Concepts of Experiments at Future Colliders II

PD Dr. Oliver Kortner

14.06.2024

Examples of important probability distributions

Recapitulation of the previous lecture

Interval estimation

Goal: Determination of an interval which contains the true value of a parameter with a given probability.

Limit case of the normal distribution

Let us assume the variable $x \in \mathbb{R}$ is normally distributed, i.e.

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

If μ and σ are known, then

$$p(a < x < b) = \int_a^b N(x; \mu, \sigma)dx =: \beta.$$

If μ is unknown, one can calculate $p(\mu + c < x < \mu + d)$:

$$\begin{aligned}\beta = p(\mu + c < x < \mu + d) &= \int_{\mu+c}^{\mu+d} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx = \int_c^d \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{y^2}{\sigma^2}} dy \\ &= p(c - x < -\mu < d - x) = p(x - d < \mu < x - c).\end{aligned}$$

Interval estimation with the normal distribution

$$\beta = p(\mu + c < x < \mu + d) = \int_{\mu+c}^{\mu+d} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx = \int_{c}^{d} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{y^2}{\sigma^2}} dy$$
$$= p(c - x < -\mu < d - x) = p(x - d < \mu < x - c).$$

That is, if x has been measured, the probability that the desired value of μ lies between $x - d$ and $x - c$ is equal to β.

- If x is a parameter $\hat{\alpha}$ from a point estimation conducted using the method of maximum likelihood or the method of least squares, then $\hat{\alpha}$ is asymptotically normally distributed, and the above formulas can be applied for interval estimation.
- The intervals $[a, b]$ or $[x - d, x - c]$ are called confidence intervals. β is the confidence level corresponding to the confidence level.

Generalization to the multidimensional case

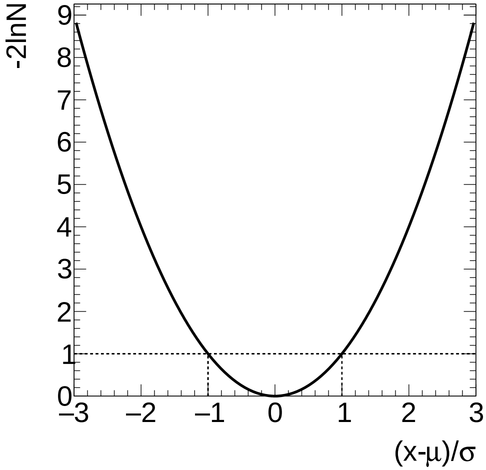
$$Q(x;\mu,\Sigma) := (x-\mu)^t\Sigma^{-1}(x-\mu),\ \ x,\mu \in |\mathrm{R}.$$
$$p(Q) = \frac{1}{(2\pi)^{d/2}} \cdot \frac{1}{\sqrt{det(\Sigma)}} \exp\left(-\frac{1}{2}Q(x;\mu,\Sigma)\right).$$

In multiple dimensions, the confidence interval becomes a confidence region corresponding to the confidence level β:

$$p(Q(x;\mu,\Sigma) < K_\beta^2) = \beta.$$

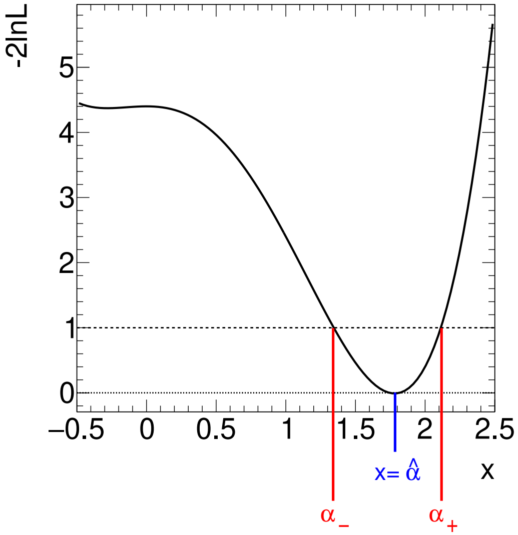
Likelihood-based confidence intervals

$$-2\ln N(x=\mu\pm\sigma;\mu,\sigma) - [-2\ln N(x=\mu;\mu,\sigma] = 1.$$



Likelihood-based confidence intervals

Generalization



Confidence Interval: $[\alpha_-, \alpha+]$.

Recapitulation of the previous lecture

Hypothesis testing

Goal, to determine which hypothesis (for a probability distribution) describes the recorded data point distributions (data).

Nomenclature. H_0: null hypothesis.
H_1: alternative hypothesis.

Simple and Composite Hypotheses

- When the hypotheses H_0 and H_1 are given completely without free parameters, the hypotheses are called simple hypotheses.
- If a hypothesis contains at least one free parameter, it is referred to as a composite hypothesis.

Procedure

For hypothesis testing, W must be chosen such that

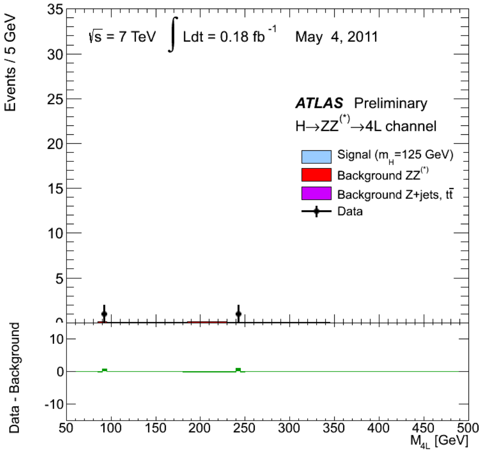
$$p(\text{data} \in W | H_0) = \alpha$$

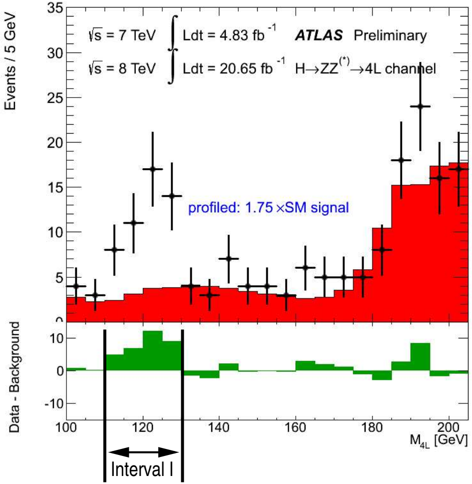
with a small value of α and simultaneously

$$p(\text{data} \in D \setminus W | H_1) = \beta$$

with the smallest possible β.

Discovery of the Higgs boson





Number of events in I

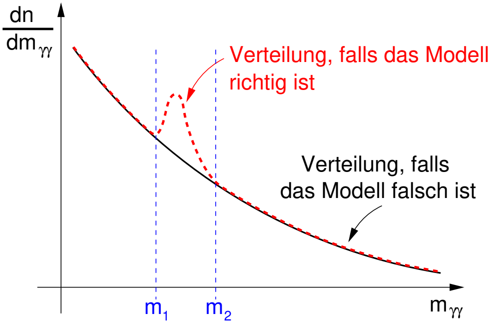
- 13.6 without a Higgs boson,
- Observed: 50.

$\Rightarrow$ Excess of 36.4 Events $> 13,6 + 5 \cdot \sqrt{13,6} = 32.$

- So the probability that the observed excess is caused by a statistical fluctuation of the "red distribution" is extremely small.

Introductory example of hypothesis testing

A theoretical model predicts the existence of a particle with mass M, the production cross-section, and the partial width for decay into a photon pair. To confirm or refute this model, one must examine the distribution of $m_{\gamma\gamma}$.



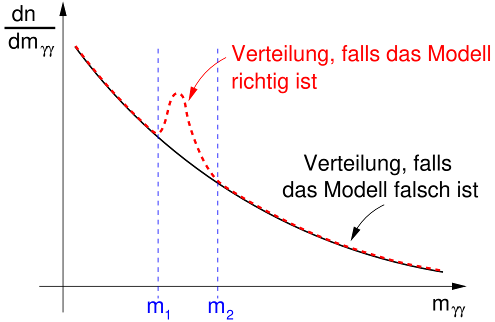
In the interval $[m_1, m_2]$, one is sensitive to the model's prediction. There are two hypotheses, namely that the theory is correct or incorrect.

H_0: Null hypothesis: TTheory is incorrect.“

H_1: Alternative hypothesis: TTheory is correct.“

With a sufficiently large amount of data, the probability that the measured $m_{\gamma\gamma}$ distribution looks like H_0 is small if the theory is correct. At the same time, the probability that the measured mass distribution looks like H_1 is large.

Introductory example of hypothesis testing



n: Number of events measured in the interval $[m_1, m_2]$.
One must now choose a threshold value N such that

$$p(n > N|H_0) = \alpha$$

with a small value of α and

$$p(n \leq N|H_1) = \beta$$

is as small as possible if the theory, i.e., H_1, is correct.

Recapitulation of the previous lecture

Introductory example – experimental practice

n: Number of events measured in the interval $[m_1, m_2]$.
One must now choose a threshold value N such that

$$p(n > N|H_0) = \alpha$$

with a small value of α and

$$p(n \leq N|H_1) = \beta$$

is as small as possible if the theory, i.e., H_1, is correct.

Experimental Practice

- $\alpha = 5.7 \cdot 10^{-7}$, which corresponds to 5$\sigma$ of a normal distribution, to claim the discovery of a particle.
- With a value of $\alpha = 0.3\%$, which corresponds to 3σ of a normal distribution, one says there is evidence for the existence of a new particle.

Type I and type II errors

The confidence level α is defined as the probability that $x \in W$ if the null hypothesis H_0 is correct:

$$p(x \in W | H_0) = \alpha.$$

The probability β represents the likelihood of incorrectly rejecting the alternative hypothesis H_1:

$$p(x \in D \backslash W | H_1) = \beta.$$

Approach	H_0 correct	H_1 correct
$x \notin W \Rightarrow H_0$ is considered correct	Good acceptance, since $p(x \in D\backslash W \| H_0) = 1 - \alpha$ is large	Contamination Type II error $p(x \in D\backslash W \| H_1) = \beta$.
$x \in W \Rightarrow H_0$ is rejected, H_1 is considered correct	Wrong decision Type I error $p(x \in W \| H_0) = \alpha$ is small	Rejecting H_0 good, since $p(x \in W \| H_1) = 1 - \beta$ is large.

The Neyman-Pearson test for simple hypotheses

$x = (x_1, \ldots, x_n)$: Random variable with the probability density $f_N(x;\theta)$.

- $\theta = \theta_0$: Null hypothesis.
- $\theta = \theta_1$: Alternative hypothesis.

$$\begin{aligned}
\alpha &= \int\limits_{W_\alpha} f_N(x;\theta_0)\,dx. \\
1-\beta &= \int\limits_{W_\alpha} f_N(x;\theta_1)\,dx = \int\limits_{W_\alpha} f_N(x;\theta_1)\cdot\frac{f_N(x;\theta_0)}{f_N(x;\theta_0)}\,dx = \int\limits_{W_\alpha} \frac{f_N(x;\theta_1)}{f_N(x;\theta_0)}\cdot f_N(x;\theta_0)\,dx \\
&= E_{W_\alpha}\left(\frac{f_N(x;\theta_1)}{f_N(x;\theta_0)}\right).
\end{aligned}$$

- E_{W_α} becomes particularly large when W_α contains those points x for which $\frac{f_N(x;\theta_1)}{f_N(x;\theta_0)}$ becomes particularly large.
- The best critical region is chosen by requiring

$$\ell_N(x;\theta_0,\theta_1) := \frac{f_N(x;\theta_1)}{f_N(x;\theta_0)} \geq c_\alpha$$

for the points $x \in W_\alpha$.

The Neyman-Pearson test for simple ypotheses

$x = (x_1, \ldots, x_n)$: Random variable with the probability density $f_N(x;\theta)$.

- $\theta = \theta_0$: Null hypothesis.
- $\theta = \theta_1$: Alternative hypothesis.
- The best critical region is chosen by requiring

$$\ell_N(x;\theta_0,\theta_1) := \frac{f_N(x;\theta_1)}{f_N(x;\theta_0)} \geq c_\alpha$$

  for the points $x \in W_\alpha$.

Neyman-Pearson Test. The likelihood ratio $\ell_N(x;\theta_0,\theta_1)$ is used as the decision criterion:

- $\ell_N(x;\theta_0,\theta_1) \geq c_\alpha \Rightarrow H_1$ is accepted, H_0 is rejected.
- $\ell_N(x;\theta_0,\theta_1) < c_\alpha \Rightarrow H_0$ is accepted, H_1 is rejected.

Generalization to composite hypotheses

H_0 and H_1 contain unknown free parameters θ.

- Θ: Set of all possible θ values.
- ν: Subset of Θ.

Two families of hypotheses are introduced:

- H_0: $\theta \in \nu$.
- H_1: $\theta \in \Theta \backslash \nu$.

Example. Coupling strength g. H_0: $g = 0$. H_1: $g > 0$.

ℓ_N is replaced by λ, which is taken as the ratio of two maximized likelihood functions:

$$\lambda := \frac{\max\limits_{\theta \in \nu} L(x;\theta)}{\max\limits_{\theta \in \Theta} L(x;\theta)}$$

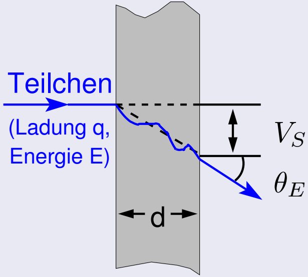
That is, using the method of maximum likelihood, the value of θ that best describes the experimental data is determined, and the corresponding likelihood value is compared with the result of the likelihood maximization for the null hypothesis. If the null hypothesis yields a significantly worse likelihood than the best description of the data, the null hypothesis is rejected.

Monte Carlo method

- The probability density describing, for example, the outcome of a proton-proton collision measurement is composed of many probability densities and generally cannot be given analytically.
- The probability distribution can be obtained using the so-called Monte Carlo method.
    - In the Monte Carlo method, the overall process is broken down into subprocesses $T_1, \ldots T_n$, for which the probability densities are known.
    - Using a random number generator, an outcome of T_1 is generated according to its probability density.
    - For this outcome of T_1, the outcome of T_2 is generated accordingly, and this procedure is continued up to T_n.
    - If this is repeated very often, the probability distribution for the overall process is gradually obtained.

Example: Scattering of muons in thick layers

Scattering of heavy charged particles in thin layers



Strahlungslänge der Schicht: X_0

- Energy loss in the layer is negligible.
- $V_S \approx \frac{1}{2} d\theta_E$.
- θ_E is approximately normally distributed around 0 with the standard deviation

$$\theta_0 := \frac{13.6\ \mathrm{MeV}}{E}\sqrt{\frac{d}{X_0}}.$$

Example: Scattering of muons in thick layers

Goal: Scattering in thick layers

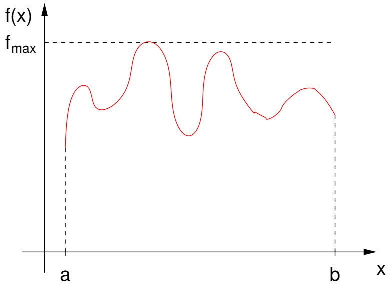
Differences

- $V_S \neq \frac{1}{2} d\theta_E$.
- Energy loss is not negligible.

Solution

- Divide the thick layer into many thin layers.
- Describe the passage through thick layers as a sequence of random processes, namely as a sequence of scatterings in the thin layers.

- (Deterministic) random number generators refer to computer programs that can generate a sequence of pseudorandom numbers.
- They are called pseudorandom numbers because, although the generated numbers appear random, they are produced using a deterministic algorithm.
- If you have a random number generator that produces random numbers uniformly distributed in an interval $[a, b]$, you can generate random numbers according to any probability distribution.



- First, generate a uniformly distributed random number $x \in [a, b]$.
- Then, generate a uniformly distributed random number $y \in [0, f_{max}]$.
- If $y < f(x)$, keep the random number x; otherwise, discard it and generate a new number $x \in [a, b]$ until $y < f(x)$.

Interaction of particles with matter — A recapitulation

Interaction of heavy charged particles with matter

Two effects when charged particles pass through matter

- Energy loss.
- Deflection from the original trajectory.

Responsible processes

- Inelastic collisions with atomic electrons in the material.
- Elastic scattering off the atomic nuclei in the material.
- Emission of Čerenkov radiation.
- Nuclear reactions.
- Bremsstrahlung.

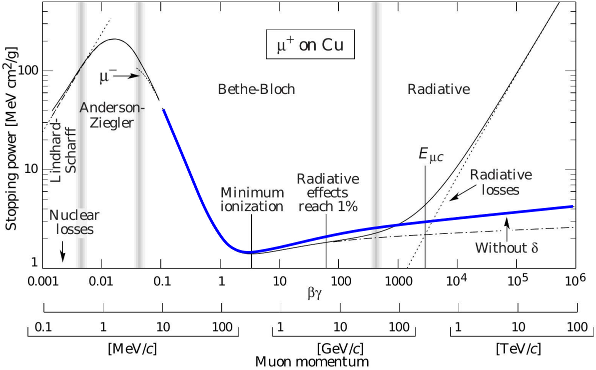
For heavy charged particles, the first two processes are dominant.

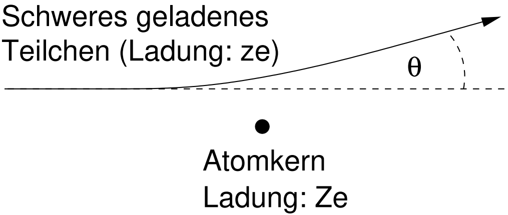
Heavy charged particles include $\mu^{\pm}$, $\pi^{\pm}$, p, $\bar{p}$, α particles, and light nuclei.

Heavy charged particles lose energy through excitation and ionization of atoms. The energy loss per unit path length is described by the Bethe-Bloch formula:

$$-\frac{dE}{dx} = \frac{4\pi n z^2}{m_e c^2 \beta^2} \cdot \left(\frac{e^2}{4\pi\epsilon_0}\right)^2 \cdot \left[\ln\left(\frac{2m_e c^2 \beta^2}{I(1-\beta^2)} - \beta^2\right)\right];$$

$\beta = v/c$: Velocity of the particle. E: Energy of the particle.
z: Charge of the particle. e: Elementary charge. n: Electron density of the material.
I: Mean excitation potential of the material.

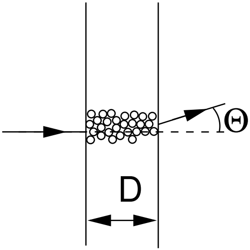




Scattering off a single nucleus:

$$\theta = \frac{\Delta p}{p} \propto \frac{z \cdot Z}{p}.$$

$$< \theta >= 0, 0 \neq \theta_0^2 := Var(\theta) \propto \frac{z^2 \cdot Z^2}{p^2}.$$



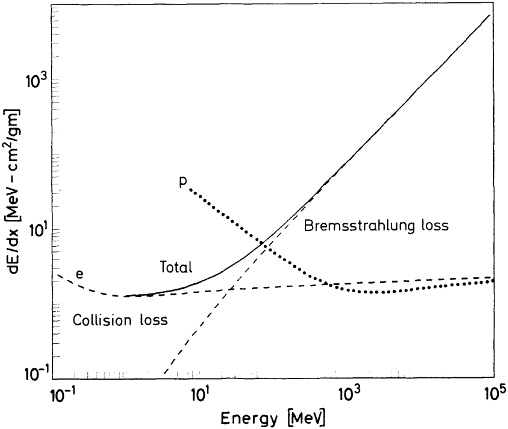
Scattering off many nuclei:

$$\begin{aligned} < \Theta > &= 0 \\ \Theta_0^2 := Var(\Theta) = \sum_{collisions} \theta_0^2 \propto D \cdot \frac{z^2 \cdot Z^2}{p^2}. \end{aligned}$$

Thus, one obtains $\boxed{\Theta_0 \propto \frac{\sqrt{D}}{p}.}$

The mass m_e is so small that the acceleration experienced by electrons or positrons in collisions with atomic nuclei is significant enough to emit bremsstrahlung photons.

$$\left.\frac{dE}{dx}\right|_{e^\pm} = \left.\frac{dE}{dx}\right|_{collisions} + \left.\frac{dE}{dx}\right|_{bremsstrahlung} .$$



Critical Energy E_k

$$\left.\frac{dE}{dx}\right|_{collisions} (E_k) = \left.\frac{dE}{dx}\right|_{bremsstrahlung} (E_k).$$

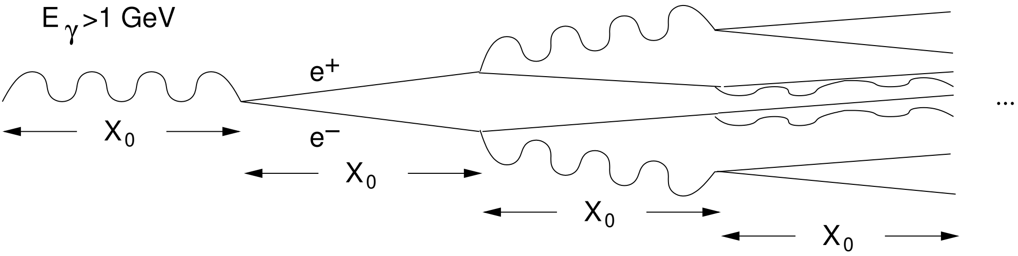
$E_k \approx \frac{800 \text{ MeV}}{Z+1/2}$, hence above $E_{e^\pm} > 1$ GeV, bremsstrahlung is dominant.

Interaction of photons with matter

Main Processes

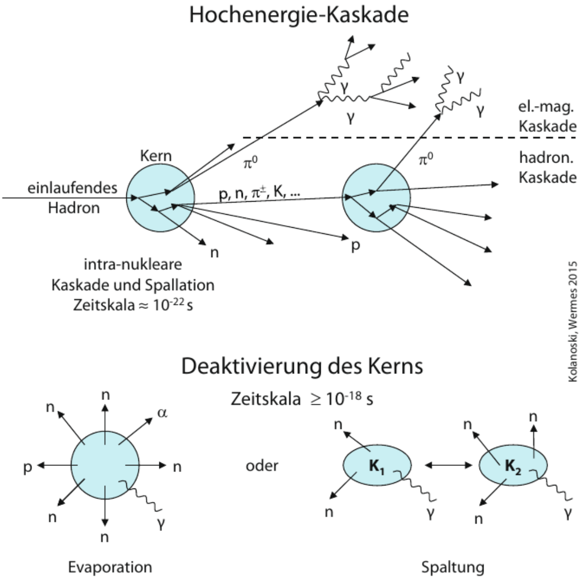
- Photoelectric effect in the energy range E_γ ~keV.
- Compton scattering in the energy range E_γ ~MeV.
- Electron-positron pair production for $E_\gamma \gtrsim 10$ MeV.

Formation of so-called electromagnetic showers in the traversed material.
Detailed examination in the next lecture.



Consequence for high-energy photon:

- After traveling a distance of $n \cdot X_0$: 2^n particles with energy $E_n \approx \frac{E_\gamma}{2^n}$.
- Cascade (shower) ends when $E_n = E_k$: $n = \frac{\ln \frac{E_\gamma}{E_k}}{\ln 2}$.
- Length of the cascade: $n \cdot X_0 = X_0 \cdot \frac{\ln \frac{E_\gamma}{E_k}}{\ln 2}$.
- Transverse extent of the cascade independent of E_γ: $L_\perp \approx 4R_M = 4X_0 \frac{21.2 \text{ MeV}}{E_k}$.



Qualitatively similar behavior to electromagnetic showers:

- Shower length proportional to $\lambda_A \approx 35\ \mathrm{g\,cm^{-2}}\, \frac{A^{1/3}}{\rho} \gg X_0$.
- Transverse extent independent of the energy of the primary hadron: λ_A.