Lecture 2: Recap of Probability Theory

Prof. Dr. Mario Fritz

2022 04 20

https://fritz.cispa.saarland
Trustworthy AI
CISPA Helmholtz Center for Information Security, Saarbrücken, Germany

Outline

Main references

- Statistics Lab notes by Prof. Wolf
- Bishop - Chapter 1.2

Outline

Why probability theory in ML course

- A key concept in ML is uncertainty.
- Source of uncertainty are diverse and include the noise in the measurements (i.e., in the observed data) and the finite sample size from the underlying data distribution.
- Probability theory gives a theoretical framework to reason under uncertainty, i.e., to quantify and manipulate uncertainty.

- **Frequentist interpretation:** Probability as the frequency or propensity of some event, i.e.,

$$\mathrm{P}(A) = \lim_{n \to \infty} \frac{n_A}{n},$$

  where n_A is the number of times A happens in n trials (usually it is assumed that $n \to \infty$).
- **Bayesian interpretation:**
  Probabilities as quantification of a belief or the uncertainty on unobserved quantities.

Outline

Discrete probability

- A random variable is used to represent the outcome of an experiment. When the number of possible outcomes is countable, then we encounter a **discrete random variable.**
- The set of all possible outcomes is called the **sample space**: $\Omega = \{\omega_1, \ldots, \omega_n\}$ (e.g., in tossing a coin experiment, $\Omega = \{H, T\}$).
- **Elementary event** is a singleton $\{\omega_r\}$ of Ω, i.e., is an event which cannot be further divided into other events.
- **The set of all possible events** is the power set 2^Ω (for the coin: $\{\emptyset, \{H\}, \{T\}, \{H, T\}\}$).
- The **probability function** P maps events $A \in 2^\Omega$ into the probability of such an event, i.e., $P : 2^\Omega \rightarrow [0, 1]$, such that
  - $P(\emptyset) = 0$ and $P(\Omega) = 1$,
  - $\sum_{\omega_i \in \Omega} P(\{\omega_i\}) = 1$,
  - $A \in 2^\Omega \implies P(A) = \sum_{\omega_i \in A} P(\{\omega_i\})$.
- **Additive rule of probabilities:**
  Let $A, B \in 2^\Omega$, then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Example: Binomial distribution I

- An experiment with two possible outcomes $Y \in \{0, 1\}$ is called **Bernoulli trial** (or binomial trial) and is defined by the "success" probability $p = \mathrm{P}(Y = 1)$.
- The **binomial distribution** models n repeated Bernoulli trials where the outcomes are independent (e.g., in a coin toss experiment) and the random variable X accounts for the number of times we observe "success" $Y = 1$ (the order does not matter), i.e.,

$$P\Big(X = k\Big) = \binom{n}{k} p^k (1 - p)^{n-k},$$

with the binomial coefficient $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

- The sample space is thus $\Omega = \{0, 1, \ldots, n\}$ and

$$\mathrm{P}(\Omega) = \sum_{k=0}^{n} \mathrm{P}\Big(X = k\Big) = \sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k} = (1 - p + p)^n = 1.$$

Example: Binomial distribution II

- Coin toss: $\Omega = \{H, T\}$, $\mathrm{P}(H) = p$. Define $Y : \{H, T\} \to \{0, 1\}$ by

$$Y = \begin{cases} 1 \text{ if } H, \\ 0 \text{ if } T. \end{cases}$$

  Y is a random variable with Bernoulli-distribution:

$$\mathrm{P}_Y(Y = 1) = \mathrm{P}(H) = p, \text{ and similarly } \mathrm{P}_Y(Y = 0) = 1 - p.$$

- Repeat the coin toss independently n times and denote by X the number of times we observe head. Let Ω be the set of all sequences of n variables with the alphabet $\{H, T\}$, then $|\Omega| = 2^n$. X is a random variable $X : \Omega \to \mathbb{Z}$ with distribution

$$\mathrm{P}_X\Big(X = k\Big) = \mathrm{P}(X^{-1}(k)) = \binom{n}{k} p^k (1-p)^{n-k}.$$

  If $n = 3$, then $X^{-1}(2) = \{HHT, HTH, THH\}$.

The Rules of Probability

- There are two fundamental rules of probability theory:

$$\textbf{Sum rule:} \quad P(X) = \sum_Y P(X, Y) \tag{1}$$

$$\textbf{Product rule:} \quad P(X, Y) = P(Y \mid X)P(X)(= P(X \cap Y)) \tag{2}$$

- Let X, Y be discrete random variables. X and Y are **independent** if,

$$\mathrm{P}_{X \times Y}(X = i, Y = j) = \mathrm{P}_X(i)\,\mathrm{P}_Y(j), \quad \forall i, j \in \mathbb{Z}.$$

- The **conditional probability** $\mathrm{P}(X = i | Y = j)$ of X given $Y = j$ is,

$$\mathrm{P}(X = i | Y = j) = \frac{\mathrm{P}_{X \times Y}(X = i, Y = j)}{\mathrm{P}(Y = j)}, \quad \forall j \text{ with } \mathrm{P}(Y = j) > 0.$$

Example: Oranges v.s Apples from Bishop

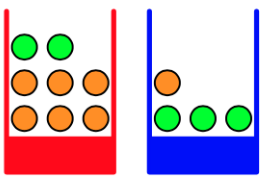


Figure 1: Figure 1.9 from Bishop

$$P(B=r)=4/10$$
$$P(B=b)=6/10$$
$$P(F=a|B=r)=1/4$$
$$P(F=o|B=r)=3/4$$
$$P(F=a|B=b)=3/4$$
$$P(F=o|B=b)=1/4$$

$$\begin{aligned}P(F=a) &= P(F=a|B=r)P(B=r)+P(F=a|B=b)P(B=b)\\ &= \frac{1}{4}\times\frac{4}{10}+\frac{3}{4}\times\frac{6}{10}=\frac{11}{20}\end{aligned}$$

$$P(B=r|F=o)=\frac{P(F=o|B=r)P(B=r)}{P(F=o)}=\frac{3/4\times 4/10}{9/20}=\frac{2}{3}$$

Outline

σ-algebra

- So far, random variables taking discrete values $X \in \{1, 2, 3, \ldots\}$, thus Ω is a countable set.
- What if we consider continuous variables, e.g., $X \in \mathbb{R}$, and thus $\Omega = \mathbb{R}$ is uncountable? How do we assign probabilities to all 2^Ω events?
- If all numbers are equally likely to occur, how do we ensure that $\sum_{\omega_i \in \Omega} P(\omega_i) = 1$?

Definition (σ-algebra)

A set $\mathcal{A} \subset 2^\Omega$ is called a σ-**algebra**:

1. If $\emptyset \in \mathcal{A}$ and $\Omega \in \mathcal{A}$,
2. If $A \in \mathcal{A}$, then also the complement A^c is contained in $\mathcal{A}$,
3. If $\mathcal{A}$ is closed under **countable** unions, that is if $A_1, A_2, \ldots$ is a sequence of events in $\mathcal{A}$, then $\cup_{i=1}^\infty A_i \in \mathcal{A}$.

Probability measure

Definition (Probability measure)

A **probability measure** defined on a σ-algebra $\mathcal{A}$ of Ω is a function $\mathrm{P} : \mathcal{A} \to [0, 1]$ that satisfies:

1. $\mathrm{P}(\Omega) = 1$,
2. For every countable sequence $(A_n)_{n \geq 1}$ of elements of $\mathcal{A}$, pairwise disjoint (that is $A_m \cap A_n = \emptyset$ whenever $m \neq n$), one has

$$\mathrm{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathrm{P}(A_n).$$

- Any discrete **probability space** $(\Omega, 2^{\Omega}, P)$ is a probability measure, since 2^{Ω} is a σ-algebra and P is a probability measure.

Borel σ-Algebra

Let $C \subset 2^{\Omega}$. The σ-algebra generated by C is the smallest σ-algebra containing C.

Definition (Borel σ-algebra)

The **Borel σ-algebra** $\mathcal{B}$ in $\mathbb{R}^d$ is the σ-algebra generated by the open sets in $\mathbb{R}^d$.

Lebesgue Measure on $\mathbb{R}^d$

- The Lebesgue measure $\mu : \mathcal{B} \to \mathbb{R}_+$ is now just the usual measure of volume. For the one-dimensional case, we have
$$\mu(]a, b[) = b - a,$$
- A set $A \in \mathcal{B}$ has **measure zero** if $\mu(A) = 0$. Any countable set of points has Lebesgue measure zero.

Warning: The Lebesgue measure works on its own (larger) σ-algebra but the difference is for our purposes negligible.

Probability on continuous spaces

In the case $\Omega = \mathbb{R}^d$ we will work with measures which have a density with respect to the **Lebesgue measure**.

Let $\mathcal{B}$ be the Borel σ-algebra in $\mathbb{R}^d$. A probability measure P on $(\mathbb{R}^d, \mathcal{B})$ has a **density** p if p is a non-negative (Borel measurable) function on $\mathbb{R}^d$ satisfying for all $A \in \mathcal{B}$ that:

$$\mathrm{P}(A) = \int_A p(x)dx = \int_A p(x_1, \ldots, x_d)\, dx_1 \ldots dx_d,$$

where $dx = dx_1 \ldots dx_d$.

- This implies: $\mathrm{P}(\mathbb{R}^d) = \int_{\mathbb{R}^d} p(x)dx = 1$.
- Observation: Not all probability measures on $\mathbb{R}^d$ have a density.

Example of a probability measure with density

The **Gaussian distribution** or normal distribution on $\mathbb{R}$ has two parameters μ (mean) and σ^2 (variance). The associated density function is denoted by $\mathcal{N}(\mu, \sigma^2)$ and defined as:

$$p(X = x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

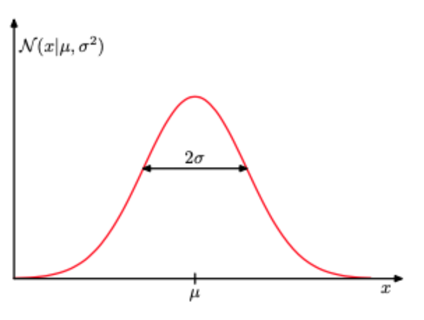


Figure 2: Figure 1.13 from Bishop

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}\left(x \mid \mu, \sigma^2\right) x \, \mathrm{d}x = \mu$$

$$\mathbb{E}\left[x^2\right] = \int_{-\infty}^{\infty} \mathcal{N}\left(x \mid \mu, \sigma^2\right) x^2 \, \mathrm{d}x = \mu^2 + \sigma^2$$

$$\mathrm{var}[x] = \mathbb{E}\left[x^2\right] - \mathbb{E}[x]^2 = \sigma^2$$

Other densities

- **Multivariate Gaussian** $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ is uniquely determined by the mean $\boldsymbol{\mu} \in \mathbb{R}^d$ and the covariance matrix $\Sigma \in \mathbb{R}^{d \times d}$ (positive-definite) as

$$p(\mathrm{x}) = \frac{1}{(2\pi)^{\frac{d}{2}} |\det \Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathrm{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathrm{x}-\boldsymbol{\mu})}$$

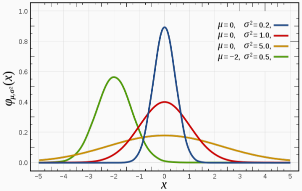
- **Laplace distribution Laplace**(μ, b) is given by

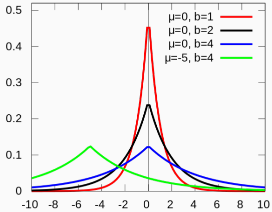
$$p(x) = \frac{1}{2b} e^{-\frac{1}{b}|x-\mu|}$$

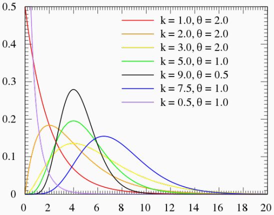
- **Gamma distribution** $\Gamma(\alpha, \beta)$ given by:

$$p(x) = \frac{x^{\alpha-1} \beta^{\alpha} e^{-\beta x}}{\Gamma(\alpha)}, \text{ where } \Gamma(\cdot) \text{ is the Gamma function.}$$

Other densities







Cumulative distribution function

- The **(cumulative) distribution function** of a probability measure P on $(\mathbb{R}, \mathcal{B})$ is the function

$$F(x) = \mathrm{P}\Big(X \in (-\infty, x]\Big) = \mathrm{P}\Big(X \leq x\Big) = \int_{-\infty}^{x} p(t)dt.$$

If the distribution function F is sufficiently differentiable, then

$$p(x) = \frac{\partial F}{\partial x}\Big|_x.$$

- The distribution function of P on $(\mathbb{R}^d, \mathcal{B})$ is the function

$$F(x_1, \ldots, x_d) = \mathrm{P}\Big(X_1 \leq x_1, \ldots, X_d \leq x_d\Big).$$

If the distribution function F is sufficiently differentiable, then

$$p(x_1, \ldots, x_d) = \frac{\partial^d F}{\partial x_1 \ldots \partial x_d}\Big|_{x_1, \ldots, x_d}.$$

Quantile

Quantiles: Quantiles are only defined for distributions on $\mathbb{Z}$ and $\mathbb{R}$.

Definition

The α-**quantile** of a probability measure on $\mathbb{Z}$ or $\mathbb{R}$ is the real number q_α such that

$$F(q_\alpha) = \mathrm{P}(]-\infty, q_\alpha]) = \alpha.$$

The **median** is the $\frac{1}{2}$-quantile.

- Median and mean agree if the distributions are symmetric (and unimodal).
- The median is more robust to changes of the probability measure.

Cumulative distribution and Quantiles

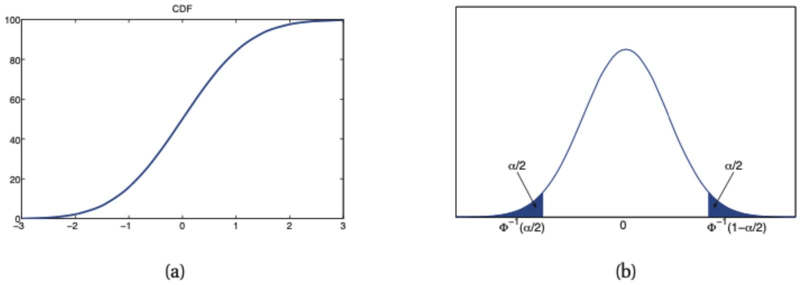


Figure 2.3 (a) Plot of the cdf for the standard normal, $\mathcal{N}(0,1)$. (b) Corresponding pdf. The shaded regions each contain $\alpha/2$ of the probability mass. Therefore the nonshaded region contains $1-\alpha$ of the probability mass. If the distribution is Gaussian $\mathcal{N}(0,1)$, then the leftmost cutoff point is $\Phi^{-1}(\alpha/2)$, where Φ is the cdf of the Gaussian. By symmetry, the rightost cutoff point is $\Phi^{-1}(1-\alpha/2) = -\Phi^{-1}(\alpha/2)$. If $\alpha = 0.05$, the central interval is 95%, and the left cutoff is -1.96 and the right is 1.96. Figure generated by `quantileDemo`.

Joint density and marginals

Let $X = (X_1, X_2)$ be a $\mathbb{R}^2$-valued random variable with density p_X on $\mathbb{R}^2$. Then the densities p_{X_1} of X_1 and p_{X_2} of X_2 are given as

$$p_{X_1}(x_1) = \int_{\mathbb{R}} p_X(x_1, x_2) dx_2, \qquad p_{X_2}(x_2) = \int_{\mathbb{R}} p_X(x_1, x_2) dx_1.$$

- $p_X(x_1, x_2)$ denotes the **joint density**.
- p_{X_1} and p_{X_2} are called **marginal densities** of X and are associated to the probability measures of X_1 respectively X_2.

Observation: The joint measure can in general not be reconstructed from the knowledge of the marginal densities (only if X_1 and X_2 are independent).

Independence and conditional density

Let X, Y be $\mathbb{R}$-valued random variables with joint-density $p_{X \times Y}$ and marginal densities p_X and p_Y, then X and Y are **independent** if

$$p_{X \times Y}(x, y) = p_X(x)\ p_Y(y), \quad \forall x, y \in \mathbb{R}.$$

The **conditional density** $p(x|Y = y)$ of X given $Y = y$ is defined as,

$$p(x|y) = \frac{p(x, y)}{p(y)}, \quad \forall\, y \text{ with } p(y) > 0.$$

Example: Joint, marginals and conditionals

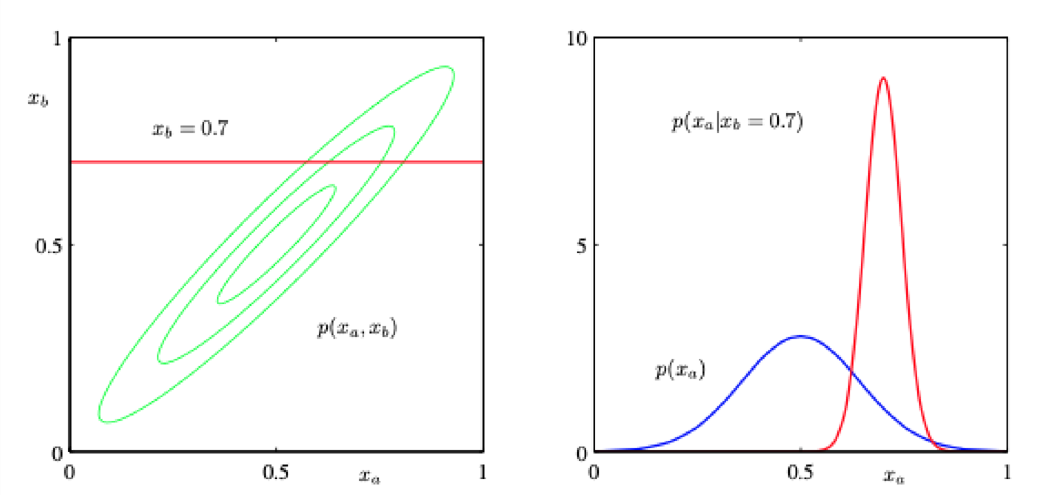


Figure 2.9 from Bishop

Transformation of Random Variables

Theorem

Let $X = (X_1, \ldots, X_d)$ *have joint density* p_X*. Let* $g : \mathbb{R}^d \to \mathbb{R}^d$ *be continuously differentiable and injective, with non-vanishing Jacobian. Then* $Y = g(X)$ *has density*

$$p_Y(y) = \quad p_X(g^{-1}(y))|\det J_{g^{-1}}(y)|$$

- The Jacobian $J_g(x)$ of a function $g : \mathbb{R}^d \to \mathbb{R}^d$ at value x is the $d \times d$- matrix

$$J_g(x)_{ij} = \frac{\partial g_i}{\partial x_j}\Big|_x, \quad i, j = 1, \ldots, d$$

- This result allows us to generate samples from complicated densities from simple ones.

Example: Sampling from an exponential distribution

$$p_\lambda(y) = \lambda \, \mathsf{exp}(-\lambda \, y), \text{ for } y \geq 0.$$

1. We can first sample from a uniform distribution on $[0, 1]$.
2. Apply a function $g : [0, 1] \to \mathbb{R}_+$ (resp. g^{-1}) such that

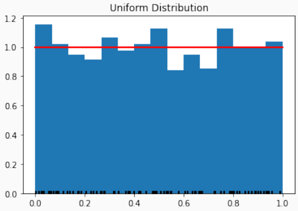
$$p_\lambda(y) = \lambda \, \mathsf{exp}(-\lambda \, y) = p_X(g^{-1}(y)) \left| \frac{\partial g^{-1}}{\partial y} \right| = \left| \frac{\partial g^{-1}}{\partial y} \right|.$$

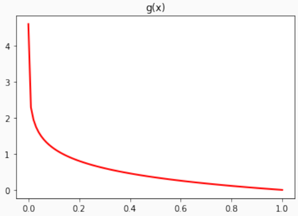
General case: complicated differential equation.

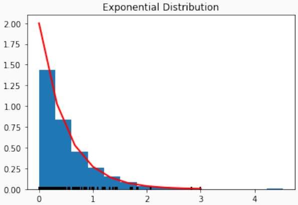
This case: $g^{-1}(y) = \mathsf{exp}(-\lambda \, y) \implies g(x) = -\frac{\log(x)}{\lambda}$

- X_i samples from the uniform distribution on $[0, 1]$,
- $Y_i = g(X_i) = -\frac{\log(X_i)}{\lambda}$ are samples from the exponential distribution.

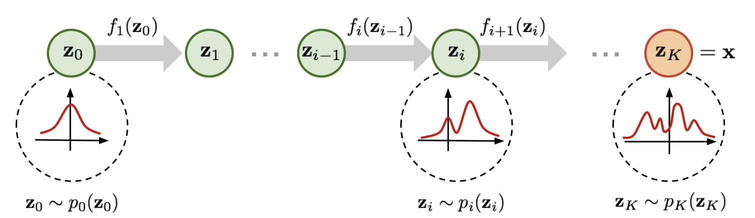
Example: Sampling from an exponential distribution

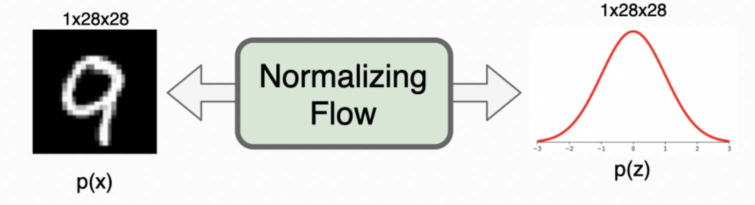






Outlook: State of the Art Normalizing Flow Models





e.g. Kingma'18, Flow++, GLOW

Outline

Expectation

The **expected value** or **expectation** of a $\mathbb{R}^d$-valued random variable X is defined as

$$(\mathbb{E}[X])_i = \int_{\mathbb{R}^d} x_i \; p(x) \, dx \; = \int_{\mathbb{R}^d} x_i \; p(x_1, \ldots, x_d) \, dx_1 \ldots dx_d \,,$$

and for a discrete random variable X taking values in $\mathbb{Z}$ it is defined as,

$$\mathbb{E}[X] = \sum_{n=-\infty}^{\infty} n \, \mathrm{P}(X = n).$$

Expectation of functions of random variables
We can also define the expectation of functions of random variables.

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}^d} f(x) \, p(x) \, dx \; = \int_{\mathbb{R}^d} f(x_1, \ldots, x_d) \, p(x_1, \ldots, x_d) \, dx_1 \ldots, dx_d \,.$$

Variance, Covariance and Correlation

The **variance** $\mathrm{Var}[X]$ (also $\sigma^2(X)$) of an $\mathbb{Z}$- or $\mathbb{R}$-valued random variable X is defined as

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

The standard deviation of X is $\sigma(X) = \sqrt{\mathrm{Var}[X]}$.
The covariance matrix Σ of an $\mathbb{R}^d$-valued random variable X is given as $\Sigma_{ij} = \mathrm{Cov}(X_i, X_j)$ or in matrix form

$$\Sigma = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T].$$

The **covariance** $\mathrm{Cov}(X, Y)$ of two $\mathbb{R}$-valued random variables X and Y is defined as,

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = \mathbb{E}[X\,Y] - \mathbb{E}[X]\,\mathbb{E}[Y].$$

The **correlation** $\mathrm{Corr}(X, Y)$ of two $\mathbb{R}$-valued random variables X and Y is then defined as,

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Cov}(X, X)\,\mathrm{Cov}(Y, Y)}} = \frac{\mathrm{Cov}(X, Y)}{\sigma(X)\sigma(Y)}.$$

Properties

- The expectation and variance have the following properties $\forall a, b \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E}[aX+b] &= a\,\mathbb{E}[X]+b, \qquad \mathbb{E}[X+Y] = \mathbb{E}[X]+\mathbb{E}[Y], \\ \mathrm{Var}[aX+b] &= a^2\ \mathrm{Var}[X], \\ \mathrm{Var}[X+Y] &= \mathrm{Var}[X]+\mathrm{Var}[Y]+2\ \mathrm{Cov}(X,Y). \end{aligned}$$

- Correlation is a measure of **linear dependence**, and satisfies $-1 \leq \mathrm{Corr}(X,Y) \leq 1$. If X and Y are linearly dependent, that is $Y = aX + b$ with $a, b \in \mathbb{R}$, then

$$\mathrm{Corr}(X,Y) = \mathrm{Corr}(X, aX+b) = \frac{a}{|a|} = \begin{cases} 1, & \text{if } a > 0, \\ 0, & \text{if } a = 0, \\ -1, & \text{if } a < 0. \end{cases}$$

In words, linearly dependent random variables achieve maximal correlation.

Conditional expectation

Let X, Y be two $\mathbb{R}$-valued random variables. The **conditional expectation** $\mathbb{E}[X|Y=y]$ of X given $Y=y$ is defined for y with $p(y)>0$ as the quantity

$$\mathbb{E}[X|Y=y] = \int_{\mathbb{R}} x\, p(x|y)\, dx.$$

The **conditional expectation** $\mathbb{E}[X|Y]$ of X given Y is a random variable $h(Y)$ with values

$$h(y) = \mathbb{E}[X|Y=y].$$

Important properties of the conditional expectation are:

- $\mathbb{E}[X|Y] = \mathbb{E}[X]$, if X and Y are **independent**,
- $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$,
- $\mathbb{E}[f(Y)|Y] = f(Y)$ and $\mathbb{E}[f(Y)g(X)|Y] = f(Y)\mathbb{E}[g(X)|Y]$.

Outline

Law of total probability

Assume that we have a finite or countably infinite number of events $\mathcal{A} = \{A_1, A_2, A_3, \ldots\}$ and $\Omega = A_1 \cup A_2 \cup A_3 \cup \ldots$.

Definition

A collection of events $(A_n)_{n \geq 1}$ is called a **partition** of Ω if $A_n \in \mathcal{A}$ for each n, they are pairwise disjoint, $A_n \cap A_m = \emptyset$ for $m \neq n$, $\mathrm{P}(A_n) > 0$ for each n, and $\cup_n A_n = \Omega$.

Theorem (Law of total probability)

Let $(A_n)_{n \geq 1}$ be a finite or countable partition of Ω. Then if $B \in \mathcal{A}$,

$$\mathrm{P}(B) = \sum_n \mathrm{P}(B|A_n)\mathrm{P}(A_n).$$

Bayes' theorem

Theorem (Bayes' theorem)

Let A, B be two events and $P(B) > 0$, then

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

- The above definition follows from the definition of conditional probability.
- Implication: Let $(A_n)_{n\geq 1}$ be a finite or countable partition of Ω, and suppose $\mathrm{P}(B) > 0$. Then

$$\mathrm{P}(A_i|B) = \frac{\mathrm{P}(B|A_i)\mathrm{P}(A_i)}{\sum_n \mathrm{P}(B|A_n)\mathrm{P}(A_n)}.$$

Bayes' theorem

$$\mathrm{P}(A, B) = \mathrm{P}(A, B)$$

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

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

Bayes' theorem

What is the probability of having Covid given that a test is positive? $\mathrm{P}(\text{covid}|\text{positive})$

- Sensitivity, e.g. 96.5% $\rightarrow \mathrm{P}(\text{positive}|\text{covid}) = 0.965$
  False-negative-rate $1-$ sensitivity $= 3.5\%$
- Specificity, e.g. 99.7% $\rightarrow \mathrm{P}(\text{negative}|\text{healthy}) = 0.997$
  False-positive-rate is $1-$ specificity $= 0.3\%$
- Prevalence, e.g. 0.25% $\rightarrow \mathrm{P}(\text{covid}) = 0.0025$ (Note: we pick a **random** person here!)

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

$$\mathrm{P}(\text{covid}|\text{positive}) = \frac{\mathrm{P}(\text{positive}|\text{covid})\mathrm{P}(\text{covid})}{\mathrm{P}(\text{positive})}$$

$$\mathrm{P}(\text{covid}|\text{positive}) = \frac{\mathrm{P}(\text{positive}|\text{covid})\mathrm{P}(\text{covid})}{\mathrm{P}(\text{positive}|\text{covid})\mathrm{P}(\text{covid}) + \mathrm{P}(\text{positive}|\text{healthy})\mathrm{P}(\text{healthy})}$$

$$\mathrm{P}(\text{covid}|\text{positive}) = \frac{0.965 * 0.0025}{0.965 * 0.0025 + 0.003 * 0.9975} = 0.446$$

Lecture 2: Recap of Probability Theory

Prof. Dr. Mario Fritz

2022 04 20

https://fritz.cispa.saarland
Trustworthy AI
CISPA Helmholtz Center for Information Security, Saarbrücken, Germany