random numbers

statistics and data analysis - (chapter 1)

Björn Malte Schäfer

Graduate School for Fundamental Physics
Fakultät für Physik und Astronomie, Universität Heidelberg

April 17, 2016

outline

- topics of the lectures (monday morning, 09.00h - 11.00h)
  1. random numbers
  2. operations on random numbers
  3. distributions
  4. descriptive statistics
  5. Gaussian magic
  6. statistical tests
  7. linear regression
  8. maximum likelihood and nonlinear fitting
  9. statistical errors and posterior statistics
  10. Monte-Carlo Markov chain techniques
  11. Bayesian statistics and model selection
  12. extreme value statistics
  13. random fields
- transparencies: on moodle, subscribe with password ”roulette”

additional information

- exercises: exercise sheets with questions and a computer task
- course material, additional material and exercise sheets on moodle
- tutor: Alessio Spurio Mancini
- final exam in the study week

literature

- probability theory
  1. Kersting, Wakolbinger: elementare Stochastik
  2. Kersting, Wakolbinger: Zufallsvariable und stochastische Prozesse
  3. Krengel: Einführung in die Wahrscheinlichkeitstheorie und Statistik
  4. Georgii: Stochastik
  5. Lee: Bayesian statistics - an introduction
- data analysis
  1. Held: Methoden der statistischen Inferenz
  2. Cox: principles of statistical inference
  3. Gregory: Bayesian logical data analysis for the physical sciences
  4. Sivia, Skilling: data analysis - a Bayesian tutorial
- statistical physics
  1. Kerson-Huang: statistical physics
  2. Schwabl: statistische Mechanik
  3. Sethna: entropy, order parameters and complexity
  4. Krauth: algorithms and computations
  5. MacKay: information theory, inference and learning algorithms

contents

statistics and data analysis

aims:

- random processes and their description
- statistical background of data analysis
- fitting of data, and parameter inference
- statistics in physical processes
- philosophical questions concerning inference and model selection
- understand the role of randomness in data

the lecture is not...

... providing cookbook recipes for dealing with data

instead...

I hope that you develop intuition about randomness and statistics

progress in science: inference and deduction

- **deduce** phenomena and relations from first principles and model axioms (mathematically exact)
- test models with noisy and incomplete data, **infer** their parameters and validity (look for the most probable model, and the most probable set of parameters)

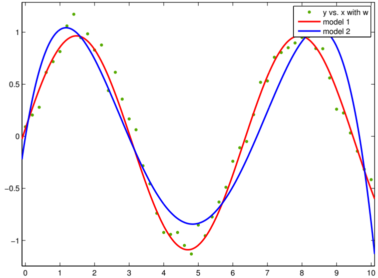
inference...

...is never exact, but a probabilistic statement, due to the noise in the experiment.

perhaps this helps you to remember...

Sherlock Holmes is actually doing inference: he looks for the most probable explanation, but he (wrongly) claims that he'd be doing deduction

motivation



data (x_i, y_i) with error Δy_i, models $y(x)$ with parameters p_μ

- what defines the error Δy_i?
- what is the best model? what are the best model parameters?
- how independent are the parameters, and what are their errors?

10 questions about statistics....

question 1

Is science about finding the truth?

question 2

Are laws of Nature constructed or discovered?

question 3

When would you stop believing in a law of Nature?

question 4

When would you start believing in a law of Nature?

question 5

Can you prove a law of Nature positively?

10 questions about statistics....

question 1

Is science about finding the truth?

question 2

Are laws of Nature constructed or discovered?

question 3

When would you stop believing in a law of Nature?

question 4

When would you start believing in a law of Nature?

question 5

Can you prove a law of Nature positively?

10 questions about statistics....

question 1

Is science about finding the truth?

question 2

Are laws of Nature constructed or discovered?

question 3

When would you stop believing in a law of Nature?

question 4

When would you start believing in a law of Nature?

question 5

Can you prove a law of Nature positively?

10 questions about statistics....

question 1

Is science about finding the truth?

question 2

Are laws of Nature constructed or discovered?

question 3

When would you stop believing in a law of Nature?

question 4

When would you start believing in a law of Nature?

question 5

Can you prove a law of Nature positively?

10 questions about statistics....

question 1

Is science about finding the truth?

question 2

Are laws of Nature constructed or discovered?

question 3

When would you stop believing in a law of Nature?

question 4

When would you start believing in a law of Nature?

question 5

Can you prove a law of Nature positively?

10 questions about statistics....

question 6

Can you prove in science that something does not exist?

question 7

Do you believe in determinism?

question 8

Do you believe in reductionism?

question 9

Would you always prefer a simple model over a complicated one?

question 10

Would you prefer a kind of science without statistics?

10 questions about statistics....

question 6

Can you prove in science that something does not exist?

question 7

Do you believe in determinism?

question 8

Do you believe in reductionism?

question 9

Would you always prefer a simple model over a complicated one?

question 10

Would you prefer a kind of science without statistics?

10 questions about statistics....

question 6

Can you prove in science that something does not exist?

question 7

Do you believe in determinism?

question 8

Do you believe in reductionism?

question 9

Would you always prefer a simple model over a complicated one?

question 10

Would you prefer a kind of science without statistics?

10 questions about statistics....

question 6

Can you prove in science that something does not exist?

question 7

Do you believe in determinism?

question 8

Do you believe in reductionism?

question 9

Would you always prefer a simple model over a complicated one?

question 10

Would you prefer a kind of science without statistics?

10 questions about statistics....

question 6

Can you prove in science that something does not exist?

question 7

Do you believe in determinism?

question 8

Do you believe in reductionism?

question 9

Would you always prefer a simple model over a complicated one?

question 10

Would you prefer a kind of science without statistics?

context 1: probability

in the first three lectures, we will learn to quantify random processes. a random experiment is a random selection of a subset of elements from a set of possible outcomes. how likely it is to obtain a certain selection can be quantified with the probability, whose mathematical definition is given by the Kolmogorov axioms. a random experiment can depend on the outcome of a previously carried out random experiment - in this case one speaks of conditional probabilities. Bayes' law tells us how to combine individual probabilities in the case of conditional probabilities.

context 2: random variables

it is sensible to assign a value to the outcome of a random experiment, which defines a random variable. ultimately, one is interested in the properties of the random variable itself and in its distribution rather than the probability of a certain outcome of a random experiment. possible quantifications of the random variable can be the expectation value and the variance, and in case of two random variables, their covariance. the independency of random variables can be traced back to the independency of their random experiments, and be quantified with the correlation coefficient.

context 3: properties of random variables

the Cauchy-Schwarz inequality quantifies the expecation value of a product of two random variables in relation to the individual expectation values. it is used in the proof of the Chebyshev-inequality, which quantifies the deviation of a random variable from its expectation value. finally, the law of large numbers gives an estimate, how well the expecation value of a random variable can be approximated by the arithmetic average

theory of sets

- sets are not defined in a mathematical sense
  $\rightarrow$ intuitive understanding
- important logical contradictions (B. Russel)
  - set of all possible sets (complement of the empty set)
  - set of all sets which don't contain themselves as elements
- defining a set: $Q = \{x |\, x$ has required property$\}$, meaning:
  x has required property $\leftrightarrow x \in Q$
- empty set $\{\,\}$ is a set with no element

operations on sets

- subset $A \subset Q$: $x \in A \rightarrow x \in Q$, but not necessarily the opposite
  $A \subset A$ is always true
- intersection: $A \cap B = \{x|x:\ x \in A \wedge x \in B\}$, $\wedge$ means **and**
- union $A \cup B = \{x|x:\ x \in A \vee x \in B\}$, $\vee$ means **or**, non-exclusive
- difference $A \setminus B = \{x|x:\ x \in A \wedge x \notin B\}$
- complement $C(A) = \{x|x:\ x \notin A\}$, $\rightarrow A \setminus B = A \cap C(B)$

calculations with sets

operations on sets are defined as logical selection on the set members. complicated operations can always be reduced by logic operations. visualisation is possible with Venn-diagrams

probability: Kolmogorov axioms

- random experiment: a random selection from a finite, non-empty set of events Ω consisting of individual outcomes ω
- power set $\mathcal{P}(\Omega)$: set of all subsets (all possible selections from Ω)
- cardinality $\sharp(\Omega)$: number of elements of Ω
- probability measure: mapping $P : \mathcal{P}(\Omega) \rightarrow [0 \ldots 1]$
- **Kolmogorov-axioms**
  1. $P(\Omega) = 1$
  2. $P(A) \geq 0, \forall A \subset \Omega$
  3. $P(A \cup B) = P(A) + P(B), \forall A \cap B = \{\,\}$
- $P(A)$ is called the probability of an event $\omega \in A \subset \Omega$ happening
- (Ω, P) form a **probability space**

question

if $\sharp(\Omega) = n$, show that $\sharp(\mathcal{P}(\Omega)) = 2^n$ (induction or combinatorics)

conclusions from Kolmogorov axioms

- $P(C(A)) + P(A) = 1$, in particular $P(\{\ \}) = 0$
- $A \subset B \rightarrow P(A) \leq P(B)$
- $P(A \setminus B) = P(A) - P(A \cap B)$, with $A \setminus B = A \cap C(B)$
- $P(A \cup B) + P(A \cap B) = P(A) + P(B)$

question

prove these relations, and visualise them with Venn-diagrams

probability function

- probability function: look at individual events $\omega \in \Omega$
- third Kolmogorov axiom: $P(A \cup B) = P(A) + P(B)$ for disjunct A, B

$$P(\cup_i A_i) = \sum_i P(A_i) \rightarrow P(A) = \sum_{\omega \in \Omega} P(\{\omega\}) \tag{1}$$

- mapping $\omega \rightarrow P(\omega)$ is called **probability function**
- properties:
  - $P(\omega) \geq 0 \ \forall \omega \in \Omega$
  - normalisation $\sum_{\omega \in \Omega} P(\omega) = 1$
- determination of $P(\omega) \equiv P(\{\omega\})$: counting of events ω, which are assumed to be equally likely

$$P(\omega) = \frac{1}{\sharp(\Omega)} \quad \forall \omega \in \Omega \tag{2}$$

$$P(B) = \frac{\sharp(B)}{\sharp(\Omega)} \quad \rightarrow \text{Laplacian probability} \tag{3}$$

conditional probability

- all events $\omega \in \Omega$ have equal probability
- subset $\omega \in A$ is selected by a random process
  $\rightarrow$ events $\omega \in C(A)$ have probability 0
- **conditional probability** of an event $\omega \in B$ under the condition $\omega \in A$

$$P(B|A) = \frac{\sharp(A \cap B)}{\sharp(A)} \tag{4}$$

- be careful: $P(A|B) \neq P(B|A)$ in general
  (A =female, B =pregnant $\rightarrow P(B|A) \simeq 10^{-2}$ but $P(A|B) = 1$)
- **Bayes law**:

$$P(A) = \frac{\sharp(A)}{\sharp(\Omega)} \rightarrow P(B|A) = \frac{P(A \cap B)}{P(A)} \tag{5}$$

- visualisation: with a tree diagram

proporties of the conditional probability

- $A \subset C(B) \rightarrow P(A|B) = 0$
- partition of B, $B = \cup_i B_i$:

$$P(A) = \sum_i P(A|B_i)P(B_i) \quad \text{with} \quad A = \cup_i (A \cap B_i) \tag{6}$$

- **Bayes law**: switch the condition with the random experiment

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_i P(A|B_i)P(B_i)} \quad \text{with} \quad P(B_i|A)P(A) = P(A|B_i)P(B_i) \tag{7}$$

- independency: A and B are independent, if $P(A \cap B) = P(A)P(B)$, equivalent with $P(A) = P(A|B)$
- alternatively: use symmetry of $P(A \cap B) \rightarrow P(A|B)P(B) = P(B|A)P(A)$

question

show that $P(A|B) = P(B|A)$ for independent outcomes A and B

random variables

- random variable: mapping $X : \Omega \to \mathbb{R}$ (or $\mathbb{C}$, $\mathbb{R}^n$, $\mathbb{C}^n$) is an $\mathbb{R}$-valued random variable (should actually be called random mapping)
- motivation: subsets $A \subset \Omega$ are outcomes of a random experiment, every subset of Ω is given a **value** X
  - the drawing of a ticket is a random experiment, the money one wins is the random variable associated with the random experiment
  - transitions in a hydrogen atom are random experiments, but the frequency of the light observed is a random variable
  - **random variables have units**, probabilities don't
- distribution: how probable are values of X in a random experiment selecting $A \subset \Omega$?
  - if Ω is countable $\to \{X(\omega)|\omega \in \Omega\}$ is countable
  - $P(x) = P(\{\omega \in \Omega : X(\omega) = x\})$ is called **distribution**
  - visualisation: histogram of values
  - NB: we're not getting into complications with uncountable Ω ($\to$ Borel)
- independency of random variables are traced back to their random processes selecting the subsets $A, B \subset \Omega$

expectation value

- expectation value $E(X)$ of a random variable X

$$E(X) = \sum_{\omega\in\Omega} X(\omega)P(\omega) \tag{8}$$

- for all values x_i (remember, finite!) of X we get:

$$E(X) = \sum_i \sum_{\omega:X(\omega)=\omega_i} X(\omega)P(\omega) = \sum_i x_i P(X = x_i) \tag{9}$$

- properties of the expectation value
  - $E(\lambda X) = \lambda E(X)$
  - $E(X + Y) = E(X) + E(Y)$
  - X, Y independent $\rightarrow E(XY) = E(X)E(Y)$

variance, covariance and correlation coefficient

- variance $\mathrm{Var}(X)$ of a random variable X

$$\mathrm{Var}(X) = E\left([X - E(X)]^2\right) \tag{10}$$

- covariance $\mathrm{Cov}(X, Y)$ **between** two random variables X, Y

$$\mathrm{Cov}(X, Y) = E\left([X - E(X)]\,[Y - E(Y)]\right) \tag{11}$$

- correlation coefficient ρ_{XY} between X and Y

$$\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \tag{12}$$

- standard deviation σ_X

$$\sigma_X = \sqrt{\mathrm{Var}(X)} \tag{13}$$

but be careful: σ_X is defined for any random variable, and does not imply a Gaussian process. we'll show that for a Gaussian random variable σ_X is equal to the parameter σ

properties of the (co)variance

- properties of the variance
  - $\mathrm{Var}(X) = E(X^2) - E(X)^2$
  - $\mathrm{Var}(\lambda X) = \lambda^2 \mathrm{Var}(X)$
  - $\mathrm{Var}(X + c) = \mathrm{Var}(X)$
- properties of the covariance
  - $\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y)$
  - $\mathrm{Cov}(\lambda X, \mu Y) = \lambda\mu \mathrm{Cov}(X, Y)$
- interpretation of the correlation coefficient
  - X, Y independent $\rightarrow \rho_{XY} = 0$
  - X, Y independent $\rightarrow \mathrm{Var}(X, Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

question

derive the relations for the (co)variance

Cauchy-Schwarz inequality

- Cauchy-Schwarz inequality

$$|E(XY)|^2 \leq E(|X|^2)\, E(|Y|^2) \tag{14}$$

- proof: set $\alpha = E(Y^2) > 0$, and $\beta = -E(XY)$

$$0 \leq E\left(|\alpha X + \beta Y|^2\right) = \ldots = \alpha\left[E(|X|^2)E(|Y|^2) - |E(XY)|^2\right] \tag{15}$$

- divide out α, which does not change the sign, and do $\sqrt{\ldots}$ which is monotonic

question

carry out the proof of the Cauchy-Schwarz inequality

question

show that $|\rho_{XY}| \leq 1$ using the Cauchy-Schwarz inequality

Chebyshev inequality

- X is a random variable following from a random process defined in (Ω, P), with finite variance. then, the **Chebyshev**-inequality applies

$$P(|X - E(X)| \geq \epsilon) \leq \frac{\mathrm{Var}(X)}{\epsilon^2} \quad \text{for all } \epsilon > 0 \tag{16}$$

- proof: abbreviate $Z = X - E(X)$ (center the random variable)
- set

$$Y(\omega) = \begin{cases} 0 & \text{for } \omega \text{ with } |Z(\omega)| < \epsilon \\ \epsilon^2 & \text{for } \omega \text{ with } |Z(\omega)| \geq \epsilon \end{cases} \tag{17}$$

- then, $Y \leq |Z|^2$ and consequently:

$$\mathrm{Var}(X) = E(|Z|^2) \geq E(Y) = \epsilon^2 P(Y \geq \epsilon^2) = \epsilon^2 P(|X - E(X)| \geq \epsilon) \tag{18}$$

because $Y \geq \epsilon^2$ implies $|Z| \geq \epsilon$

- $\rightarrow$ the Chebyshev inequality quantifies, how often large values appear in relation to the variance of a distribution

law of large numbers (weak formulation)

- $X_1, X_2, \ldots, X_n$ are **independent** random numbers with equal expectation values and finite variances $\mathrm{Var}(X_i) \leq M < \infty$. then, the **law of large numbers** applies

$$P\left(\left|\frac{X_1 + \ldots + X_n}{n} - E(X_1)\right| \geq \epsilon\right) \leq \frac{M}{n\epsilon^2} \quad \text{for all } \epsilon > 0 \tag{19}$$

- proof: set $X = (X_1 + \ldots + X_n)/n$, such that $E(X) = E(X_1)$ and finally

$$\mathrm{Var}(X) = \frac{1}{n^2}\mathrm{Var}(X_1 + \ldots + X_n) = \frac{1}{n^2}\left(\mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)\right) \leq \frac{M}{n} \tag{20}$$

- apply the Chebyshev-inequality:

$$P(|X - E(X)| \geq \epsilon) \leq \frac{\mathrm{Var}(X)}{\epsilon^2} \leq \frac{M}{n\epsilon^2} \tag{21}$$

- $\rightarrow$ the expecation value of a random variable can be approximated by the **arithmetic mean**repeating the experiment, and the accuracy is related to the variance of the random variable

summary

- random experiments: random selection from a set of possible outcomes
- probability theory: definition of probability with the Kolmogorov axioms
- conditional probability for dependent random experiments, Bayes
- random variables: assigning a value to the outcome of a random experiment
- quantification of random variable: expectation value and variance
- Chebyshev-inequality: magnitude of deviation of random variable from its mean
- law of large numbers: estimate of the expectation value from the arithmetic mean of a random variable from repeated random experiments