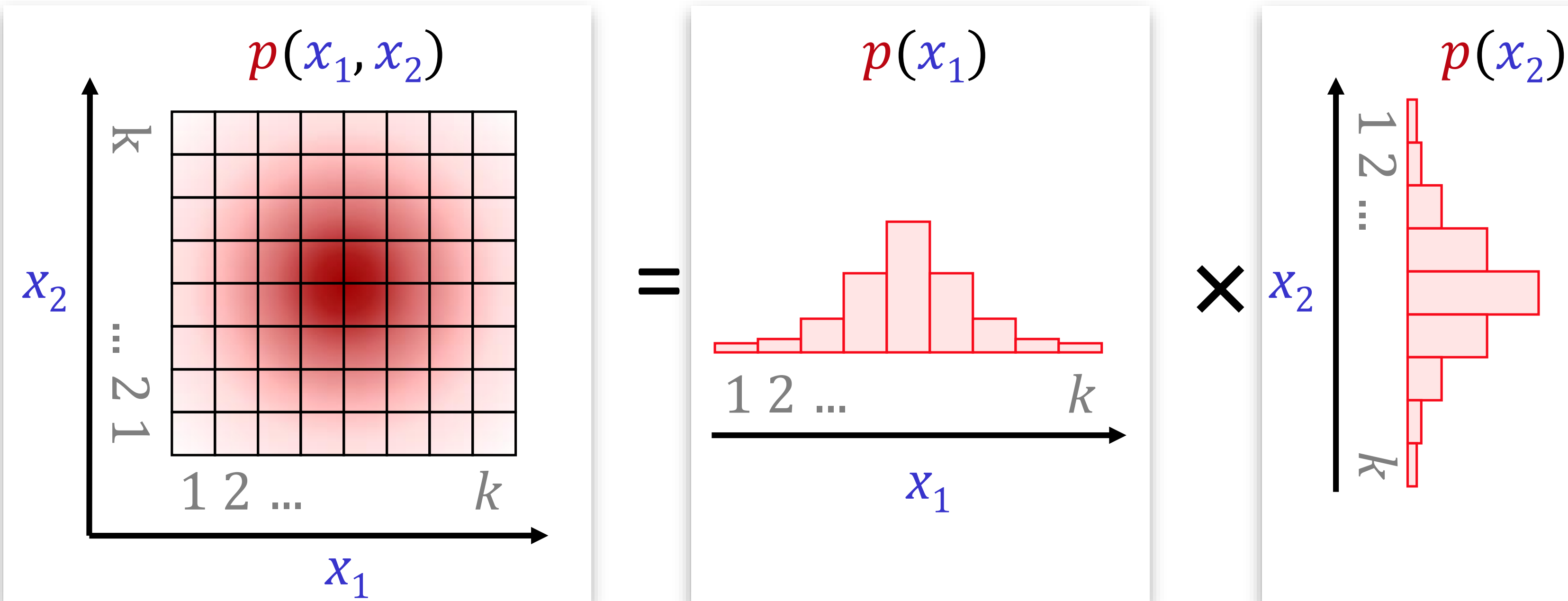
Not Independence → No Factorization



$$= \; p(x_1, x_2)$$

Factorization

Independence = Density Factorization



$$p(x_1, x_2) = p(x_1) \times p(x_2)$$

$$O(k^d) \qquad O(d \cdot k)$$

Complexity

Curbing complexity

- n pieces of information (bits)

  $\rightarrow$ up to 2^n different combinations

  $\rightarrow$ up to 2^n different probabilities

- Statistical dependencies
  - Arbitrary structure:
    all combinations might matter
  - Fully independent: linear
    $2n$ instead of 2^n
  - Truth is "in between"
    Restricted dependencies make model feasible

More Drastic Example

Random Images

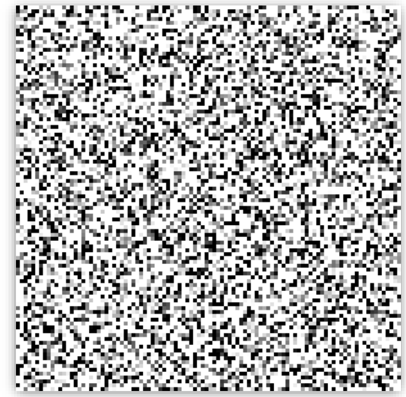
- 100 x 100 pixel
- 8 bit (256 grey values)

Independent Pixels

- $256 \times 100^2 = 2\,560\,000$ probability values

Arbitrary Dependencies

- $256^{100^2} = 2.51 \times 10^{24\,082}$ possible images / probabilities



independent

complex dependency
(M-GAN)

Modeling Examples

How to build a probability space?

Statistics appears unintuitive

- Often: Choice of Ω major problem
- Looking at events can be misleading
- Often: higher dimensionality needed

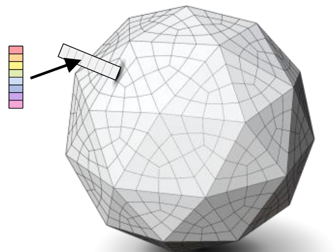
How to build a probability space?

Example: Weather in Mainz

- Interesting events: {*rain*, *sunshine*, *cloudy*}

Model 1: Low-level

- Sample space: Ω = Set of all states of the earth's atmosphere
  - ICON weather model: 265M grid cells, 10 (major) variables
- Define events by thresholds
  - Water / ice content
- Very expensive (too expensive?)
  - But captures the situation quite comprehensively



2949120×87 cells

How to build a probability space?

Example: Weather in Mainz

- Interesting events: {*rain*, *sunshine*, *cloudy*}

Model 2a: Event-level

- Problematic: $\Omega = \{rain, sunshine, cloudy\}$
- Not mutually exclusive
  - Sun can shine during rain
  - Complex dependencies need to be captured
- Not suitable for reasoning about the weather

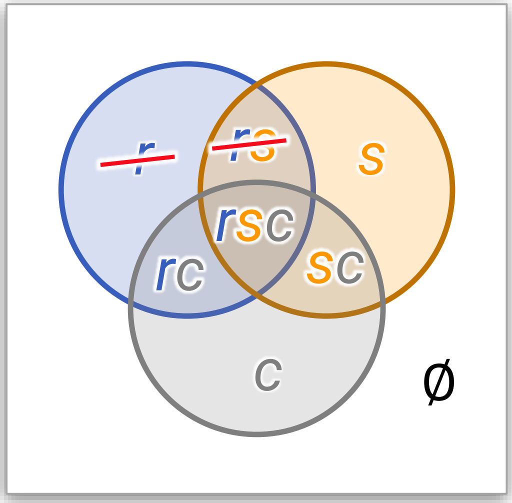
How to build a probability space?

Example: Weather in Mainz

- Interesting events: {*rain*, *sunshine*, *cloudy*}

Model 2b: Event-level

- All combinations:
  $\Omega = \{rsc, sc, rc, \cancel{rs}, \cancel{r}, s, c, \emptyset\}$
- All possible combinations of events
  - Some might be impossible, i.e., $P = 0$
- Exponential costs
  - 2^n outcomes for n Boolean variables
  - Not uncommon, if dependency structure is not known

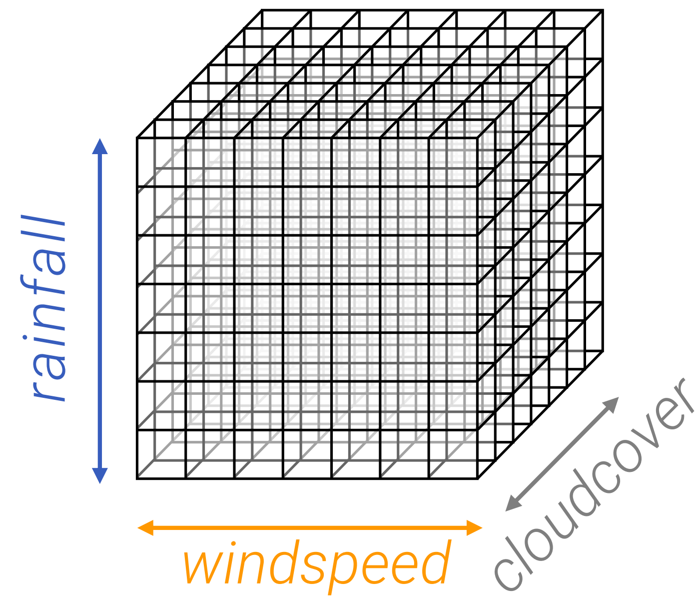


more knowledge:
no rain without clouds

How to build a probability space?

Example: Weather in Mainz

- Random variables:
  - *rainfall* [mm] ($\mathbb{R}$)
  - *windspeed* [m/s] ($\mathbb{R}$)
  - *cloudcover* [%] ($\mathbb{R}$)



Model 3: 3D Density

- Naïve discretization: Histogram/bins
- Again, exponential in number of variables
- k different values, n variables: k^n outcomes

How to build a probability space

Rules of thumb

- Define “experiment” clearly
- Collect variables
  - Observables & unobserved / latent parameters
- Assume all combinations have likelihood (densities)
  - Unless you know better
  - Model assigns probability for all relevant combinations
- If you know better
  - Restrict dependencies
  - Only then you can build a complex model

Summary

What we have seen so far…

Statistical independence

- Probability/density factorizes

$$p(x, y) = p(x) \cdot p(y)$$

- Dependency: potentially complex function structure

$$p(x, y)$$

Conditional probability

- Conditional density „x given y“: $p(x|y) = \frac{p(x,y)}{p(y)}$
  - Take joint density $p(x, y)$
  - Renormalize by $p(y)$ (because y has happened already)

Complexity

- Unrestricted dependencies lead to exponential model size

Calculus with Densities

Summary

tl;dw: Calculus

- Discussing functions $p: \Omega \to \mathbb{R}$
- Understanding them better:
  - Switch the basis /
    project on test-functions

Moments of Distributions

Density Function (1D)

- $p: \mathbb{R} \to \mathbb{R}^{\geq 0}$

Expected Value / Mean:

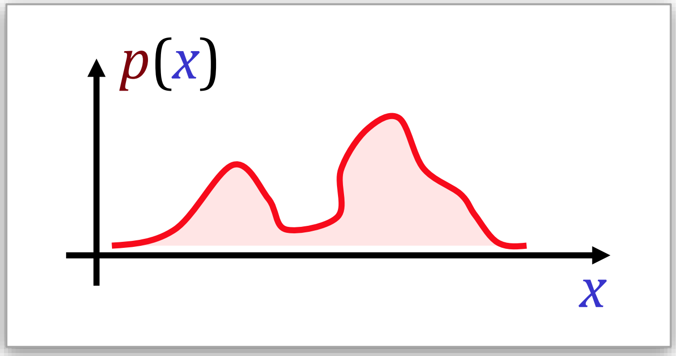
- $E(p) = \mu := \langle p, x \rangle$

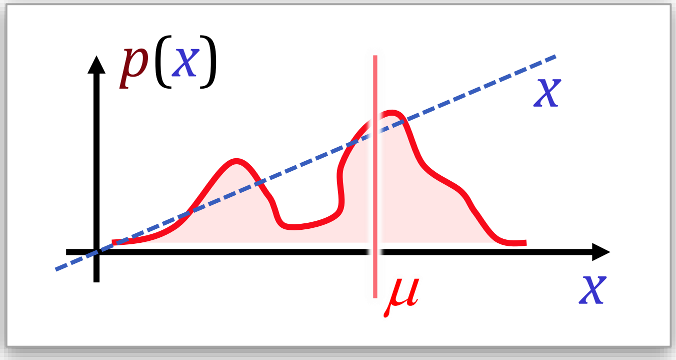
$$= \int_{\mathbb{R}} p(x) \cdot x \, dx$$

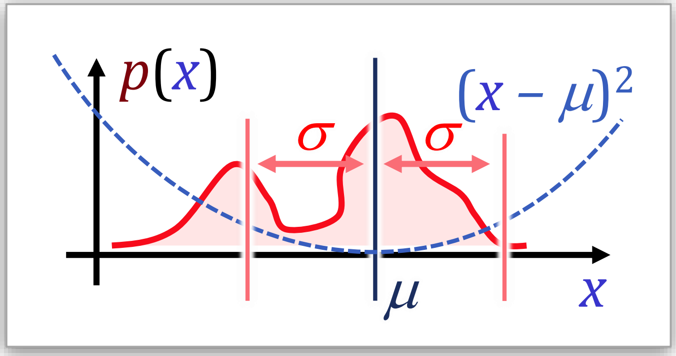
Variance:

- $Var(p) = \sigma^2 := \langle p, (x - \mu)^2 \rangle$

$$= \int_{\mathbb{R}} p(x) \cdot (x - \mu)^2 \, dx$$







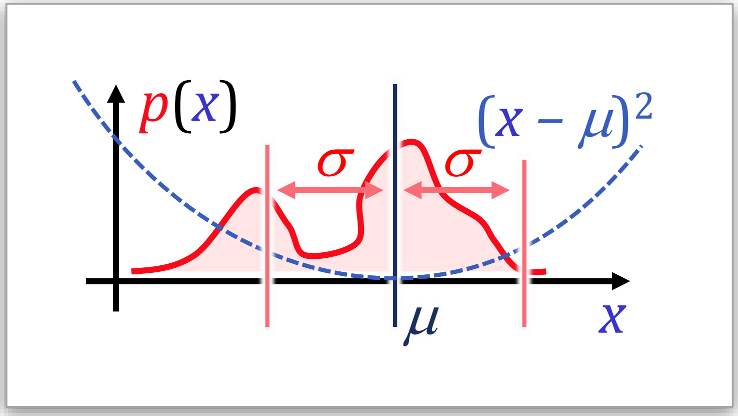
Standard Deviation

Bounds on spread

- Standard deviation

$$\sigma = \sqrt{Var(p)}$$

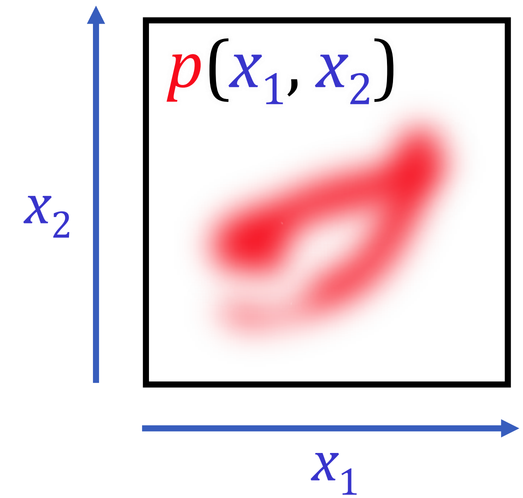
- Expected range of variation

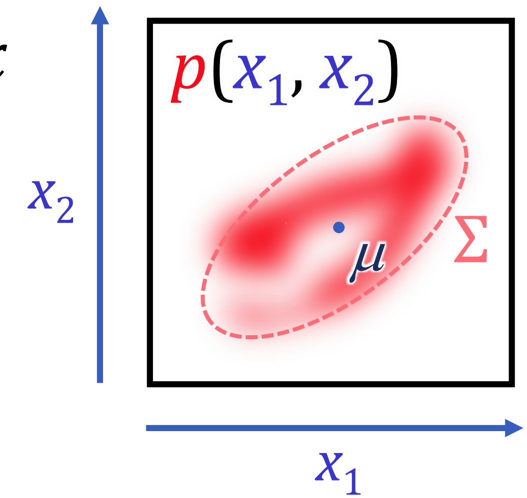


Moments of Distributions

Multi-variate density function

- Density $p\colon \mathbb{R}^d \to \mathbb{R}^{\geq 0}$
- $E(p) = \mu := \langle p, \mathbf{x} \rangle = \int_{\mathbb{R}^d} p(\mathbf{x}) \cdot \mathbf{x} \, dx$
- $\mathrm{Cov}(x_i, x_j) := \langle p, (x_i - \mu_i)(x_j - \mu_j) \rangle$
  $$= \int\limits_{\mathbb{R}^d} p(\mathbf{x}) \, (x_i - \mu_i)(x_j - \mu_j) \, dx$$
- $\Sigma = \begin{pmatrix} \ddots & \vdots & ⋰ \\ \cdots & \mathrm{Cov}(x_i, x_j) & \cdots \\ ⋰ & \vdots & \ddots \end{pmatrix}$





Properties

Expected value

- $E(X+Y) = E(X) + E(Y)$
- $E(\lambda X) = \lambda E(X)$

Variance

- $\mathrm{Var}(\lambda X) = \lambda^2 \mathrm{Var}(X)$
- Let X, Y be *independent*, then:
  $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

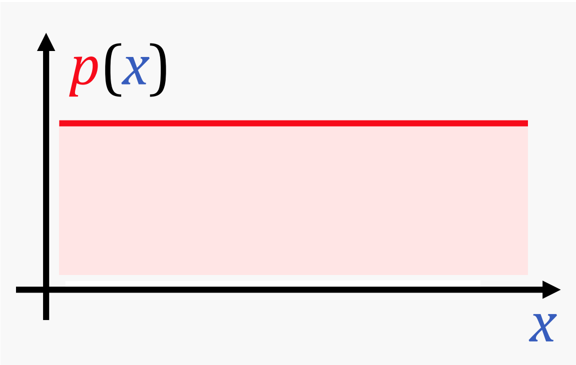
Entropy

(There will be a whole video on this)

Entropy

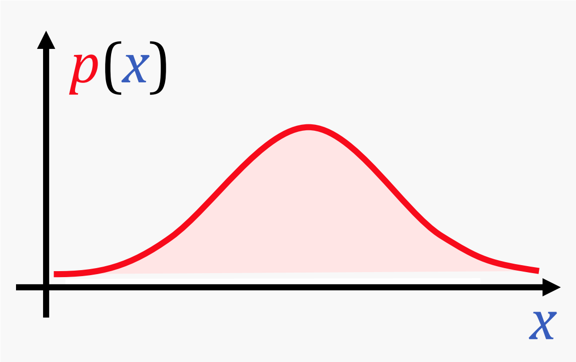
Entropy: How random?

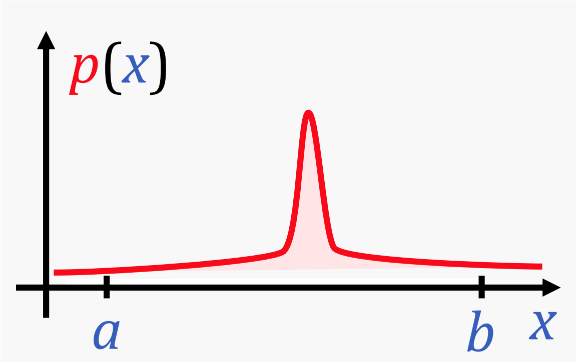
$$H(X) = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i)$$



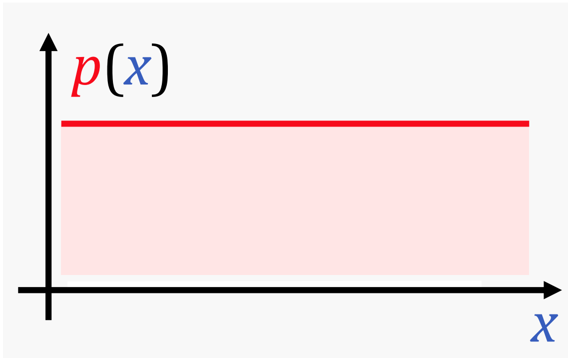
Model

- Binary coding
- $\mathcal{O}\left(\log \frac{1}{p}\right)$ bits for…
- …events with probability p

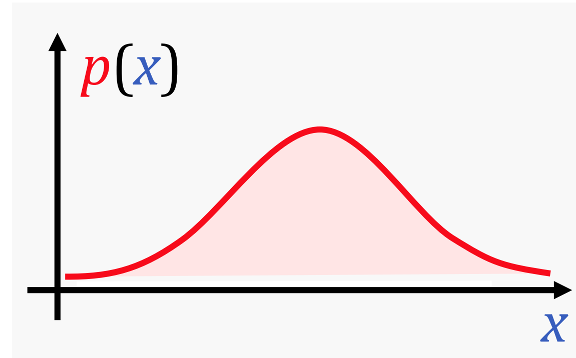


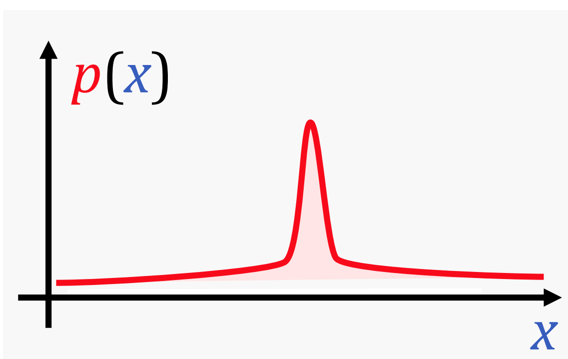


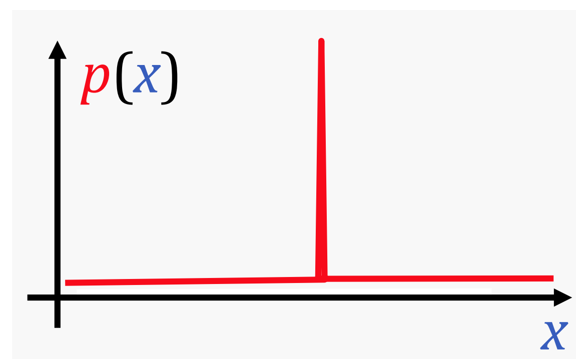
Examples



$$H = -\sum_{i=1}^{n} \frac{1}{n} \log \frac{1}{n} = \log n$$





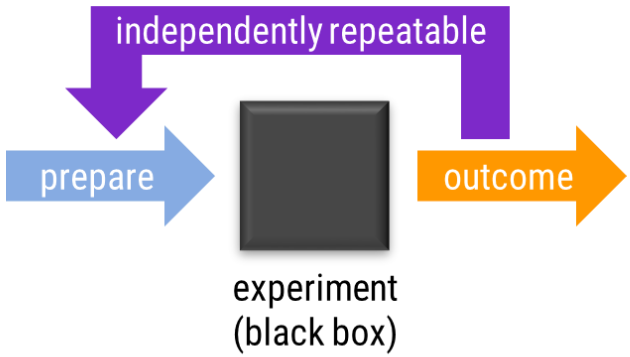


$$H = 0$$

Limits:

Repeating Experiments

Law of Large Numbers



Repeated experiment

- Experiment, outcome $x \in \mathbb{R}$
- Repeated n times

We look at the mean

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

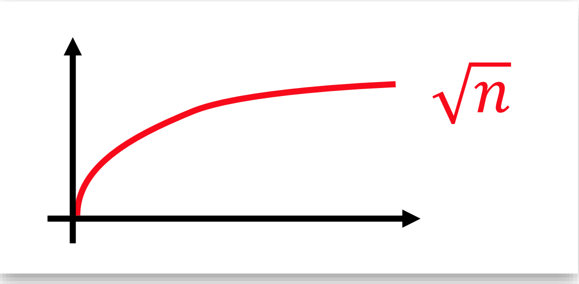
(Weak) law of large numbers

$$\lim_{n\to\infty} \Pr(|\bar{X}_n - \mu| > \epsilon) = 0$$

Stochastic Convergence

Averaging of independent trials

- Convergence rate is $\frac{1}{\sqrt{n}}$
- Lousy convergence rate



Proof

Proof: weak law of large numbers

- Additionally assumption: finite variance $\mathrm{Var}(X_i) = \sigma^2$
- The theorem then follows from
  - Additivity of variances
  - Chebyshev's bound

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

$$\Rightarrow \sigma(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$$

- Chebyshev: $\Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$

Algebra with Random Variables

Random Variable Vector Algebra

Vector algebra

- Given independent random variables X, Y
- Look at operation $Z = f(X, Y)$ with $\Omega_Z = \Omega_X \times \Omega_Y$

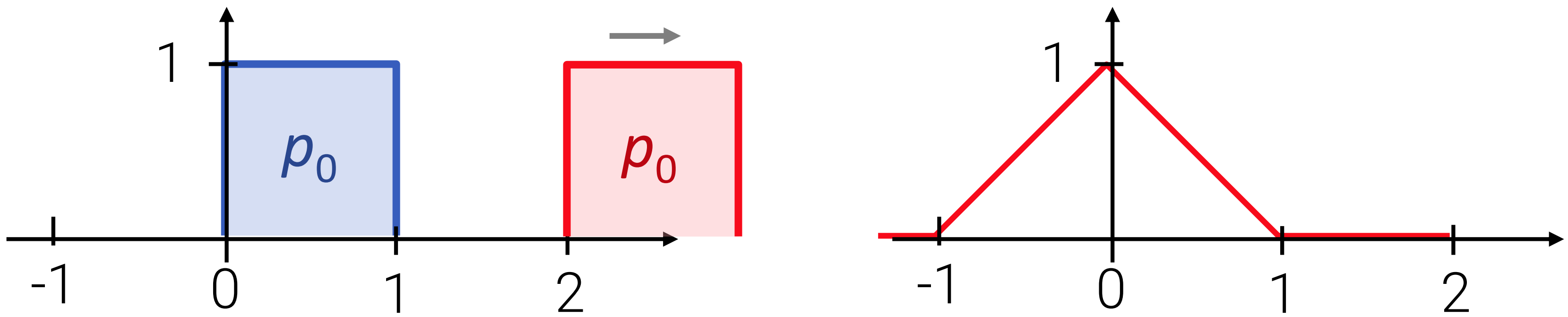
Scaling random variables

- Scaling variable: $Z = \lambda X$ (Factor λ not random)
- Scaling variable: $p_{\mathbf{z}}(\mathbf{z}) = p_{\mathbf{x}}\left(\frac{1}{\lambda}\mathbf{z}\right)$

Adding independent random variables

- Adding variables: $Z = X + Y$
- Convoling densities: $p_{\mathbf{z}}(\mathbf{z}) = p_{\mathbf{x}}(\mathbf{x}) \otimes p_{\mathbf{y}}(\mathbf{y})$

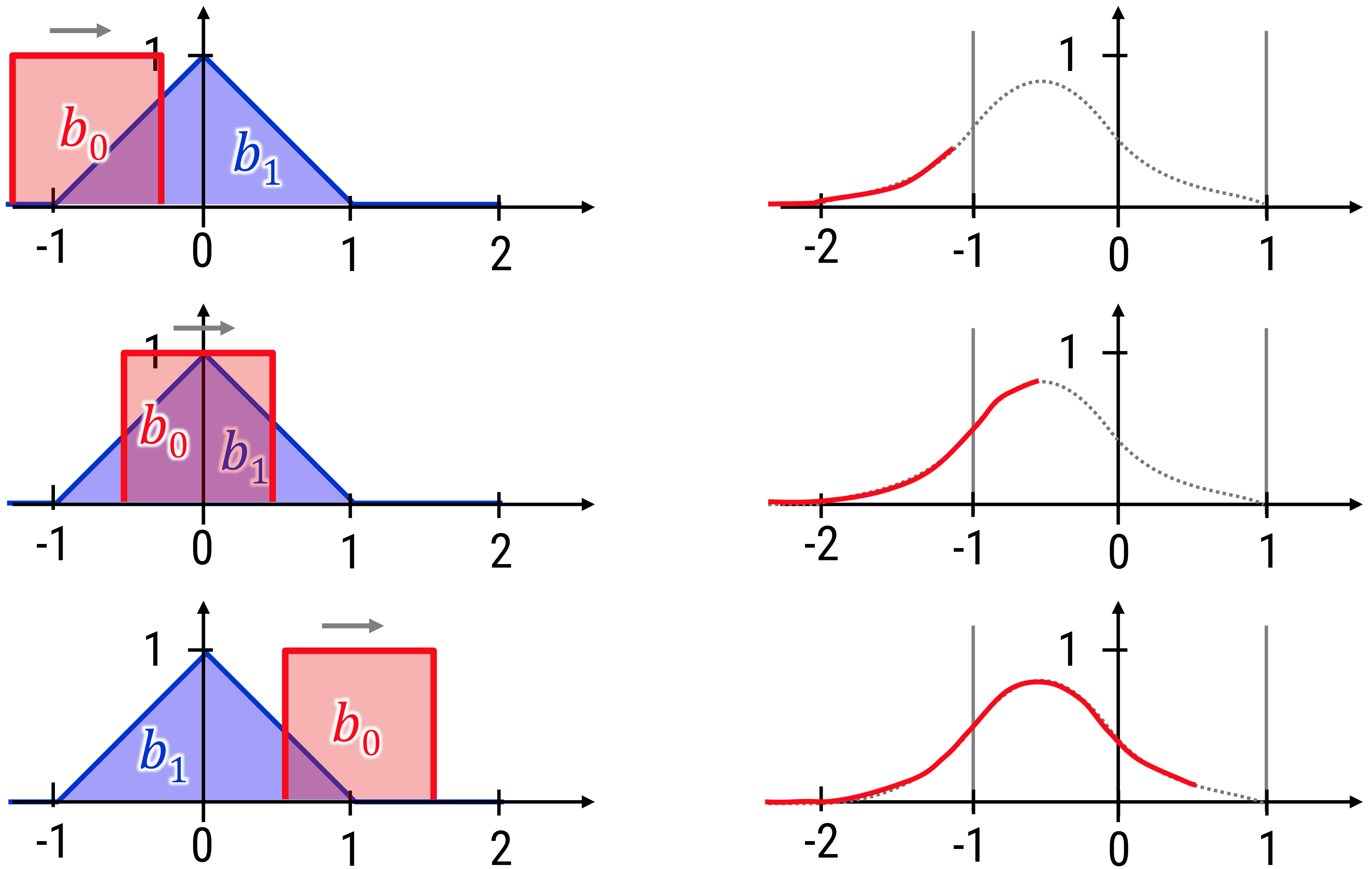
Convolution Example



Uniform distribution on $[0,1]$:

- "Box" function
- Auto-convolution yields "triangle" function
- Remark: Increases smoothness by one order

Illustration



Remarks

Repeated auto-convolution

- Of a uniform distribution
  - Yields increasingly smooth functions
  - Called “B-splines of order k” (for k-fold convolution)
  - Converges to Gaussian normal distribution
- Of general distributions
  - Converges to special limit distributions
  - Gaussian if mean and variance exist
    - Even if distributions are different (but independent)
    - “Central limit theorem”

Central Limit Theorem

Why are so many phenomena normal-distributed?

- Let $X_1, \ldots, X_n$ be real (1D) random variables with means μ_i and *finite* variances σ_i^2.
- Then the distribution of the mean

$$\frac{\sum_{i=1}^{n} X_i \; - \; \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}} \;\rightarrow\; \mathcal{N}(0,1)$$

converges to a normal distribution.

Multi-dimensional variant

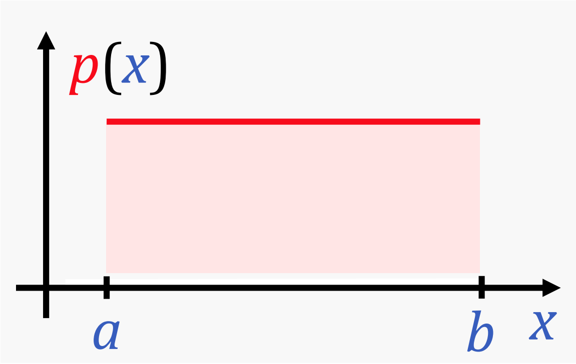
- Similar result for multi-dimensional case

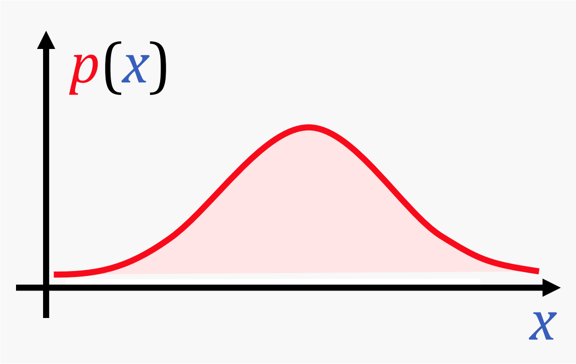
Common Parametric Distributions

Well-known probability distributions

Important distributions

- Uniform distribution
  - Only defined for finite domains
  - Maximum entropy
    among all distributions
- Binomial distribution
  - Coin-flipping
  - (one bit at a time)

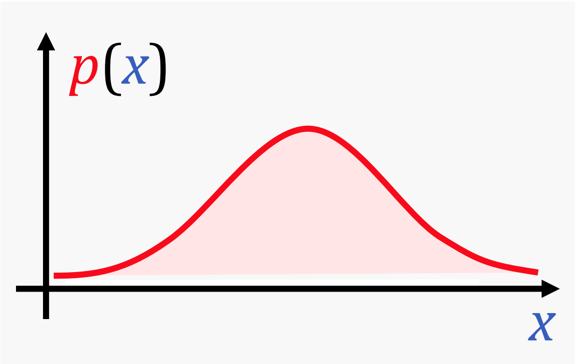


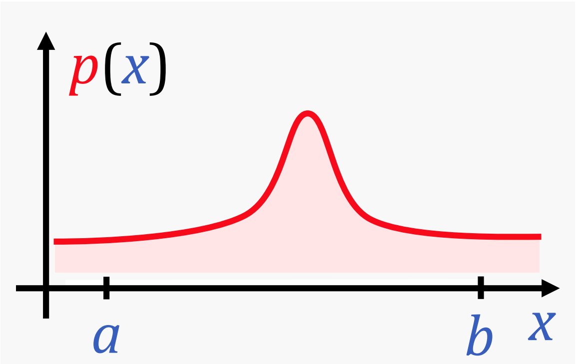


Well-known probability distributions

Important distributions

- Gaussian / normal distribution
  - Infinite domains
  - Maximizes entropy
    for fixed variance
- Heavy tail distributions
  - "Outlier robust"
  - Example: Exponential/Laplace/L1
  - Drops-off "slower than Gaussian"





Uniform distribution

What should we say?

- Fixed domain Ω with…
- …finite area $|\Omega| = \int_\Omega 1 d\mathbf{x} < \infty$
- Density

$$p(x) = \frac{1}{|\Omega|}$$

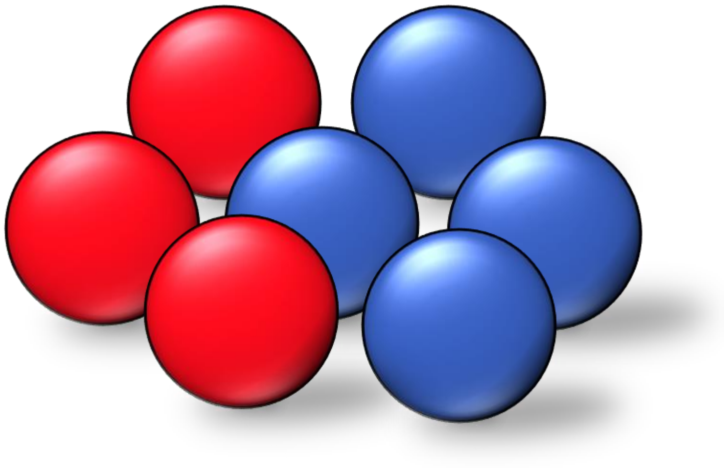
Attention

- No uniform distribution on infinite domains
- No "uniform distribution on $\mathbb{R}$"

Binomial Distriubution

Binomial Distribution

- Two possible outcomes “1”,”0”
- Probabilities p, $(1-p)$
- Repeated n times i.i.d.



Formulas

- $p(k \text{ times "1"}) = \binom{n}{k} p^k (1-p)^{n-k}$
- $\mu = np$
- $\sigma^2 = np(1-p)$
- Asymptotically ($n \to \infty$) Gaussian (CLT)

Gaussians

Gaussian Normal Distribution

- Two parameters: μ, σ
- Density:

$$\mathcal{N}_{\mu,\sigma}(x) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Mean: μ
- Variance: σ^2

Gaussian normal distribution

Log Space

Neg-log-density

$$\log \mathcal{N}_{\mu,\sigma}(x) := \frac{(x-\mu)^2}{2\sigma^2} + \frac{1}{2}\ln(2\pi\sigma^2)$$
$$\sim \frac{1}{2\sigma^2}(x-\mu)^2$$

Calculations in log-space

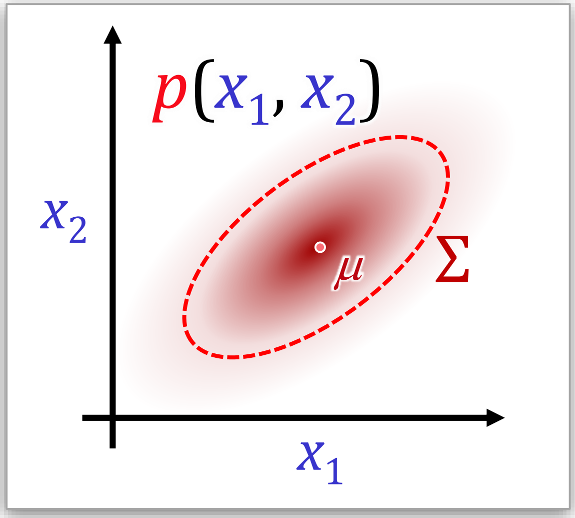
- Densities of products of Gaussians are Sums of quadratic polynomials
- Calculations simplified in log-space
  - Attention: Sum of Gaussians do not simplify!

→ Modelling 1

Multi-Variate Gaussians

Gaussian normal distribution in d dimensions

- Two parameters
  - Mean $\boldsymbol{\mu}$ (d-dim-vector)
  - Covariance matrix $\boldsymbol{\Sigma}$ ($d \times d$ matrix)



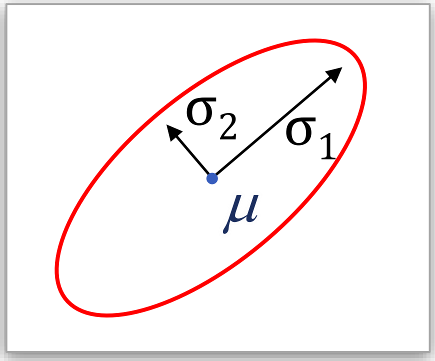
- Density

$$\mathcal{N}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{x}) := \left(\frac{1}{(2\pi)^{-\frac{d}{2}} \det(\boldsymbol{\Sigma})^{-\frac{1}{2}}}\right) e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

Log Space

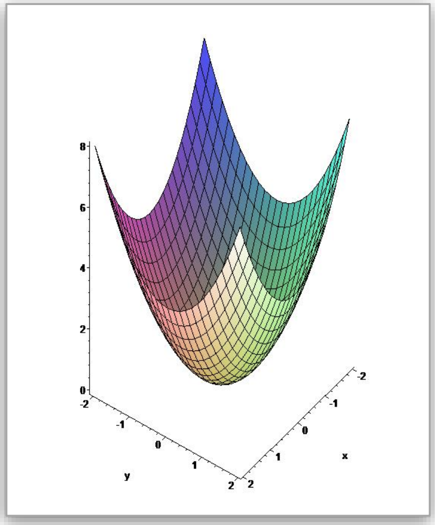
Neg-Log Density

- $\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) + const$
- Quadratic multivariate polynomial



Consequences

- Optimization (maximum density) $\rightarrow$ linear system
- Gaussians are ellipsoids
  - Eigenvectors of Σ are main axes
  - Eigenvalues are extremal variances



Example: A “Heavy Tail”-Distribution

More spread out than Gaussian

- Exponential distribution

$$p(x) := \lambda e^{-\lambda|x|}$$
$$x \geq 0$$

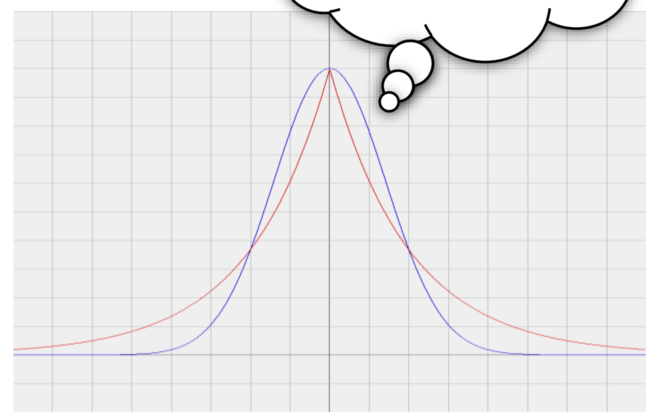
  - Mean: λ^{-1}
  - Variance: λ^{-2}

- Laplace distribution

$$p(x) := \frac{1}{2}\lambda e^{-\lambda|x-\mu|}$$
$$x \in \mathbb{R}$$

  - Mean: μ
  - Variance: $2\lambda^{-2}$

Mod-I: L_2 vs. L_1-norm fitting



Gaussian vs Laplace distribution
(height normalized)

Summary

What we have seen so far…

Moments

- Mean, variance, etc…
- Project density on polynomials

Limits

- Weak law of large numbers
- Central limit theorem (finite variance)

(Some) Standard distributions

- Binomial distribution
- Gaussian normal distribution
- Exponential / Laplace distribution