Statistische Methoden der Datenanalyse

Kapitel 1: Fundamentale Konzepte

Professor Markus Schumacher

Freiburg / Sommersemester 2009
Data analysis in particle physics

Observe events of a certain type

Measure characteristics of each event (particle momenta, number of muons, energy of jets,...)

Theories (e.g. SM) predict distributions of these properties up to free parameters, e.g., $\alpha$, $G_F$, $M_Z$, $\alpha_s$, $m_H$, ...

Some tasks of data analysis:

Estimate (measure) the parameters;

Quantify the uncertainty of the parameter estimates;

Test the extent to which the predictions of a theory are in agreement with the data.
Dealing with uncertainty

In particle physics there are various elements of uncertainty:

- theory is not deterministic
- quantum mechanics
- random measurement errors
- present even without quantum effects
- things we could know in principle but don’t
  e.g. from limitations of cost, time, ...

We can quantify the uncertainty using PROBABILITY
A definition of probability

Consider a set $S$ with subsets $A, B, \ldots$

For all $A \subset S$, $P(A) \geq 0$

$P(S) = 1$

If $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$

From these axioms we can derive further properties, e.g.

$P(\overline{A}) = 1 - P(A)$

$P(A \cup \overline{A}) = 1$

$P(\emptyset) = 0$

if $A \subset B$, then $P(A) \leq P(B)$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
Conditional probability, independence

Also define conditional probability of $A$ given $B$ (with $P(B) \neq 0$):

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

E.g. rolling dice:

$$P(n < 3 \mid n \text{ even}) = \frac{P((n<3) \cap n \text{ even})}{P(\text{even})} = \frac{1/6}{3/6} = \frac{1}{3}$$

Subsets $A$, $B$ independent if:

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

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

N.B. do not confuse with disjoint subsets, i.e. $A \cap B = \emptyset$
Interpretation of probability

I. Relative frequency

\[ P(A) = \lim_{n \to \infty} \frac{\text{times outcome is } A}{n} \]

\( A, B, \ldots \) are outcomes of a repeatable experiment
cf. quantum mechanics, particle scattering, radioactive decay...

II. Subjective probability

\( A, B, \ldots \) are hypotheses (statements that are true or false)

\[ P(A) = \text{degree of belief that } A \text{ is true} \]

- Both interpretations consistent with Kolmogorov axioms.
- In particle physics frequency interpretation often most useful, but subjective probability can provide more natural treatment of non-repeatable phenomena: systematic uncertainties, probability that Higgs boson exists,...
Bayes’ theorem

From the definition of conditional probability we have,

\[ P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{and} \quad P(B|A) = \frac{P(B \cap A)}{P(A)} \]

but \( P(A \cap B) = P(B \cap A) \), so

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]

First published (posthumously) by the Reverend Thomas Bayes (1702–1761)

The law of total probability

Consider a subset $B$ of the sample space $S$, divided into disjoint subsets $A_i$ such that $\bigcup_i A_i = S$,

\[ B = B \cap S = B \cap (\bigcup_i A_i) = \bigcup_i (B \cap A_i), \]

\[ P(B) = P(\bigcup_i (B \cap A_i)) = \sum_i P(B \cap A_i) \]

\[ P(B) = \sum_i P(B | A_i) P(A_i) \]

Bayes’ theorem becomes

\[
P(A|B) = \frac{P(B | A) P(A)}{\sum_i P(B | A_i) P(A_i)}
\]
An example using Bayes’ theorem

Suppose the probability (for anyone) to have AIDS is:

\[
P(\text{AIDS}) = 0.001 \]
\[
P(\text{no AIDS}) = 0.999
\]

Consider an AIDS test: result is + or −

\[
P(\text{+} | \text{AIDS}) = 0.98
\]
\[
P(\text{−} | \text{AIDS}) = 0.02
\]
\[
P(\text{+} | \text{no AIDS}) = 0.03
\]
\[
P(\text{−} | \text{no AIDS}) = 0.97
\]

Suppose your result is +. How worried should you be?
Bayes’ theorem example (cont.)

The probability to have AIDS given a + result is

$$P(\text{AIDS}|+) = \frac{P(+)\mid \text{AIDS} P(\text{AIDS})}{P(+)\mid \text{AIDS} P(\text{AIDS}) + P(+)\mid \text{no AIDS} P(\text{no AIDS})}$$

$$= \frac{0.98 \times 0.001}{0.98 \times 0.001 + 0.03 \times 0.999}$$

$$= 0.032 \quad \leftarrow \text{posterior probability}$$

i.e. you’re probably OK!

Your viewpoint: my degree of belief that I have AIDS is 3.2%

Your doctor’s viewpoint: 3.2% of people like this will have AIDS
In frequentist statistics, probabilities are associated only with the data, i.e., outcomes of repeatable observations (shorthand: \( \overline{x} \)).

Probability = limiting frequency

Probabilities such as

\[ P(\text{Higgs boson exists}), \]
\[ P(0.117 < \alpha_s < 0.121), \]

etc. are either 0 or 1, but we don’t know which.

The tools of frequentist statistics tell us what to expect, under the assumption of certain probabilities, about hypothetical repeated observations.

The preferred theories (models, hypotheses, ...) are those for which our observations would be considered ‘usual’.
Bayesian Statistics – general philosophy

In Bayesian statistics, use subjective probability for hypotheses:

- probability of the data assuming hypothesis $H$ (the likelihood)
- prior probability, i.e., before seeing the data
- posterior probability, i.e., after seeing the data
- normalization involves sum over all possible hypotheses

Bayes’ theorem has an “if-then” character: If your prior probabilities were $\pi(H)$, then it says how these probabilities should change in the light of the data.
No general prescription for priors (subjective!)
Random variables and probability density functions

A random variable is a numerical characteristic assigned to an element of the sample space; can be discrete or continuous.

Suppose outcome of experiment is continuous value $x$

$$P(x \text{ found in } [x, x + dx]) = f(x) \, dx$$

$\rightarrow f(x) =$ probability density function (pdf)

$$\int_{-\infty}^{\infty} f(x) \, dx = 1 \quad \text{x must be somewhere}$$

Or for discrete outcome $x_i$ with e.g. $i = 1, 2, \ldots$ we have

$$P(x_i) = p_i \quad \text{probability mass function}$$

$$\sum_i P(x_i) = 1 \quad \text{x must take on one of its possible values}$$
Cumulative distribution function

Probability to have outcome less than or equal to $x$ is

$$\int_{-\infty}^{x} f(x') \, dx' \equiv F(x)$$

cumulative distribution function

Alternatively define pdf with

$$f(x) = \frac{\partial F(x)}{\partial x}$$
Mean, median and mode

e.g.: Maxwells’s velocity distribution

![Graph showing mean, median, and mode](image-url)
pdf = histogram with
infinite data sample,
zero bin width,
normalized to unit area

\[ f(x) = \frac{N(x)}{n\Delta x} \]

\( n \) = number of entries
\( \Delta x \) = bin width
Multivariate distributions

Outcome of experiment characterized by several values, e.g. an \( n \)-component vector, \((x_1, \ldots, x_n)\)

\[
P(A \cap B) = f(x, y) \, dx \, dy
\]

Joint pdf

Normalization:

\[
\int \cdots \int f(x_1, \ldots, x_n) \, dx_1 \cdots dx_n = 1
\]
Marginal pdf

Sometimes we want only pdf of some (or one) of the components:

\[
P(A) = \sum_i P(A \cap B_i) = \sum_i f(x, y_i) \, dy \, dx \rightarrow \int f(x, y) \, dy \, dx
\]

\[
f_x(x) = \int f(x, y) \, dy
\]

\[\rightarrow \text{marginal pdf} \quad f_1(x_1) = \int \cdots \int f(x_1, \ldots, x_n) \, dx_2 \cdots dx_n\]

\(x_1, x_2\) independent if \(f(x_1, x_2) = f_1(x_1)f_2(x_2)\)
Marginal pdf (2)

Marginal pdf ~ projection of joint pdf onto individual axes.
Conditional pdf

Sometimes we want to consider some components of joint pdf as constant. Recall conditional probability:

\[ P(B | A) = \frac{P(A \cap B)}{P(A)} = \frac{\int f(x, y) \, dx \, dy}{\int f_x(x) \, dx} \]

→ conditional pdfs:

\[ h(y | x) = \frac{f(x, y)}{f_x(x)} , \quad g(x | y) = \frac{f(x, y)}{f_y(y)} \]

Bayes’ theorem becomes:

\[ g(x | y) = \frac{h(y | x) f_x(x)}{f_y(y)} \]

Recall \( A, B \) independent if

\[ P(A \cap B) = P(A)P(B) \]

→ \( x, y \) independent if

\[ f(x, y) = f_x(x) f_y(y) \]
E.g. joint pdf $f(x,y)$ used to find conditional pdfs $h(y|x_1)$, $h(y|x_2)$:

Basically treat some of the r.v.s as constant, then divide the joint pdf by the marginal pdf of those variables being held constant so that what is left has correct normalization, e.g.,

$$\int h(y|x) \, dy = 1.$$
WDF für zwei Variablen mit Abhängigkeit

(a) Punktwolke
(b) Verteilung von Y
(c) Verteilung von X
(d) Verteilung von X|Y
2.4 Functions of a random variable

A function of a random variable is itself a random variable.
Suppose $x$ follows a pdf $f(x)$, consider a function $a(x)$.
What is the pdf $g(a)$?

$$g(a) \, da = \int_{dS} f(x) \, dx$$

$dS = \text{region of } x \text{ space for which } a \text{ is in } [a, a+da]$.
For one-variable case with unique inverse this is simply

$$g(a) \, da = f(x(a)) \left| \frac{dx}{da} \right|$$
Mapping the x and a spaces

probability in a-space

\[ g(a) \, da \]

equals one in x-space

\[ f(x) \, dS \]

figure from Lutz Feld
Mapping the x and u spaces
Mapping the x and a spaces: two „branches“
If inverse of $a(x)$ not unique, include all $dx$ intervals in $dS$ which correspond to $da$:

**Example:**

$$a = x^2, \quad x = \pm \sqrt{a}, \quad dx = \pm \frac{da}{2\sqrt{a}}.$$  

$$dS = \left[ \sqrt{a}, \sqrt{a} + \frac{da}{2\sqrt{a}} \right] \cup \left[ -\sqrt{a} - \frac{da}{2\sqrt{a}}, -\sqrt{a} \right]$$  

$$g(a) = \frac{f(\sqrt{a})}{2\sqrt{a}} + \frac{f(-\sqrt{a})}{2\sqrt{a}}$$
Consider r.v.s \( \vec{x} = (x_1, \ldots, x_n) \) and a function \( a(\vec{x}) \).

\[
g(a')da' = \int \cdots \int_{dS} f(x_1, \ldots, x_n) dx_1 \cdots dx_n
\]

\( dS = \) region of \( x \)-space between (hyper)surfaces defined by

\[
a(\vec{x}) = a', \quad a(\vec{x}) = a' + da'
\]
Example: r.v.s $x, y > 0$ follow joint pdf $f(x,y)$, consider the function $z = xy$. What is $g(z)$?

\[
g(z) \, dz = \int \cdots \int_{dS} f(x,y) \, dx \, dy \\
= \int_0^\infty dx \int_{z/x}^\infty \frac{(z+dz)}{x} f(x,y) \, dy \\
\rightarrow \\
g(z) = \int_0^\infty f\left(x, \frac{z}{x}\right) \frac{dx}{x} \\
= \int_0^\infty f\left(\frac{z}{y}, y\right) \frac{dy}{y}
\]

(Mellin convolution)
More on transformation of variables

Consider a random vector \( \Vec{x} = (x_1, \ldots, x_n) \) with joint pdf \( f(\Vec{x}) \).

Form \( n \) linearly independent functions \( \Vec{y}(\Vec{x}) = (y_1(\Vec{x}), \ldots, y_n(\Vec{x})) \) for which the inverse functions \( x_1(\Vec{y}), \ldots, x_n(\Vec{y}) \) exist.

Then the joint pdf of the vector of functions is

\[
g(\Vec{y}) = |J| f(\Vec{x})
\]

where \( J \) is the Jacobian determinant:

\[
J = \begin{vmatrix}
\frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\
\frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n}
\end{vmatrix}
\]

For e.g. \( g_1(\Vec{y}_1) \) integrate \( g(\Vec{y}) \) over the unwanted components.
Consider continuous r.v. $x$ with pdf $f(x)$.

Define expectation (mean) value as

$$E[x] = \int x f(x) \, dx$$

Notation (often):

$E[x] = \mu$ \text{ ~ centre of gravity of pdf.}$

For a function $y(x)$ with pdf $g(y)$,

$$E[y] = \int y g(y) \, dy = \int y(x) f(x) \, dx$$

( equivalent )

Variance:

$$V[x] = E[x^2] - \mu^2 = E[(x - \mu)^2]$$

Notation:

$V[x] = \sigma^2$

Standard deviation:

$$\sigma = \sqrt{\sigma^2}$$

$\sigma \sim \text{width of pdf, same units as } x.$
Covariance and correlation

Define covariance \( \text{cov}[x,y] \) (also use matrix notation \( V_{xy} \)) as

\[
\text{Cov}[x, y] = E[xy] - \mu_x \mu_y = E[(x - \mu_x)(y - \mu_y)]
\]

Correlation coefficient (dimensionless) defined as

\[
\rho_{xy} = \frac{\text{Cov}[x, y]}{\sigma_x \sigma_y}
\]

If \( x, y \), independent, i.e.,

\[
f(x, y) = f_x(x)f_y(y), \quad \text{then}
\]

\[
E[xy] = \int \int xy f(x, y) \, dx \, dy = \mu_x \mu_y
\]

\( \rightarrow \) \( \text{cov}[x, y] = 0 \) \( x \) and \( y \), ‘uncorrelated’

N.B. converse not always true.
Correlations

\[ \rho = 0.75 \]

\[ \rho = -0.75 \]

\[ \rho = 0.95 \]

\[ \rho = 0.25 \]
Korrelationen

(a) \( \rho = 0.00 \)

(b) \( \rho = 0.92 \)

(c) \( \rho = -0.60 \)

(d) \( \rho = 0.00 \)