Statistical Methods and Learning in Computer Vision

Claudia Nieuwenhuis

Organization

- Lectures every week, Thursday 11.15am
- Exercises every other week, Tuesday or Wednesday 10.15am
- Oral exam at the end of the lecture
- Topics:
    - necessary basics in measure theory and statistics
    - density estimation and sampling methods
    - subspace methods (PCA, ICA, LDA)
    - learning and classification approaches (SVM, NN)
    - optimization (MRF, PDE)

Chapter 1
Statistical Basics
Measure and Test Theory

Statistical Methods and Learning in Computer Vision
SS 2011

Claudia Nieuwenhuis
Lehrstuhl für Computer Vision and Pattern Recognition
Fakultät für Informatik
Technische Universität München

Overview

1. Statistical Basiscs
2. Probability Spaces and Measures
3. Random Variables
4. Cumulative Distribution Functions
5. Density Functions
6. Moments
7. Test Theory
8. Parameter Estimation
9. Test Experiments

Overview

Randomized Experiment

Randomized Experiment

A randomized experiment is a process with unknown result, which can be arbitrarily often repeated.

Example: tossing two dice.

State Space

The state space Ω is the set of all possible outcomes of a randomized experiment.

Example: $\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$.

Event

An event is a property which can be observed either to hold or not to hold after the experiment is done. It is a subset of Ω.

Example: $A = \{(i, j) \in \Omega : i + j < 9\}$

Discrete Probability Space

A triple $(\Omega, \mathcal{A}, P)$ is called a discrete probability space if

- the state space Ω is not empty and countable.
- $\mathcal{A}$ is the power set $\mathcal{P}(\Omega)$.
- $P : \mathcal{A} \rightarrow \mathbb{R}$ is a mapping with the following properties
  - $P(A) \geq 0$ $(A \in \mathcal{A})$
  - $P(\Omega) = 1$
  - for each sequence of pairwise distinct sets from $\mathcal{A}$ the σ-additivity holds: $P(\sum_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} P(A_n)$

P is called a probability measure.

Dice Example:

- $\Omega = \{(i,j) : 1 \leq i,j \leq 6\}$
- $\mathcal{A} = \mathcal{P}(\Omega) = \{\{(1,1)\}, \{(1,1)(1,2)\}, \{(1,1)(1,2)(1,3)\}, ...\}$
- $P(A) = \frac{\text{number of elements in A}}{36}$

Probability Space

σ-Algebra

A σ-algebra $\mathcal{A}$ is a system of subsets of Ω if

- $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$, where $A^c = \Omega - A = \{x \in \Omega : x \notin A\}$ means the complement of A.
- For each sequence (A_n) of sets from $\mathcal{A}$ $\quad \cup_{n\in\mathbb{N}} A_n$ lies in $\mathcal{A}$.

The elements of $\mathcal{A}$ are called events or measurable sets.

A σ-algebra is closed under finite set operations

- $A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$
- $A, B \in \mathcal{A} \Rightarrow A - B \in \mathcal{A}$
- $\emptyset \in \mathcal{A}, \Omega \in \mathcal{A}$

Borel σ algebra

The Borel σ algebra $\mathcal{B}^n$ consists of all finite unions of I^n, which is the set of right half-open intervals in $\mathbb{R}^n$.
$\mathcal{B}^n = \{\sum_{i=1}^{k} A_i | k \in \mathbb{N}, A_i \in I^n\}$.

$\mathcal{B}^n$ is very often used in probability theory, since it contains almost any element in $\mathcal{P}(\mathbb{R}^n)$.

Probability Space

Probability Space

A triple $(\Omega, \mathcal{A}, P)$ is called a probability space if

- the state space Ω is not empty.
- $\mathcal{A}$ is a σ-algebra over Ω.
- $P : \mathcal{A} \to \mathbb{R}$ is a mapping with the following properties
  1. $P(A) \geq 0$ $(A \in \mathcal{A})$ (non-negative)
  2. $P(\Omega) = 1$ (normed)
  3. for each sequence of pairwise disjoint sets from $\mathcal{A}$ the σ-additivity holds: $P(\sum_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} P(A_n)$

P is a probability measure over the σ algebra $\mathcal{A}$.

In the definition of the discrete probability space the σ algebra was specified as the power set over Ω.

If instead of condition 2, $P(\emptyset) = 0$ holds and $\mathcal{A}$ is a system of subsets over Ω, then P is called a measure, $(\Omega, \mathcal{A})$ is called a measurable space and the triple $(\Omega, \mathcal{A}, P)$ a measure space.

Every probability measure is a measure.

Example of a Probability Measure

Point Probability Measure

Let $(\Omega, \mathcal{A})$ be a state space Ω with σ-algebra $\mathcal{A}$ and $\omega \in \Omega$.
Then $\mu_\omega : \mathcal{A} \to \{0, 1\}$,

$$\mu_\omega(A) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

defines a probability measure.

Proof:

- $\mu_\omega(A) \geq 0$
- $\mu_\omega(\Omega) = 1$
- Let $A_1, .., A_k, ...$ be pairwise disjoint sets in $\mathcal{A}$, then

$$\mu_\omega(\sum_{i=1}^{\infty} A_i) = \begin{cases} 1, & \exists i : \omega \in A_i \\ 0, & \text{otherwise} \end{cases} = \sum_{i=1}^{\infty} \mu_\omega(A_i)$$

Summary

- Probability Space $(\Omega, \mathcal{A}, P)$, $\mathcal{A}$ σ-algebra, P probability measure.
- Probability measure
  - $P(A) \geq 0$ $(A \in \mathcal{A})$
  - $P(\Omega) = 1$
  - for each sequence of pairwise distinct sets from $\mathcal{A}$ the σ-additivity holds: $P(\sum_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} P(A_n)$
- For a measure, condition 2 is replaced by $P(\emptyset) = 0$ and $\mathcal{A}$ does not have to be a σ-algebra.
- Every probability measure is a measure.

Random Variables

We now aim at transferring a probability measure from one σ-algebra to another.

Dice Example:
We have the state space $\Omega = \{(k,l) : 1 \leq k, l \leq 6\}$ with event system $\mathcal{P}(\Omega)$ and the uniform probability measure $P : P\{(k,l)\} = \frac{1}{36}$, which make up the probability space $(\Omega, \mathcal{P}(\Omega), P)$.

We are interested in the sum of the dice defined by the mapping $T : \Omega \rightarrow \Omega'$ with $T(k,l) = k + l$ and $\Omega' = \{2, .., 12\}$.

The mapping leads to the probability space $(\Omega', \mathcal{P}(\Omega'), P')$. We are interested in the probability measure P'.
For $A' \in \mathcal{P}(\Omega')$, $P'(A')$ is understood as the probability of $P(\{(k,l) : T(k,l) \in A'\}) \in P(\Omega)$.

For example: $P'(\{11, 12\}) = P\{(5,6), (6,5), (6,6)\} = P\{(5,6)\} + P\{(6,5)\} + P\{(6,6)\} = \frac{3}{36}$.

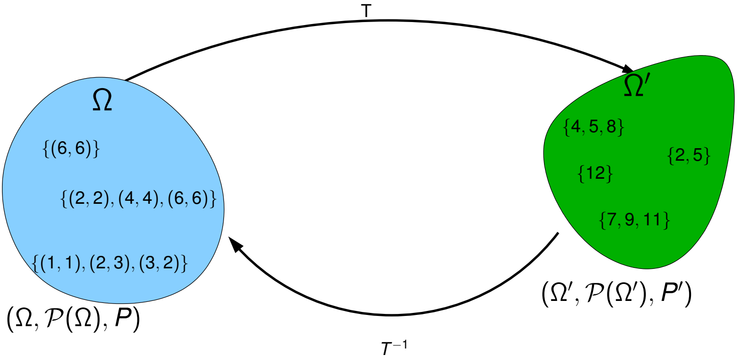
Random Variables

Preimage Mapping

Let $T : \Omega \to \Omega'$ be an arbitrary mapping. Then the preimage mapping $T^{-1} : \mathcal{P}(\Omega') \to \mathcal{P}(\Omega)$ is defined by

$$T^{-1}(A') = \{\omega \in \Omega : T(\omega) \in A'\}, A' \in \mathcal{P}(\Omega')$$

The preimage of a σ-algebra is a σ-algebra.



Random Variable

Let $(\Omega, \mathcal{A})$ and $(\Omega', \mathcal{A}')$ be measurable spaces. A mapping $X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ is called $\mathcal{A} - \mathcal{A}'$ measurable if

$$X^{-1}(A') = \{\omega \in \Omega : X(\omega) \in A'\} \in \mathcal{A}$$

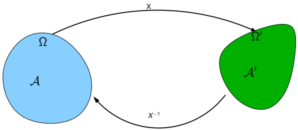
Such a measurable mapping X is called a random variable.

Image Measure

Let $X : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$ be a random variable and P a measure over $\mathcal{A}$. Then

$$P'(A') := P_X(A') := P(X^{-1}(A')), \; A' \in \mathcal{A}$$

defines a measure over $\mathcal{A}'$. Is P a probability measure then P_X is a probability measure over $\mathcal{A}'$. P_X is called the image measure of P by X.



Example

Dice Example:
State spaces $\Omega = \{(k,l) : 1 \leq k,l \leq 6\}$ and $\Omega' = \{2,..,12\}$ with power sets as event system and the uniform probability measure on $(\Omega, \mathcal{P}(\Omega))$.

Define the mapping $X : (\Omega, \mathcal{P}(\Omega)) \to (\Omega', \mathcal{P}(\Omega'))$, $X((k,l)) = k + l$. Is X a random variable?
It holds that $X^{-1}(A') = \{\omega \in \Omega : X(\omega) \in A'\} \in \mathcal{P}(\Omega)$ since for every value in Ω' we can find two dice results which sum up to this value. It follows that X is $\mathcal{P}(\Omega) - \mathcal{P}(\Omega')$-measurable and, thus, a random variable.
Furthermore, P is a probability measure. Hence, the image measure P_X under X which is defined by $P_X(A') = P(X^{-1}(A'))$ is a probability measure on $(\Omega', \mathcal{P}(\Omega'))$.

For example
$P_X\{2,4,5\} = P(X^{-1}\{2,4,5\}) = P\{(k,l) : X(k,l) \in \{2,4,5\}\} = P\{(1,1),(2,2),(1,3),(3,1),(2,3),(3,2)\} = \frac{6}{36} = \frac{1}{6}$.

Probability Distribution

The notion 'random variable' is just a name. It is neither a variable nor is it random, but a measurable mapping. By means of a random variable measures can be transferred from one σ-algebra to another.

Probability Distribution

Let $X : (\Omega, \mathcal{A}, P) \rightarrow (\Omega', \mathcal{A}')$ be a random variable. Then the image measure P_X of P by X is called probability distribution.

Every probability measure can be understood as a distribution, since there is always a random variable having P as its image measure: the identical mapping $Id : \Omega \rightarrow \Omega$. Hence, the notions probability distribution and probability measure are often used equivalently.

Summary

- A random variable is a measurable mapping, which transfers probability measures from one space to another
- The image measure of a random variable is called probability distribution
- Every probability measure is a probability distribution by means of the random variable defined by the identity mapping.

Cumulative Distribution Function

Let P be a probability measure on the Borel σ-algebra $\mathcal{B}$. Then $F_P : \mathbb{R} \to \mathbb{R}$ is called cumulative distribution function of P if

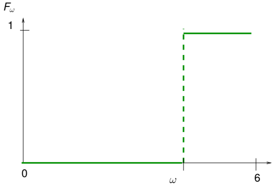
$$F_P(x) = P((-\infty, x)), x \in \mathbb{R}$$

Example:

Point Measure on $\mathcal{B}$

$$\mu_\omega(A) = \begin{cases} 0, & \omega \notin A \\ 1, & \omega \in A \end{cases}$$

Cumulative Distribution Fct.



$$F_\omega(x) = \begin{cases} 0, & x \leq \omega \\ 1, & x > \omega \end{cases}$$

Note: a probability distribution is a probability measure defined on a specific system of sets (a σ-algebra). The cumulative distribution function is defined for points in $\mathbb{R}$.

Cumulative Distribution Function

$$F_P(x) = P((-\infty, x)), x \in \mathbb{R}$$

Properties of Cumulative Distribution Functions

The cumulative distribution function F_P of a probability measure P over $\mathcal{B}$ has the following properties

- F_P is monotonously increasing
- F_P is left-continuous
- $\lim_{x \to -\infty} F_P(x) = 0$
- $\lim_{x \to \infty} F_P(x) = 1$

Each function with these properties uniquely describes a probability measure.
Each probability measure is defined uniquely by its distribution function.
$P([a, b)) = P(-\infty, b) - P(-\infty, a) = F_P(b) - F_P(a)$

Summary

- The cumulative distribution function is defined as: $F_P(x) = P((-\infty, x)), x \in \mathbb{R}$
- The probability of an element of the Borel σ-algebra $\mathcal{B}$ can be determined by $P([a, b)) = F_P(b) - F_P(a)$
- A probability measure and its cumulative distribution function uniquely determine each other.

Density Function

Let $F_P : \mathbb{R} \to \mathbb{R}$ be the cumulative distribution function of a probability measure P over $\mathcal{B}$. A measurable function $f : \mathbb{R} \to \mathbb{R}_+ \cup \{\infty\}$ is called a density function if and only if

$$F_P(t) = \int_{-\infty}^{t} f(x)\,\mathrm{d}x,\ \ t \in \mathbb{R}$$

Densities are not unique. However, different densities of the same probability measure differ only over null sets.

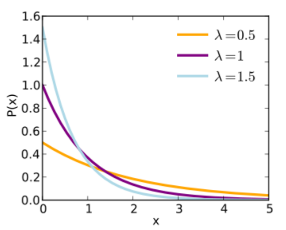
Relationship between probability measures, cumulative distribution functions and density functions:

$$\begin{aligned} P([a,b)) &= P((-\infty,b)) - P((-\infty,a)) \\ &= F_P(b) - F_P(a) \\ &= \int_{-\infty}^{b} f_P(x)\,\mathrm{d}x - \int_{-\infty}^{a} f_P(x)\,\mathrm{d}x \\ &= \int_{a}^{b} f_P(x)\,\mathrm{d}x \end{aligned}$$

Exponential Distribution

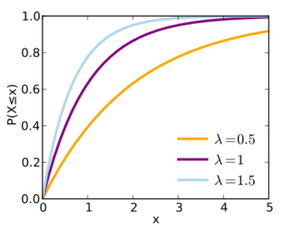
Density Function

$$f_\lambda(x) = \begin{cases} \lambda \exp^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$



Cumulative Distribution Function

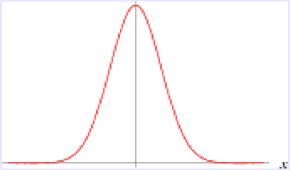
$$F_\lambda(x) = \int_{-\infty}^{x} f_\lambda(t)\,\mathrm{d}t = 1 - \exp^{-\lambda x}$$



Normal Distribution

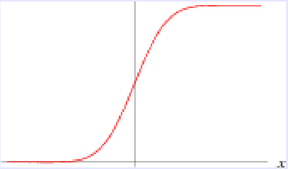
Density Function

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp^{-\frac{x^2}{2}}, \ x \in \mathbb{R}$$



Cumulative Distribution Function

$$F(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp^{-\frac{x^2}{2}} \, \mathrm{d}x, \ t \in \mathbb{R}$$



Density of the Image Measure

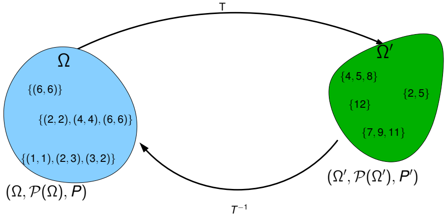
We now want to transfer the density of a probability measure P from one σ-algebra to another.
Let P be a probability measure over $\mathcal{B}^n$ with density f and $T : (\mathbb{R}^n, \mathcal{B}^n) \to (\mathbb{R}^n, \mathcal{B}^n)$ a measurable mapping. When does a density of the image measure P_T exist?

We have

$$P_T(B') = P(T^{-1}(B')) = \int_{T^{-1}(B')} f\, dx \; (B' \in \mathcal{B}^n)$$

Does a density $g : \mathbb{R}^n \to \mathbb{R}_+$ of the image measure P_T exist with

$$P(T^{-1}(B')) = \int_{T^{-1}(B')} f\, \mathrm{d}x = \int_{B'} g\, \mathrm{d}x = P_T(B'), \quad \text{for } B' \in \mathcal{B}^n?$$



Density of the Image Measure: Density Transformation Theorem

Let P be a probability measure over $\mathcal{B}^n$ with a density function f, and $T : \mathbb{R}^n \to \mathbb{R}^n$ a measurable mapping for which holds

- T is continuously differentiable.
- The functional determinant $\Delta_T(x) = \frac{\partial T}{\partial x} \neq 0$ for $x \in \mathbb{R}^n$.
- T is injective meaning that the inverse mapping $T^{-1} : T(\mathbb{R}^n) \to \mathbb{R}^n$ exists.

Then T^{-1} is measurable and a density function g of the image measure P_T is given by

$$g(y) = \begin{cases} \frac{f(T^{-1}(y))}{|\Delta_T(T^{-1}(y))|} & y \in T(\mathbb{R}^n) \\ 0 & y \in \mathbb{R}^n - T(\mathbb{R}^n) \end{cases}$$

Density Transformation Example

Let X be a random variable distributed according to the standard normal distribution, i.e. $P_X = \mathcal{N}(0, 1)$, meaning that the density function of X is given by

$$f(y) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{y^2}{2}\}$$

Let $\mu, \sigma \in \mathbb{R}, \sigma > 0$. The mapping

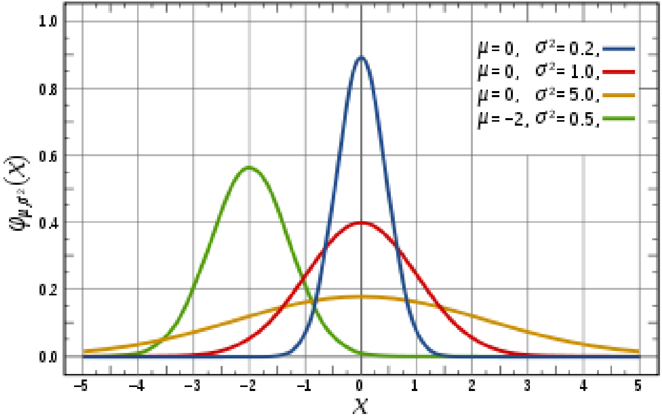
$$T(x) = \sigma x + \mu$$

is measurable. It is continuously differentiable, $\Delta_T(x) = \sigma > 0$ and the inverse transform exists $T^{-1}(y) = \frac{y-\mu}{\sigma}$. Hence, a density of the image measure P_T is given by

$$g(y) = \frac{f(T^{-1}(y))}{|\Delta_T(T^{-1}(y))|} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}$$

g is the density of a probability measure, namely of the normal distribution $\mathcal{N}(\mu, \sigma^2)$.

The Normal Distribution



The smaller σ^2 the steeper rises the density function. It is centered at μ.

90% of the area under curve lie within the interval $\mu \pm 1.6\sigma$.

95% of the area under curve lie within the interval $\mu \pm 2\sigma$.

99% of the area under curve lie within the interval $\mu \pm 3\sigma$.

Summary

- The density f of a probability measure P is defined by $F_P(t) = \int_{-\infty}^{t} f(x)\,\mathrm{d}x,\ t \in \mathbb{R}$
- $P([a,b)) = \int_a^b f(x)\,\mathrm{d}x$
- A density of an image measure P_X under a random variable X is given by the transformation theorem.
- The density of the normal distribution $\mathcal{N}(\mu, \sigma)$ is given as the transformed standard normal density under the random variable $T(x) = \sigma x + a$

Moments

Expectation Value

Let $(\Omega, \mathcal{A}, P)$ be a probability space, f a density of P_X and $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ a random variable. Then the expectation value of the random variable X is given by

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x$$

Examples

Exponential Distribution: $E(X) = \int_{-\infty}^{\infty} x\, \lambda \exp^{-\lambda x}\, \mathrm{d}x = \frac{1}{\lambda}$

Normal Distribution: $E(X) = \int_{-\infty}^{\infty} x\, \frac{1}{\sqrt{2\pi\sigma}} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}\, \mathrm{d}x = \mu$

Moments

Moments

Let $(\Omega, \mathcal{A}, P)$ be a probability space, f a density of P and $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ a random variable. Then the k-th moment of the random variable X is given by

$$E(X^k) := \int_{-\infty}^{\infty} x^k \, f(x) \, \mathrm{d}x.$$

The k-th centralized moment is given by

$$E[(X - E(X))^k] := \int_{-\infty}^{\infty} (x - E(x))^k \, f(x) \, \mathrm{d}x$$

The first moment is the expectation value, the 0-th equals 1.
The second centralized moment is called the variance $V_P(X)$.
$\sqrt{V_P(X)}$ is called the standard deviation.
The k-th (centralized) moment of the probability measure P is understood as that of the identity random variable $Id_{\mathcal{B}}$.

Properties of the Expectation Value

Let $\lambda \in \mathbb{R}, A \in \mathbb{R}^{n \times n}$, then

$$\begin{aligned} E(c) &= c \\ E(\lambda X) &= \lambda E(X) \\ E(X+Y) &= E(X)+E(Y) \\ E(AX) &= AE(X) \end{aligned}$$

Translation Theorem for the Variance

Let $(\Omega, \mathcal{A}, P)$ be a probability space and X a random variable:

$$V(X) = E(X^2) - [E(X)]^2$$

Proof

$$\begin{aligned} V(X) &= E[(X-E(X))^2] \\ &= E[X^2 - 2XE(X) + E(X)^2] \\ &= E(X^2) - 2E(X)E(X) + E(X)^2 \\ &= E(X^2) - E(X)^2 \end{aligned}$$

The variance measures the expected squared deviation of a random variable from its expectation value.

Covariance and Correlation Coefficient

Let $X, Y : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B})$ be two random variables with variance $V(X), V(Y)$. Then

- The covariance of X and Y is defined as

$$Cov(X,Y) := E[(X - E(X))(Y - E(Y))]$$

- The correlation coefficient of X and Y is defined as

$$Cor(X,Y) := \frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

If the covariance is 0, X and Y are called uncorrelated.
The larger the covariance the more vary X and Y together.
The correlation coefficient is a normalized version of the covariance and lies between -1 and 1.

Corollary

- $Cov(X, Y) = E(XY) - E(X)E(Y)$
- $Cov(aX + c, bY + d) = abCov(X, Y)$
- $Cov(X, X) = V(X)$
- $V(aX + c) = a^2 V(X)$
- $V(X + Y) = V(X) + V(Y) + 2Cov(X, Y)$

Uncorrelated Random Variables

For two random variables X and Y the following statements are equivalent

- $Cor(X, Y) = 0$ (X and Y uncorrelated)
- $Cov(X, Y) = 0$
- $E(XY) = E(X)E(Y)$
- $V(X + Y) = V(X) + V(Y)$ (equation by Bienaymé)

Summary

- $E_P(X) = \int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x$
- $V(X) = E(X - E(X))^2 = E(X^2) - E(X)^2$
- $Cov(X, Y) := E[(X - E(X))(Y - E(Y))]$
- $Cor(X, Y) := \frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$
- $Cov(X, Y) = 0 \Leftrightarrow E(XY) = E(X)E(Y) \Leftrightarrow V(X + Y) = V(X) + V(Y)$

Independence

Let $(\Omega, \mathcal{A}, P)$ be a probability space and $(X_i)_{i\in\mathbb{N}}$ a set of random variables $X_i : (\Omega, \mathcal{A}) \rightarrow (\Omega_i, \mathcal{A}_i)$. The random variables X_i are called stochastically independent, if and only if

$$P(\{\omega \in \Omega | X_1(\omega) \in A_1, ..., X_n(\omega) \in A_n\}) = \prod_{i=1}^{n} P(\{\omega \in \Omega | X_i(\omega) \in A_i\}), A_i \in \mathcal{A}_i$$

Independent random variables are uncorrelated, but uncorrelated random variables are not necessarily independent.

Marginal Density

Let $X = (Y, Z)$ be a random variable and f a density of P_x on $\mathbb{R}^2$. Then both P_Y and P_Z have marginal densities on $(\mathbb{R}, \mathcal{B})$:

$$f_Y(y) = \int_{-\infty}^{\infty} f(y,z)\, dz \qquad f_Z(z) = \int_{-\infty}^{\infty} f(y,z)\, \mathrm{d}y$$

Y and Z are independent if and only if $f(y,z) = f_Y(y) f_Z(z)$. In general the joint density f cannot be recovered from the marginal distributions only.

Conditional Density

Let $X = (Y, Z)$ be a random variable with density and f a density of P_X on $\mathbb{R}^2$ with marginal density $f_Y(y) \neq 0$. Then the density

$$f_c(z|Y = y) = \frac{f(y,z)}{f_Y(y)}$$

is called the conditional density of Z given $Y = y$.

Test Theory

Statistical Experiment

Let $(\mathbb{H}, \mathcal{H})$ be a measurable space, the sample space, and $\mathcal{W}$ a set of possible probability measures. The set $(\mathbb{H}, \mathcal{H}, \mathcal{W})$ is called statistical experiment if there is exactly one among the probability measures in $\mathcal{W}$ which manifests itself through its sample realizations $x \in \mathbb{H}$.

Statistic

Let $(\mathbb{H}, \mathcal{H})$ be a sample space and $(\mathbb{D}, \mathcal{D})$ a measurable space. An $\mathcal{H}$-$\mathcal{D}$-measurable mapping $T : \mathbb{H} \to \mathcal{D}$ is called statistic.

Test Problem

Let $(\mathbb{H}, \mathcal{H}, \mathcal{W})$ be a statistical experiment. In test theory, $\mathcal{W}$ is partitioned into two subsets: $\mathcal{W}_1 = \{\bar{P}\}$ is called the hypothesis, $\mathcal{W}_2 = \mathcal{W} - \mathcal{W}_1$ the alternative. For a given sample $x \in \mathbb{H}$ we examine if x is distributed according to the distribution in $\mathcal{W}_1$ or not.
A test is defined as a function $\varphi : \mathbb{H} \to \{0, 1\}$ for which $\varphi(x) = 1$ if the hypothesis is rejected and $\varphi(x) = 0$ otherwise.

Maximum Likelihood Estimation

Let $(\mathbb{H}, \mathcal{H}, \mathcal{W})$ be a statistical experiment, where $\mathcal{W}$ denotes a family of probability measures parameterized by Γ, $\mathcal{W} = \{P_\gamma | \gamma \in \Gamma\}$. We want to find out the parameter describing the true probability measure behind a given sample realization. An estimator T is a mapping from the sample space $\mathbb{H}$ to the set of parameters Γ, which assigns to each sample realization $x \in \mathbb{H}$ an estimated value of the unknown parameter γ.

Likelihood Function

For $x \in \mathbb{H}$ the mapping $L_x : \gamma \to \mathbb{R}^+, L_x(\gamma) = \prod_{i=1}^n f_\gamma(x_i), \gamma \in \Gamma$ is called the likelihood function for the sample realization x.

Maximum Likelihood Estimator

A maximum likelihood estimator T is an estimator, which for a given sample realization $x \in \mathbb{H}$ finds the most likely probability distribution P_γ by maximizing the likelihood function L_x:

$$\text{Find } T(x) \text{ s.t. } L_x(T(x)) = \sup_{\gamma \in \Gamma} L_x(\gamma)$$

Maximum A Posteriori

If additionally a prior over the parameters Γ is given, then we obtain a maximum a posteriori estimator.

Maximum A Posteriori

Maximum Likelihood:

$$\hat{\gamma}_{ML}(x) = \underset{\gamma \in \Gamma}{\operatorname{argmax}} \prod_{i=1}^{n} f_\gamma(x_i)$$

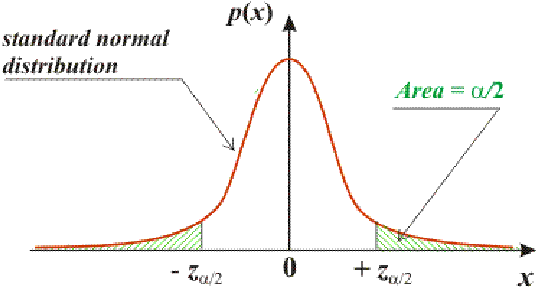
Maximum A Posteriori:

$$\hat{\gamma}_{MAP}(x) = \underset{\gamma \in \Gamma}{\operatorname{argmax}} \prod_{i=1}^{n} \frac{f_\gamma(x_i) p(\gamma)}{\int_{\gamma \in \Gamma} f_\gamma(x_i) p(\gamma) d\gamma}$$

Test Experiment

Let $(\mathbb{H}, \mathcal{H}, \mathcal{W} = \mathcal{W}_1 \cup \mathcal{W}_2)$ be a statistical experiment. A statistical test proceeds as follows

- Define a test statistic T with known image measure $\bar{P}_T$ under the hypothesis.
- Compute the critical region for values of T under the hypothesis using the distribution specific quantiles.
- Decide to either fail to reject the hypothesis or reject the hypothesis.



χ^2-Test

The χ^2 test examines if a given sample $X = (X_1, ..., X_n)$ is distributed according to a probability measure P.

χ^2-Test

Let $(A_1, ..., A_k)$ be a partition of $\mathbb{H}$ into k disjoint subsets. Then we compute the random vector

$$Y_i = \sum_{j=1}^{n} 1_{A_i} \circ X_j,$$

which counts the number of samples falling into A_i. The expected value of Y_i under the hypothesis distribution P is

$$E_P(Y_i) = \sum_{j=1}^{n} E_P(1_{A_i} \circ X_j) = nP_X(A_i)$$

χ^2-Test

The Pearson Test Statistic measures the deviation of the expected value of Y_i from the observed value. Small values support the hypothesis.

$$T(x) = \sum_{i=1}^{n} \frac{(Y_i(x) - nP_X(A_i))^2}{nP_X(A_i)}$$

The statistic T is asymptotically distributed according to the $\chi^2(k-1)$ distribution. Let t be the $(1-\alpha)$-quantile of this distribution. Then

$$P(\{x \in \mathbb{H} : T(x) \leq t\}) = 1 - \alpha$$

If $T(x) > t$ we reject the hypothesis, otherwise we do not reject the hypothesis.

P-Values

Sometimes we do not want a binary decision against a hypothesis.

P-Values

A p-value is the minimum significance level α for which the null-hypothesis is rejected given the value of the test statistic $T(x)$.

$$v(X) = \inf\{\alpha \in [0,1] : \varphi_\alpha(T(X)) = 1\}$$

The larger $v(X)$ the more conform is the value $T(X)$ with the assumed probability measure.