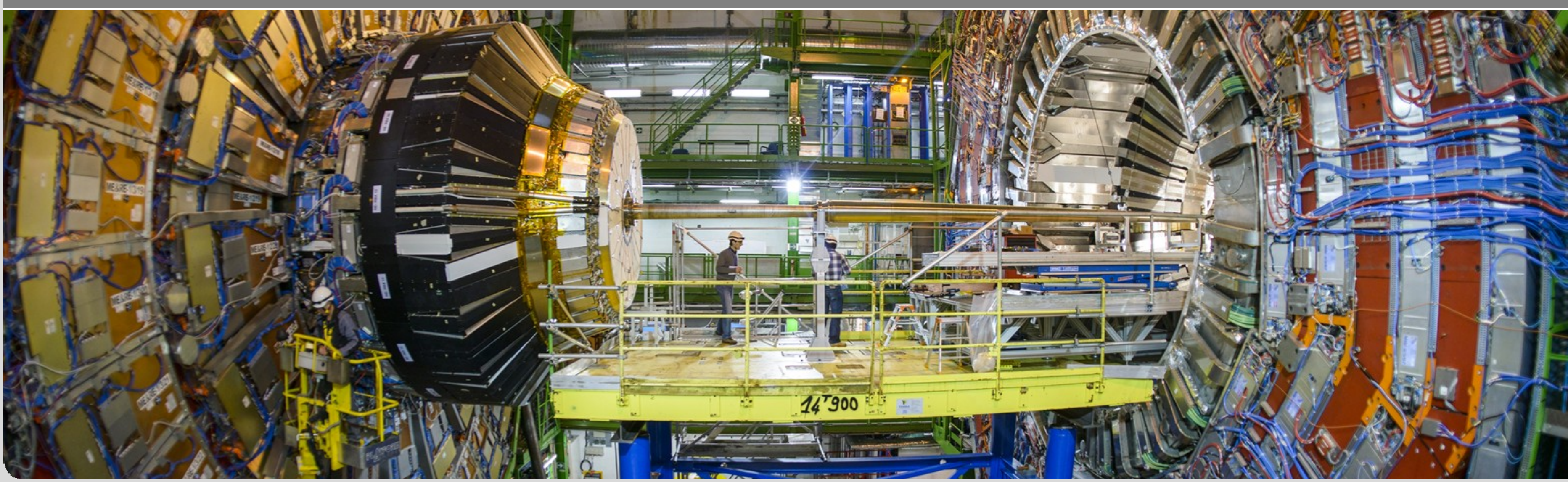
Statistical Methods used for Higgs Boson Searches

Roger Wolf

11. June 2015

INSTITUTE OF EXPERIMENTAL PARTICLE PHYSICS (IEKP) – PHYSICS FACULTY



Schedule for Today

3

Limits, p-values, significances.

2

Parameter estimates
(=fits).

1

Probability distributions
& Likelihood functions.

Schedule for Today

Walk through statistical methods that will appear in the next lectures:

- You will see all these methods acting in real life during the next lectures.
- To learn about the interiors of these methods check KIT lectures of Moderne Methoden der Datenanalyse.

3

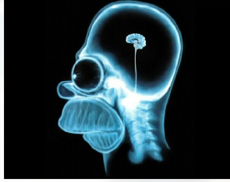
Limits, p-values, significances.

2

Parameter estimates (=fits).

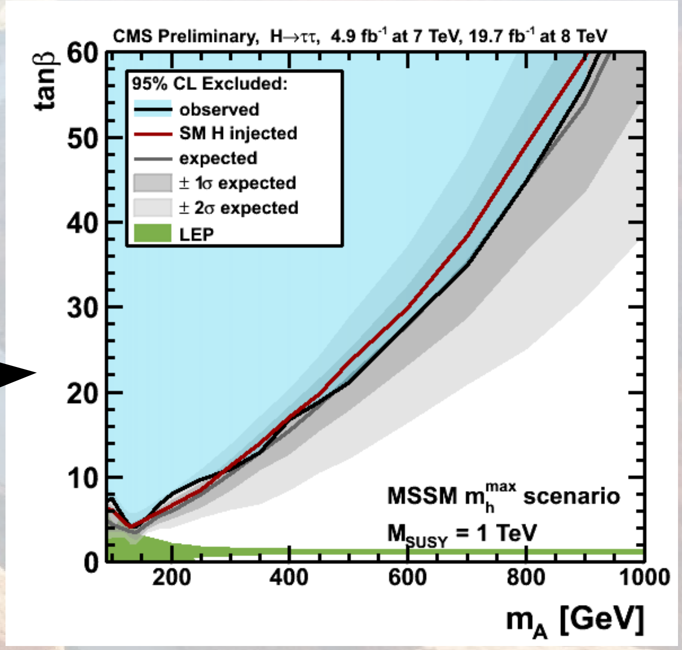
1

Probability distributions & Likelihood functions.



Outcome of the Day

- Relation between the Binomial, Gaussian & Poisson distribution.
- Relation between a minimal χ^2 fit and a Maximum Likelihood fit.
- Understand the meaning of this plot.
- Understand the meaning of a "3σ evidence" or a "5σ discovery".



Statistics ↔ Particle Physics

Theory:

- QM wave functions are interpreted as probability density functions.
- The Matrix Element, $\mathcal{S}_{fi}$,gives the probability to find final state f for given initial state i.
- Each of the statistical processes $pdf \rightarrow ME \rightarrow hadronization \rightarrow$ $energy\ loss\ in\ material \rightarrow digitization$ are statistically independent.
- Event by event simulation using Monte Carlo integration methods.

Statistics ↔ Particle Physics

Theory:

- QM wave functions are interpreted as probability density functions.
- The Matrix Element, $\mathcal{S}_{fi}$,gives the probability to find final state f for given initial state i.
- Each of the statistical processes *pdf* → *ME* → *hadronization* → *energy loss in material* → *digitization* are statistically independent.
- Event by event simulation using Monte Carlo integration methods.

Experiment:

- All measurements we do are derived from rate measurements.
- We record millions of trillions of particle collisions.
- Each of these collisions is independent from all the others.

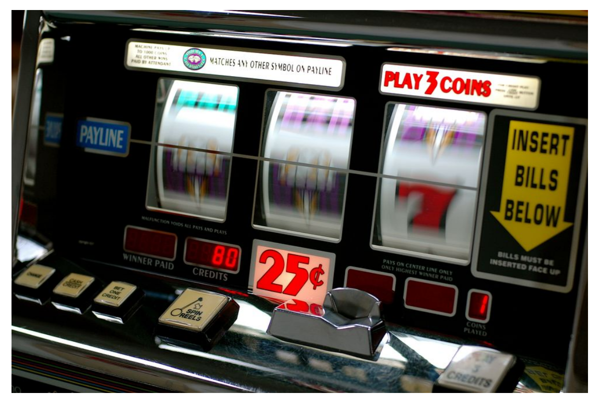
Statistics $\leftrightarrow$ Particle Physics

Theory:

- QM wave functions are interpreted as probability density functions.
- The Matrix Element, $\mathcal{S}_{fi}$,gives the probability to find final state f for given initial state i.
- Each of the statistical processes $pdf \rightarrow ME \rightarrow hadronization \rightarrow energy\ loss\ in\ material \rightarrow digitization$ are statistically independent.
- Event by event simulation using Monte Carlo integration methods.

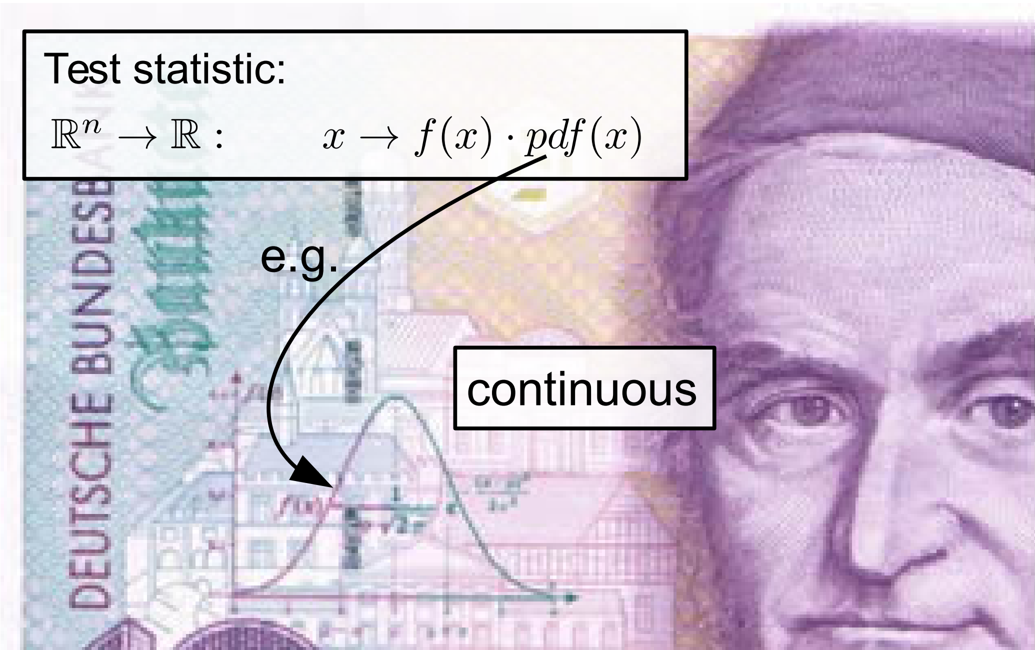
Experiment:

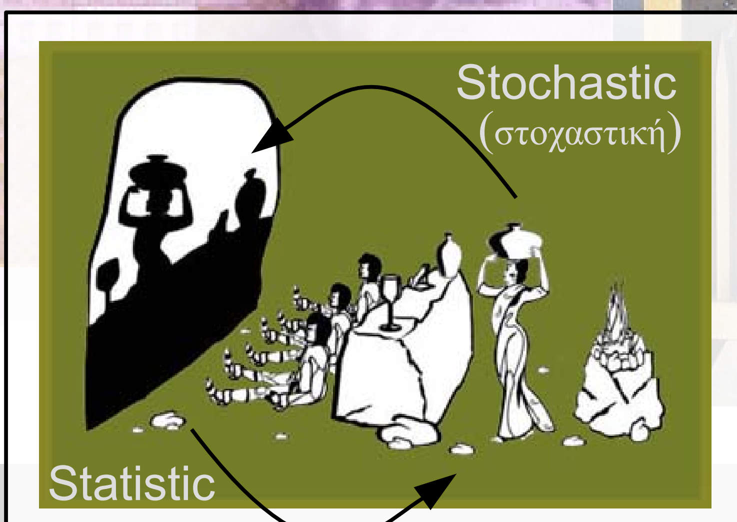
- All measurements we do are derived from rate measurements.
- We record millions of trillions of particle collisions.
- Each of these collisions is independent from all the others.



- Particle physics experiments are a perfect application for statistical methods.

Probability Distributions & Likelihood Functions





Characterization of Probability Distributions

Karlsruhe Institute of Technology

- Expectation:

$$E[x] = \int_{-\infty}^{\infty} x \cdot pdf(x)\mathrm{d}x = \mu$$

- Variance:

$$V[x] = \int_{-\infty}^{\infty} (x-\mu) \cdot pdf(x)\mathrm{d}x = \sigma^2$$

$$= E[(x - E[x])^2] = E[x^2 - 2xE[x] + E^2[x]] = E[x^2] - E^2[x]$$

- Covariance:

$$cov[x,y] = E[(x-\mu(x)(y-\mu(y))] = \int_{-\infty}^{\infty} x \cdot y \cdot pdf(x,y)\mathrm{d}x = E[xy] - \mu(x)\mu(y)$$

- Correlation coefficient:

$$\rho(x,y) = \frac{cov[x,y]}{\sqrt{V[x]V[y]}}$$

Probability Distributions

	Expectation:	Variance:
$$\mathcal{P}(k,n,p) = \binom{n}{k} p^k \cdot (1-p)^{n-k}$$ (Binomial distribution)	$\mu = np$	$\sigma^2 = np(1-p)$

Probability Distributions

	Expectation:	Variance:
$\mathcal{P}(k,n,p) = \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{1}{2}\left(\frac{k-np}{np(1-p)}\right)^2}$ (Gaussian distribution)	$\mu = np$	$\sigma^2 = np(1-p)$
$\mathcal{P}(k,n,p) = \binom{n}{k} p^k \cdot (1-p)^{n-k}$ (Binomial distribution)	$\mu = np$	$\sigma^2 = np(1-p)$

From the Binomial to the Gaussian distribution: $n \to \infty$, p fixed

Central limit theorem of de Moivre & Laplace.

Probability Distributions

	Expectation:	Variance:
$\mathcal{P}(k,n,p) = \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{1}{2}\left(\frac{k-np}{np(1-p)}\right)^2}$ (Gaussian distribution)	$\mu = np$	$\sigma^2 = np(1-p)$
$\mathcal{P}(k,n,p) = \binom{n}{k} p^k \cdot (1-p)^{n-k}$ (Binomial distribution)	$\mu = np$	$\sigma^2 = np(1-p)$
$\mathcal{P}(k,n,p) = \frac{(np)^k}{k!} e^{-np}$ (Poisson distribution)	$\mu = np$	$\sigma^2 = \mu = np$

$$\mathcal{P}(k,n,p) = \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{1}{2}\left(\frac{k-np}{np(1-p)}\right)^2}$$

(Gaussian distribution)

↑ $n \to \infty$, p fixed

Central limit theorem of de Moivre & Laplace.

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

(Binomial distribution)

↓ $n \to \infty$, np fixed

Will be shown on next slide.

$$\mathcal{P}(k,n,p) = \frac{(np)^k}{k!} e^{-np}$$

(Poisson distribution)

Probability Distributions

	Expectation:	Variance:
$\mathcal{P}(k,n,p) = \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{1}{2}\left(\frac{k-np}{np(1-p)}\right)^2}$ (Gaussian distribution)	$\mu = np$	$\sigma^2 = np(1-p)$
$\mathcal{P}(k,n,p) = \binom{n}{k} p^k \cdot (1-p)^{n-k}$ (Binomial distribution)	$\mu = np$	$\sigma^2 = np(1-p)$
$\mathcal{P}(k,n,p) = \frac{(np)^k}{k!} e^{-np}$ (Poisson distribution)	$\mu = np$	$\sigma^2 = \mu = np$

Binomial → Gaussian: $n \to \infty$, p fixed

Central limit theorem of de Moivre & Laplace.

Binomial → Poisson: $n \to \infty$, np fixed

Will be shown on next slide.

$\sigma^2 = \mu = np$: motivation for $\sqrt{k}$ uncertainty.

Binomial ↔ Poisson Distribution

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

$$= \frac{n(n-1)(n-2)\cdot\ldots\cdot(n-k+1)}{k!} \cdot \frac{\mu^k}{n^k} \cdot \frac{\left(1-\frac{\mu}{n}\right)^n}{\left(1-\frac{\mu}{n}\right)^k}$$

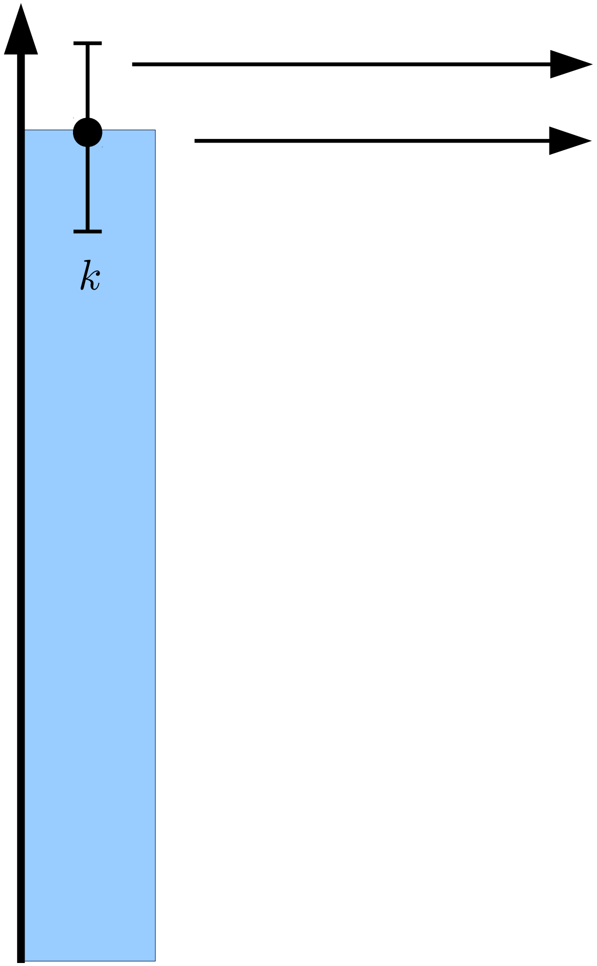
$$= \frac{1\cdot(1-\frac{1}{n})(1-\frac{2}{n})\cdot\ldots\cdot(1-\frac{k-1}{n})}{\left(1-\frac{\mu}{n}\right)^k} \cdot \frac{\mu^k}{k!} \cdot \left(1-\frac{\mu}{n}\right)^n$$

$$= \underbrace{\frac{1}{(1-\frac{\mu}{n})} \cdot \frac{(1-\frac{2}{n})}{(1-\frac{\mu}{n})} \cdot \frac{(1-\frac{2}{n})}{(1-\frac{\mu}{n})} \cdot \ldots \cdot \frac{(1-\frac{k-1}{n})}{(1-\frac{\mu}{n})}}_{\to 1} \cdot \frac{\mu^k}{k!} \cdot \underbrace{\left(1-\frac{\mu}{n}\right)^n}_{\to e^{-\mu}}$$

$$= \frac{\mu^k}{k!} e^{-\mu}$$

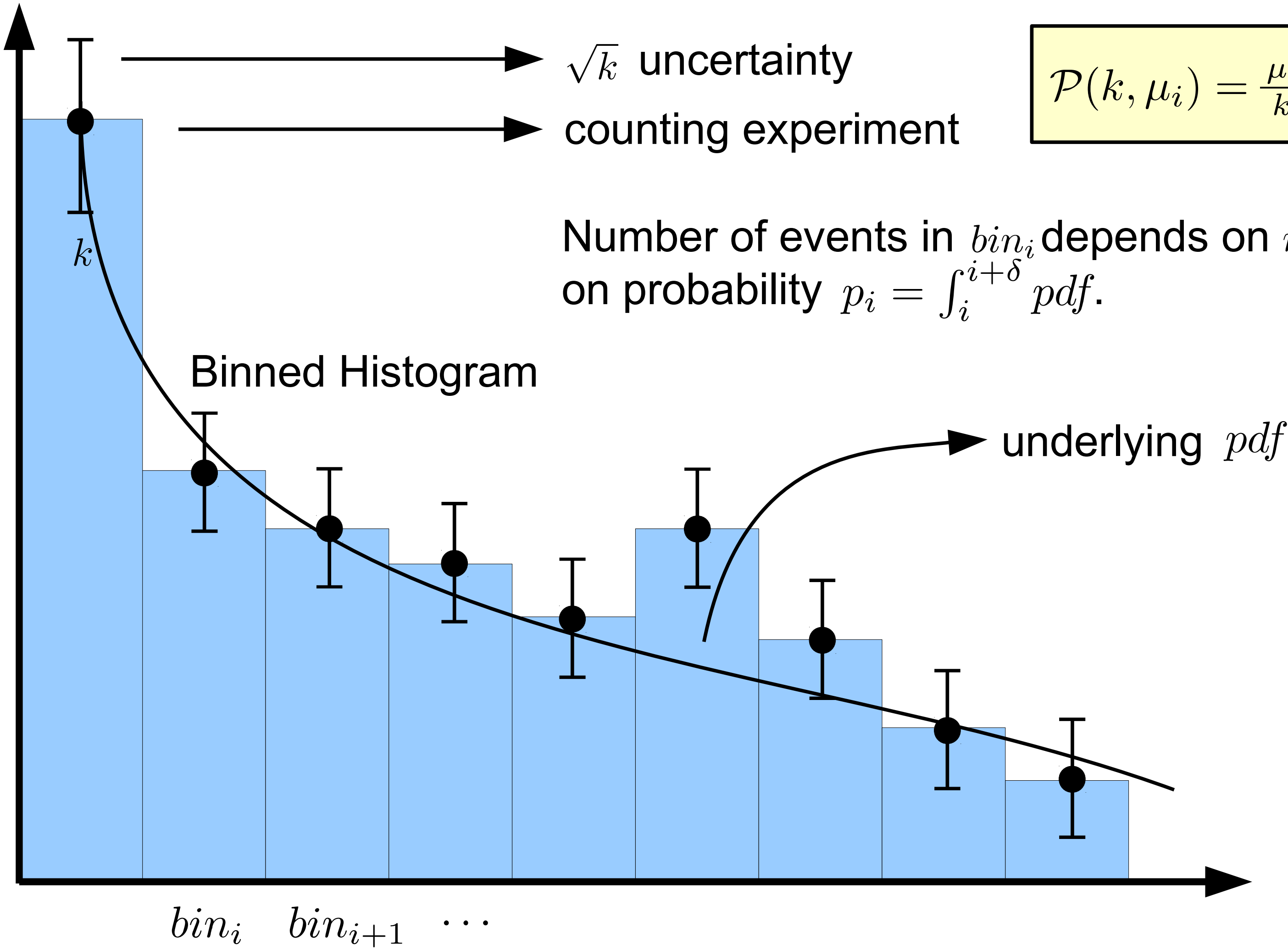
$$\mu = const, n \to \infty$$

Uncertainties on Counting Experiments



$$\mathcal{P}(k,\mu_i) = \frac{\mu_i^k}{k!}\,e^{-\mu_i}$$

Uncertainties on Counting Experiments



$$\mathcal{P}(k, \mu_i) = \frac{\mu_i^k}{k!} e^{-\mu_i}$$

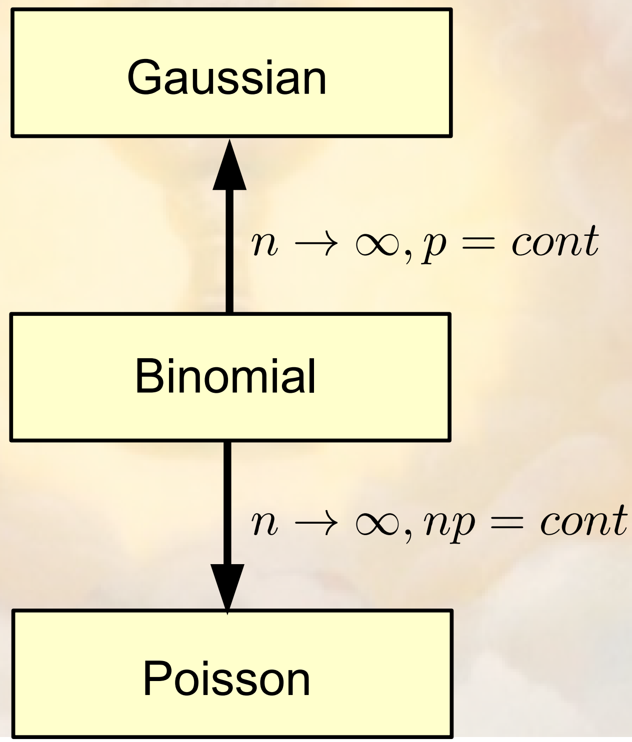
Number of events in bin_i depends on n and on probability $p_i = \int_i^{i+\delta} pdf$.

Relations between Probability Distributions

$$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Central Limit Theorem:

Random variable variable made up of a sum of many single measurements.



Look for something that is very rare very often.

Relations between Probability Distributions

$$\frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2}$$

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

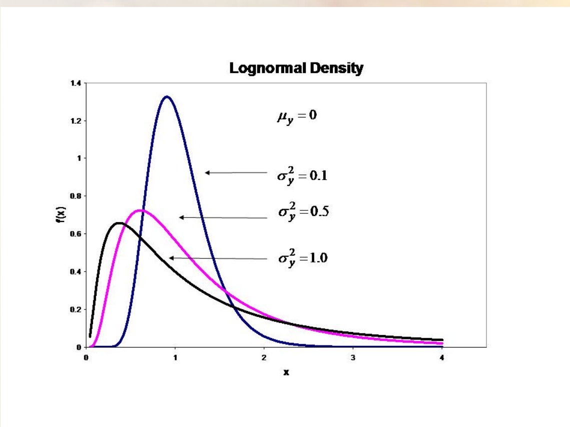
Central Limit Theorem:

Random variable variable made up of a sum of many single measurements.

Log-normal

log

Random variable variable made up of a product of many single measurements.



Gaussian

$n \to \infty, p = cont$

Binomial

$n \to \infty, np = cont$

Poisson

Look for something that is very rare very often.

Relations between Probability Distributions

Karlsruhe Institute of Technology

$$\frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2}$$

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

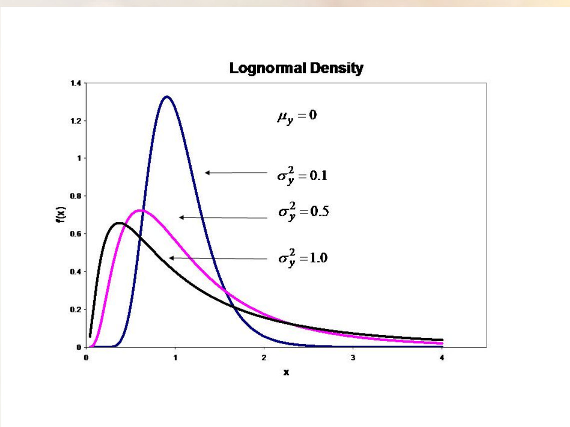
$$-\ln\left(\sqrt{2\pi}\sigma\right) - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$$

Log-normal ← log — **Gaussian** — log → **χ^2 Distribution**

Central Limit Theorem:

Random variable variable made up of a sum of many single measurements.

Random variable variable made up of a product of many single measurements.



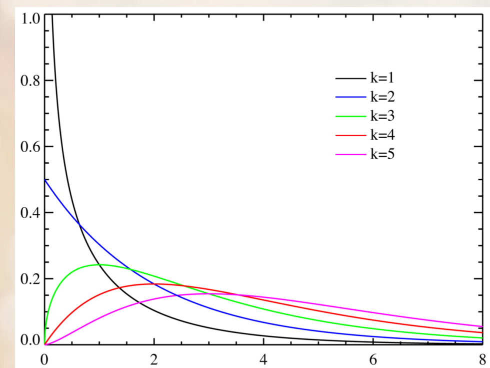
Gaussian

↑ $n \to \infty, p = cont$

Binomial

↓ $n \to \infty, np = cont$

Poisson



What does the parameter k correspond to in the χ^2 distributions?

Look for something that is very rare very often.

Relations between Probability Distributions

$\frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2}$

$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

$-\ln\left(\sqrt{2\pi}\sigma\right) - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$

Log-normal

χ^2 Distribution

Central Limit Theorem:

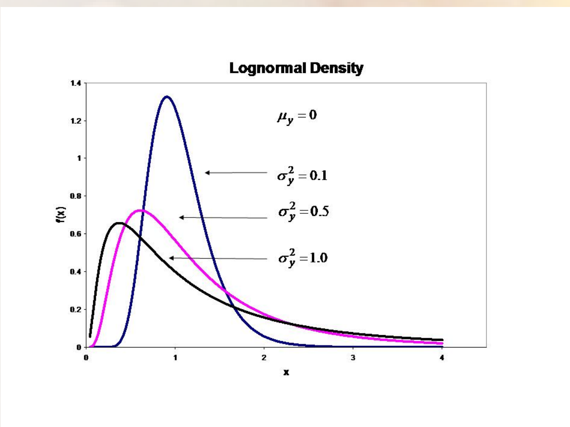
Random variable variable made up of a sum of many single measurements.

log

log

Gaussian

Random variable variable made up of a product of many single measurements.

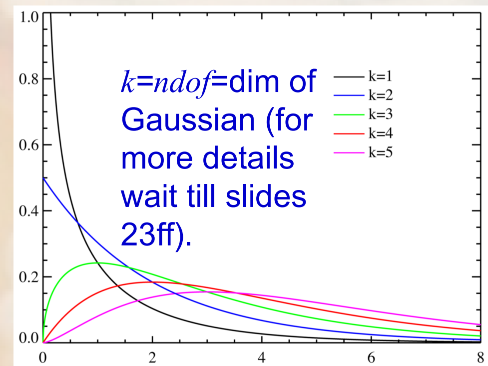


$n \to \infty, p = cont$

Binomial

$n \to \infty, np = cont$

Poisson



What does the parameter k correspond to in the χ^2 distributions?

Look for something that is very rare very often.

Likelihood Functions

- **Problem**: truth is not known!
- Deduce "truth" from measurements (usually in terms of models).
- Likeliness of a model to be true quantified by *likelihood function* $\mathcal{L}(\{k_i\}, \{\kappa_j\})$.

model parameters.

measured number of events (e.g. in bins i).

Likelihood Functions

Karlsruhe Institute of Technology

- **Problem**: truth is not known!
- Deduce "truth" from measurements (usually in terms of models).
- Likeliness of a model to be true quantified by *likelihood function* $\mathcal{L}(\{k_i\}, \{\kappa_j\})$.

  $\{\kappa_j\}$ → model parameters.

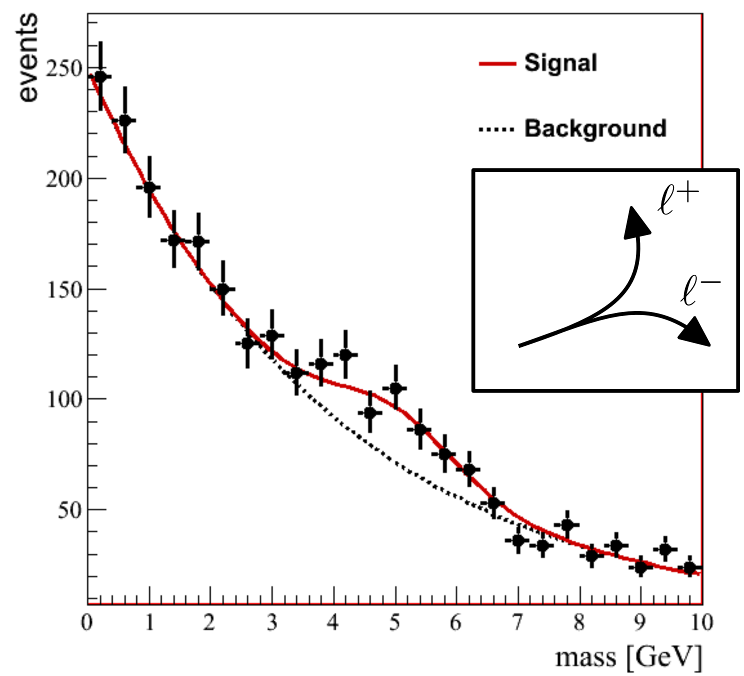
  $\{k_i\}$ → measured number of events (e.g. in bins i).

- Example:
  signal on top of known background in a binned histogram:

$$\mathcal{L}(\{k_i\}, \{\kappa_j\}) = \prod_i \underbrace{\mathcal{P}(k_i, \mu_i(\kappa_j))}$$

Product of *pdf*s for each bin (Poisson).

$$\mu_i(\kappa_j) = \underbrace{\kappa_0 \cdot e^{-\kappa_1 x_i}}_{\text{background}} + \underbrace{\kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}}_{\text{signal}}$$



Likelihood Functions

- **Problem**: truth is not known!
- Deduce "truth" from measurements (usually in terms of models).
- Likeliness of a model to be true quantified by *likelihood function* $\mathcal{L}(\{k_i\},\{\kappa_j\})$.

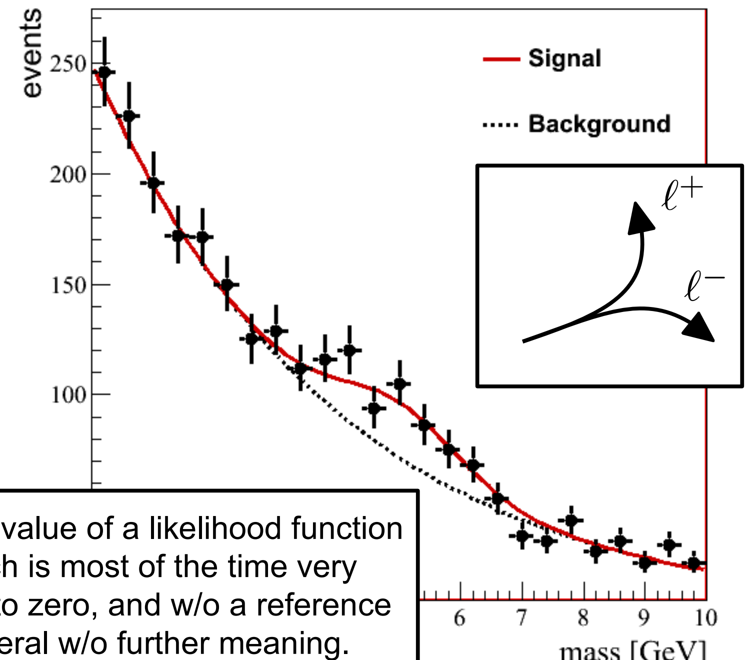
$\{\kappa_j\}$ → model parameters.

$\{k_i\}$ → measured number of events (e.g. in bins i).

- Example:
  signal on top of known background in a binned histogram:

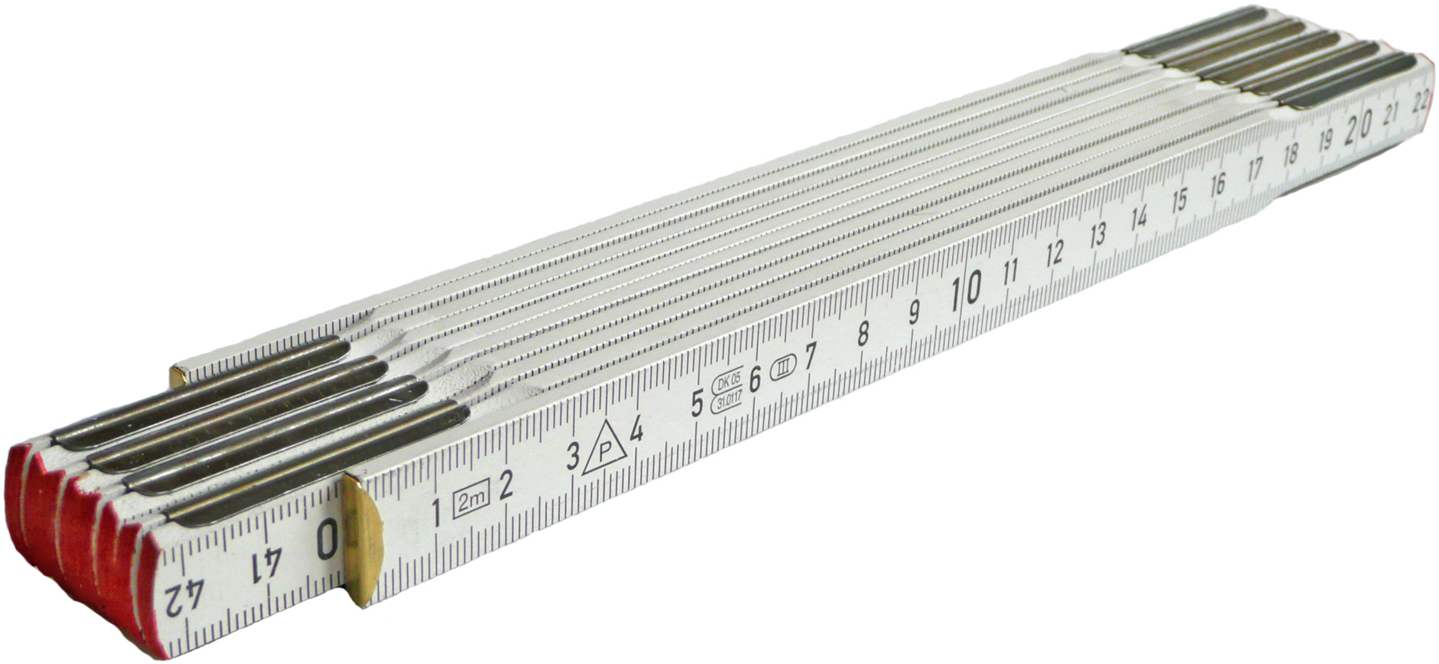
$$\mathcal{L}(\{k_i\},\{\kappa_j\}) = \prod_i \underbrace{\mathcal{P}(k_i, \mu_i(\kappa_j))}_{\text{Product of } pdf\text{s for each bin (Poisson).}}$$

$$\mu_i(\kappa_j) = \underbrace{\kappa_0 \cdot e^{-\kappa_1 x_i}}_{\text{background}} + \underbrace{\kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}}_{\text{signal}}$$



NB: a value of a likelihood function as such is most of the time very close to zero, and w/o a reference in general w/o further meaning.

Parameter Estimates



Parameter Estimates

- **Problem**: find most probable parameter(s) κ_j of a given model.
- Usually minimization of negative log likelihood function (NLL):
  - log is a monotonic function and very often numerically easier to handle.
  - e.g. products of probability distributions turn into sums.
  - e.g. if probability distributions are Gaussians NLL turns into χ^2 minimization:

Parameter Estimates

- **Problem**: find most probable parameter(s) κ_j of a given model.
- Usually minimization of negative log likelihood function (NLL):
  - log is a monotonic function and very often numerically easier to handle.
  - e.g. products of probability distributions turn into sums.
  - e.g. if probability distributions are Gaussians NLL turns into χ^2 minimization:

Clear to everybody?

Parameter Estimates

- **Problem**: find most probable parameter(s) κ_j of a given model.
- Usually minimization of negative log likelihood function (NLL):
  - log is a monotonic function and very often numerically easier to handle.
  - e.g. products of probability distributions turn into sums.
  - e.g. if probability distributions are Gaussians NLL turns into χ^2 minimization:

$$NLL = -\ln\left(\prod_i e^{-\frac{1}{2}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2}\right) \propto \sum_i \left(\frac{x_i-\mu_i}{\sigma_i}\right)^2$$

Clear to everybody?

Number of x_i's determines dimension of the Gaussian distribution.

Parameter Estimates

- **Problem**: find most probable parameter(s) κ_j of a given model.
- Usually minimization of negative log likelihood function (NLL):
  - log is a monotonic function and very often numerically easier to handle.
  - e.g. products of probability distributions turn into sums.
  - e.g. if probability distributions are Gaussians NLL turns into χ^2 minimization:

Clear to everybody?

$$NLL = -\ln\left(\prod_i e^{-\frac{1}{2}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2}\right) \propto \sum_i \left(\frac{x_i-\mu_i}{\sigma_i}\right)^2$$

Number of x_i's determines dimension of the Gaussian distribution.

- The minimization usually performed:
  - analytically (like in an optimization exercise at school).
  - numerically (usually the more general solution).
  - by scan of the NLL (for sure the most robust method, but can be time consuming).

Parameter(s) of Interest (POI)

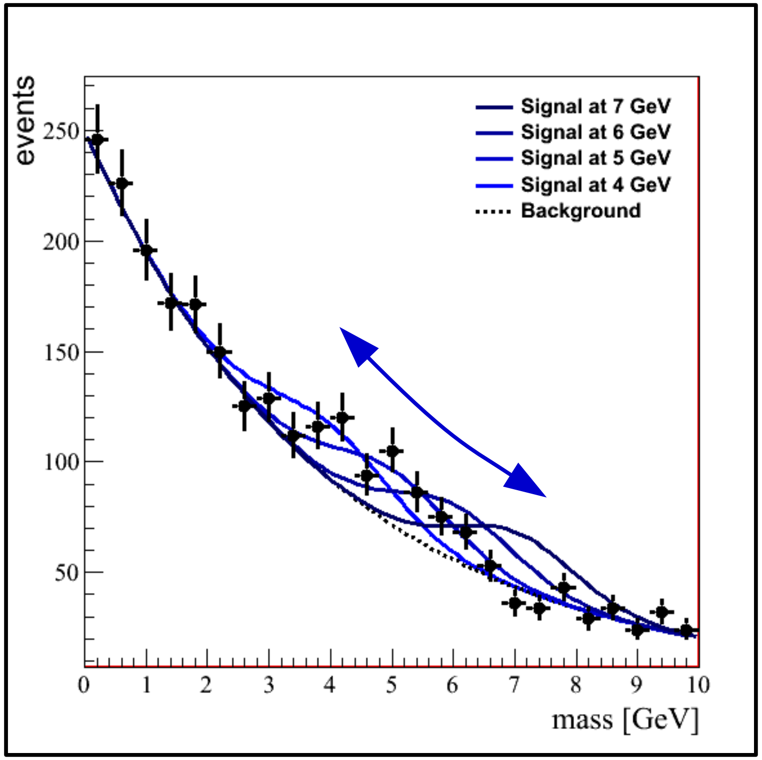
- Each case/problem defines its own *parameter(s) of interest* (POI's):
  - POI could be the mass κ_3.

- Example:
  signal on top of known background in a binned histogram:

$$\mathcal{L}(\{k_i\},\{\kappa_j\}) = \underbrace{\prod_i \mathcal{P}(k_i,\mu_i(\kappa_j))}$$

Product of *pdf*s for each bin (Poisson).

$$\mu_i(\kappa_j) = \underbrace{\kappa_0 \cdot e^{-\kappa_1 x_i}}_{\text{background}} + \underbrace{\kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}}_{\text{signal}}$$



Parameter(s) of Interest (POI)

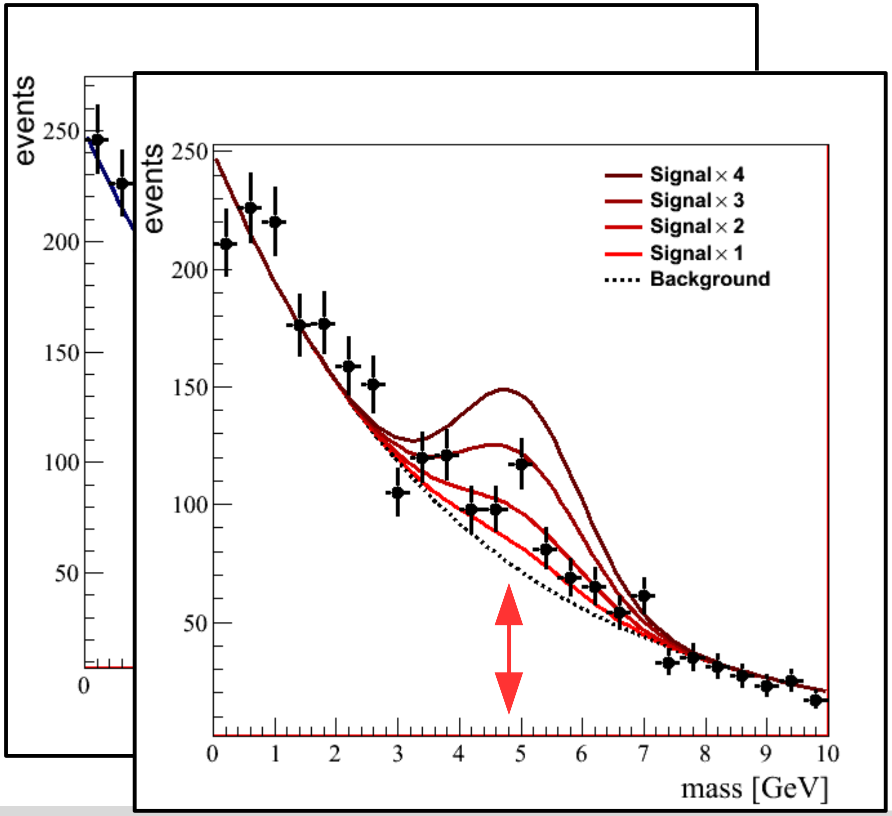
- Each case/problem defines its own *parameter(s) of interest* (POI's):
  - POI could be the mass κ_3.
  - In our case POI usually is the signal strength κ_2 for a fixed value for κ_3.

- Example:
  signal on top of known background in a binned histogram:

$$\mathcal{L}(\{k_i\},\{\kappa_j\}) = \prod_i \underbrace{\mathcal{P}(k_i,\mu_i(\kappa_j))}$$

Product of *pdf*s for each bin (Poisson).

$$\mu_i(\kappa_j) = \underbrace{\kappa_0 \cdot e^{-\kappa_1 x_i}}_{\text{background}} + \underbrace{\kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}}_{\text{signal}}$$



Systematic Uncertainties

- Systematic uncertainties are usually incorporated as *nuisance parameters*:
  - Example: assume background normalization κ_0 is not precisely known, but with an uncertainty $\sigma(\kappa_0)$:

$$\mu_i(\kappa_j) = \mathcal{P}'(\tilde{\kappa}_0, \kappa_0, \sigma(\kappa_0)) \cdot e^{-\kappa_1 x_i} + \kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}$$

uncertainty

expected value

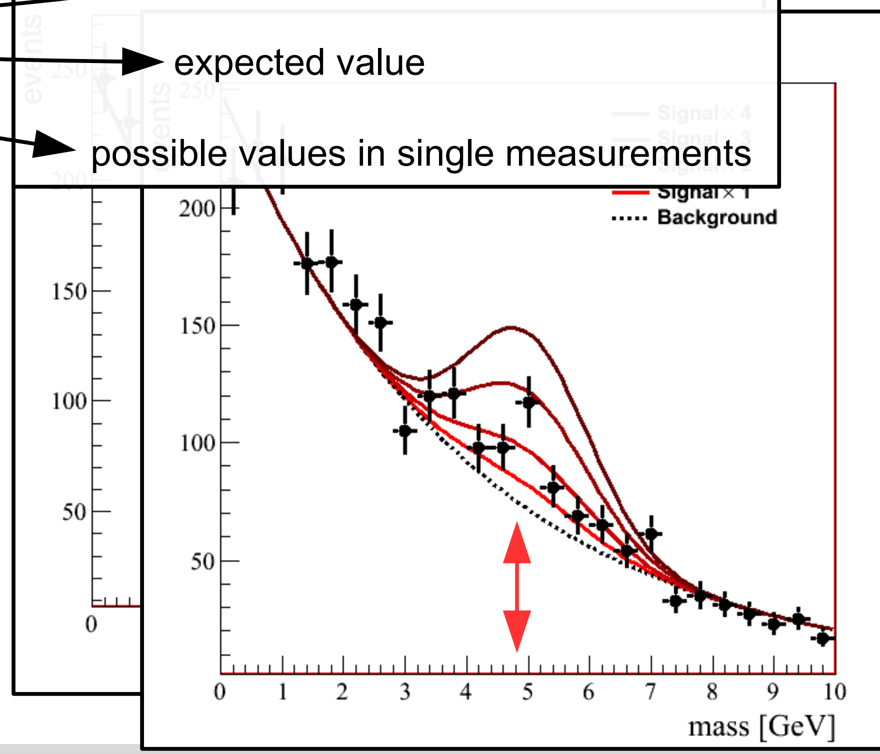
possible values in single measurements

- Example:
  signal on top of known background in a binned histogram:

$$\mathcal{L}(\{k_i\},\{\kappa_j\}) = \prod_i \underbrace{\mathcal{P}(k_i, \mu_i(\kappa_j))}$$

Product of *pdf*s for each bin (Poisson).

$$\mu_i(\kappa_j) = \underbrace{\kappa_0 \cdot e^{-\kappa_1 x_i}}_{\text{background}} + \underbrace{\kappa_2 \cdot e^{-(\kappa_3 - x_i)^2}}_{\text{signal}}$$



Hypothesis Tests

Hypothesis Separation

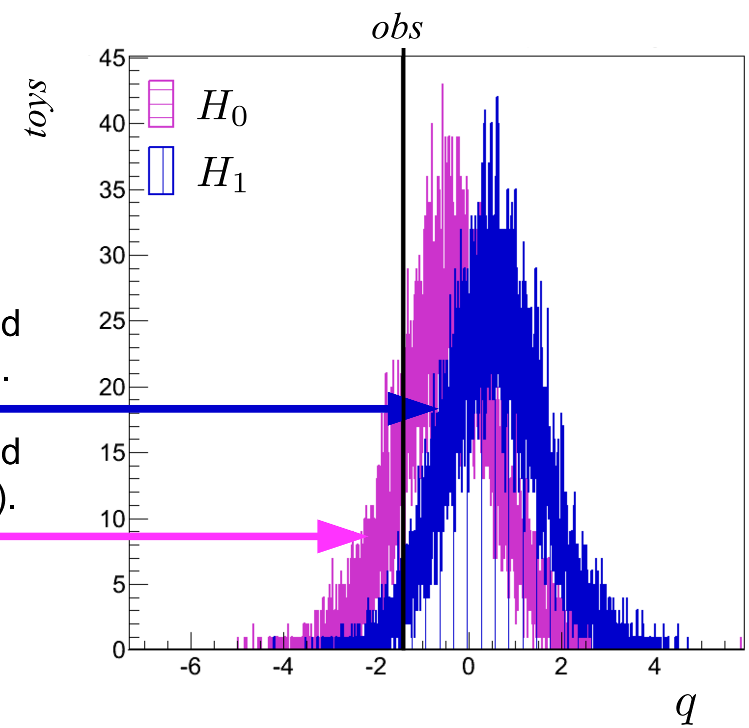
- Start with two alternative hypotheses H_0 & H_1 .
- Define a test statistic $q : \mathbb{R}^n \rightarrow \mathbb{R}$ that can distinguish these two hypotheses.
- The test statistic with the best separation power is the likelihood ratio (LR):

$$q = \ln\left(\frac{\mathcal{L}(\text{obs})|H_1}{\mathcal{L}(\text{obs})|H_0}\right)$$

- q can be calculated for the observation (obs), for the expectation for H_0 and for the expectation for H_1 :
  - Observed is a single value (outcome of measurement).
  - Expectation = mean value with uncertainties (based on toy measurements).

pdf from toys based on H_1 (usually sig).

pdf from toys based on H_0 (usually BG).



Hypothesis Separation

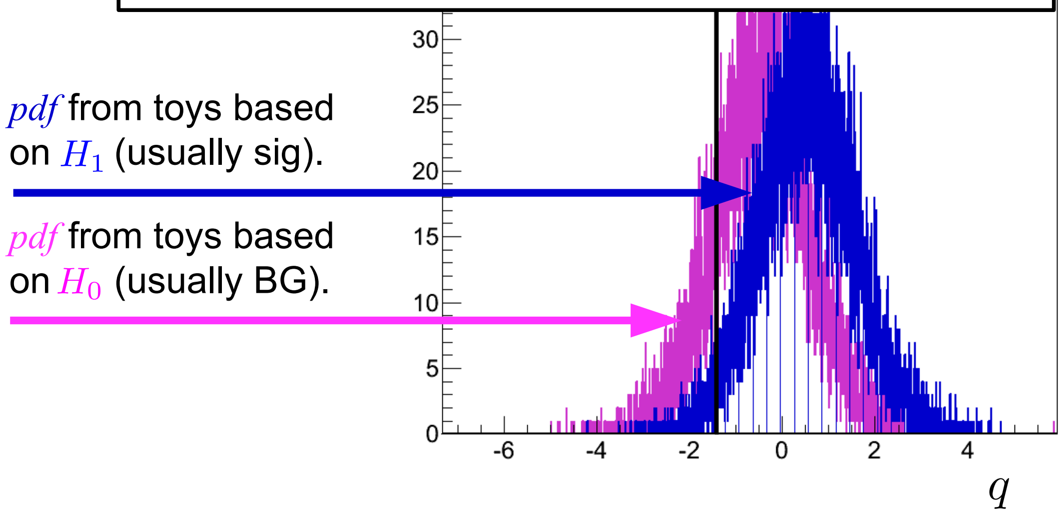
- Start with two alternative hypotheses H_0 & H_1 .
- Define a test statistic $q : \mathbb{R}^n \rightarrow \mathbb{R}$ tha
- The test statistic with the best separ

$$q = \ln\left(\frac{\mathcal{L}(\mathrm{obs})|H_1}{\mathcal{L}(\mathrm{obs})|H_0}\right)$$

- q can be calculated for the observati (*obs*), for the expectation for H_0 and t the expectation for H_1 :
  - Observed is a single value (outcome of measurement).
  - Expectation = mean value with uncertainties (based on toy measurements).

Signal on top of background!

Sorry! No price...



Test Statistics (LEP)

- Start with two alternative hypotheses H_0 & H_1.
- Define a test statistic $q : \mathbb{R}^n \to \mathbb{R}$ that can distinguish these two hypotheses.
- The test statistic with the best separation power is the likelihood ratio (LR):

$$\mathcal{L}(n|\,b(\kappa_j)) = \prod_i \mathcal{P}(n_i|b_i(\kappa_j)) \times \prod_j \mathcal{C}(\kappa_j|\tilde{\kappa}_j)$$

$$\mathcal{L}(n|\,\mu s(\kappa_j) + b(\kappa_j)) = \prod_i \mathcal{P}(n_i|\mu s_i(\kappa_j) + b_i(\kappa_j)) \times \prod_j \mathcal{C}(\kappa_j|\tilde{\kappa}_j)$$

$$q_\mu = -2\ln\left(\frac{\mathcal{L}(\,n|\mu s+b)}{\mathcal{L}(\,n|b)}\right), \quad 0 \le \mu$$

nuisance parameters $\tilde{\kappa}_j$ integrated out (by throwing toys → MC method) before evaluation of q_μ (→marginalization).

Test Statistics (Tevatron)

- Start with two alternative hypotheses H_0 & H_1 .
- Define a test statistic $q : \mathbb{R}^n \to \mathbb{R}$ that can distinguish these two hypotheses.
- The test statistic with the best separation power is the likelihood ratio (LR):

$$\mathcal{L}(n|\,b(\kappa_j)) = \prod_i \mathcal{P}(n_i|b_i(\kappa_j)) \times \prod_j \mathcal{C}(\kappa_j|\tilde{\kappa}_j)$$

$$\mathcal{L}(n|\,\mu s(\kappa_j) + b(\kappa_j)) = \prod_i \mathcal{P}(n_i|\mu s_i(\kappa_j) + b_i(\kappa_j)) \times \prod_j \mathcal{C}(\kappa_j|\tilde{\kappa}_j)$$

$$q_\mu = -2\ln\left(\frac{\mathcal{L}(\,n|\mu s(\hat{\kappa}_\mu)+b(\hat{\kappa}_\mu))}{\mathcal{L}(\,n|b(\hat{\kappa}_{\mu=0}))}\right),\quad 0 \leq \mu$$

nominator maximized for given μ before marginalization. Denominator for $\mu = 0$. Better estimates on nuisance parameters. Reduces uncertainties on nuisance parameters.

Test Statistics (LHC)

- Start with two alternative hypotheses H_0 & H_1.
- Define a test statistic $q : \mathbb{R}^n \rightarrow \mathbb{R}$ that can distinguish these two hypotheses.
- The test statistic with the best separation power is the likelihood ratio (LR):

$$\mathcal{L}(n|\, b(\kappa_j)) = \prod_i \mathcal{P}(n_i|b_i(\kappa_j)) \qquad j$$

$$\mathcal{L}(n|\, \mu s(\kappa_j) + b(\kappa_j)) = \prod_i \mathcal{P}(n_i|\mu s_i(\kappa_j) + b_i(\kappa_j)) \qquad j$$

$$q_\mu = \ln\left(\frac{\mathcal{L}(n|\mu s(\hat{\kappa}_\mu)+b(\hat{\kappa}_\mu))}{\mathcal{L}(n|\hat{\mu} s(\hat{\kappa}_{\hat{\mu}})+b(\hat{\kappa}_{\hat{\mu}}))}\right), \quad 0 \leq \hat{\mu} \leq \mu$$

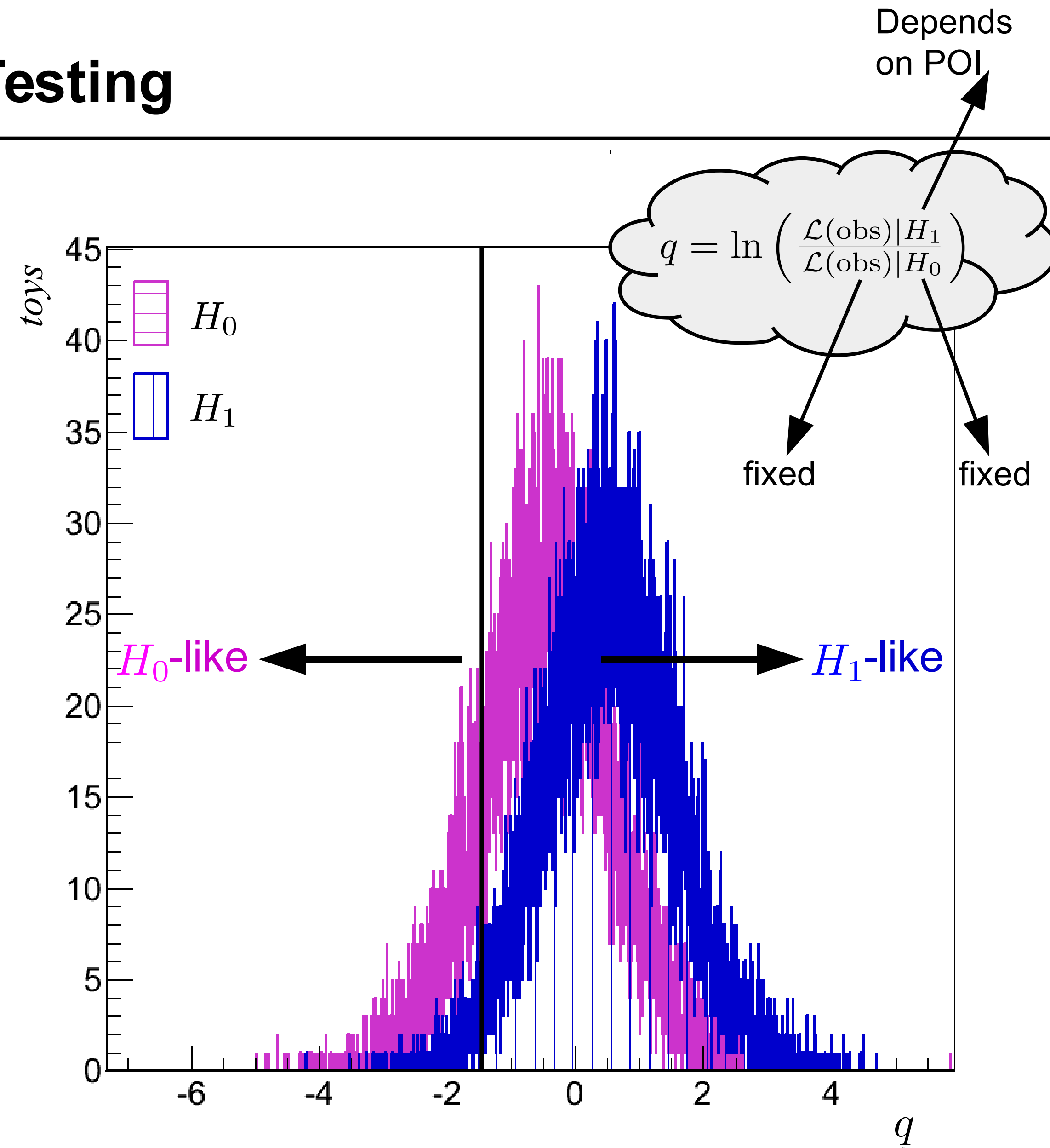
nominator maximized for given μ before marginalization. For the denominator a global maximum is searched for at $\hat{\mu}$. In addition allows use of asymptotic formulas (→ no need for toys).

Classical Hypothesis Testing

- Classical hypothesis test interested in probability to observe $q_{\rm obs}$ given that H_0 or H_1 is true:

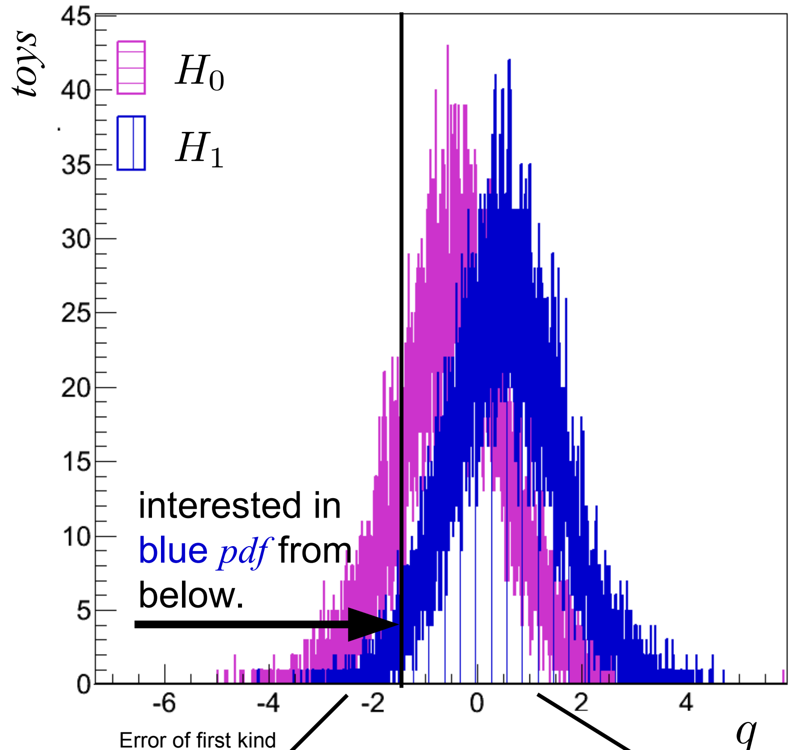
$q \leq q_{\rm obs}\|_{H_1}$	$q_{\rm obs}\|_{H_1} \leq q$
$q \leq q_{\rm obs}\|_{H_0}$	$q_{\rm obs}\|_{H_0} \leq q$
$q_{\rm obs}$ defines upper bound	$q_{\rm obs}$ defines lower bound

- We are usually interested in upper limits on the signal strength μ ($\rightarrow$ lower bound $q_{\rm obs}|_{H_1} \leq q$).



95% CL Upper Limits

- Our *pdf*'s usually depend on another parameter, which is the actual *POI* (μ in SM, $\tan\beta$ in MSSM case).
- Traditionally we set 95% CL upper limits on this POI.

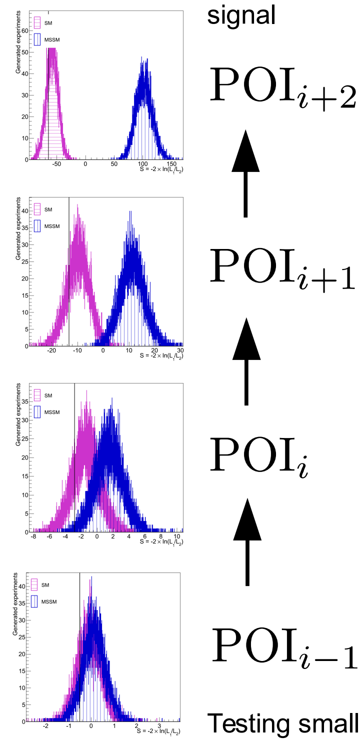


Error of first kind
(keep ≤ 5%)

Signal looks more H_0-like
(→excluded by mistake)

In ≥ 95% of all cases/toys the signal would look more signal like than observed ($q \geq q_{\mathrm{obs}}$).

Testing large signal



Testing small signal

- *pdf*'s move apart from each other.
- The more separate the *pdf*'s are the more H_0 & H_1 are distinguishable.
- Find POI_i for which:

$$\mathcal{I}_{\mathrm{POI}} = \int_{-\infty}^{q_{\mathrm{obs}}} pdf = 0.05$$

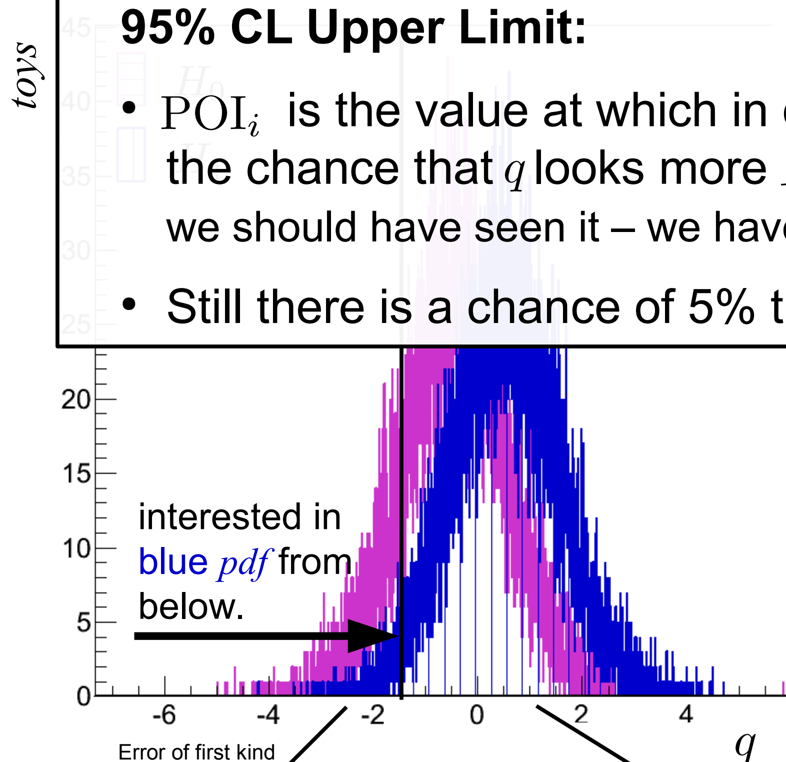
for this POI_i in 95% of all toys $q \geq q_{\mathrm{obs}}$.

95% CL Upper Limits

- Our *pdf*'s usually depend on another parameter, which is the actual *POI* (μ in SM, $\tan\beta$ in MSSM case).
- Traditionally we set 95% CL upper limits on this POI.

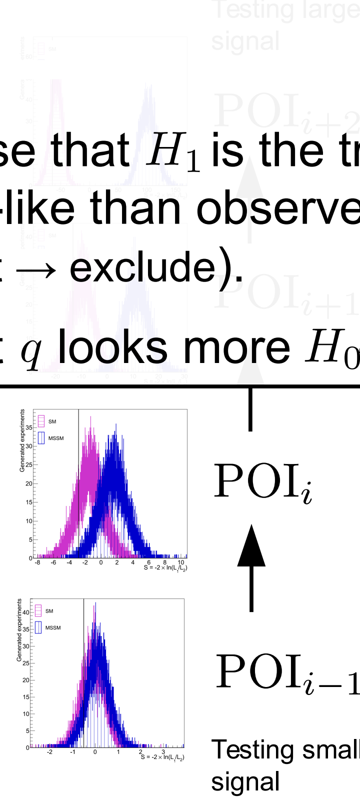
95% CL Upper Limit:

- POI_i is the value at which in case that H_1 is the true hypothesis the chance that q looks more H_1-like than observed is 95% (i.e. we should have seen it – we haven't → exclude).
- Still there is a chance of 5% that q looks more H_0-like.



Signal looks more H_0-like (→excluded by mistake)

In ≥ 95% of all cases/toys the signal would look more signal like than observed ($q \geq q_{\mathrm{obs}}$).



- *pdf*'s move apart from each other
- The more separate the *pdf*'s are the more H_0 & H_1 are distinguishable.
- Find POI_i for which:

  $$\mathcal{I}_{\mathrm{POI}} = \int_{-\infty}^{q_{\mathrm{obs}}} pdf = 0.05$$

  for this POI_i in 95% of all toys $q \geq q_{\mathrm{obs}}$.

95% CL Upper Limits

Karlsruhe Institute of Technology

- Our *pdf*'s usually depend on another parameter, which is the actual *POI* (μ in SM, $\tan\beta$ in MSSM case).
- Traditionally we set 95% CL upper limits on this POI.

95% CL Upper Limit:

- POI_i is the value at which in case that H_1 is the true hypothesis the chance that q looks more H_1-like than observed is 95% (i.e. we should have seen it – we haven't → exclude).
- Still there is a chance of 5% that q looks more H_0-like.

- *pdf*'s move apart from each other
- The more separate the *pdf*'s are the more H_0 & H_1 are distinguishable.
- Find POI_i for which: $\mathcal{I}_{\mathrm{POI}} = \int_{-\infty}^{q_{\mathrm{obs}}} pdf = 0.05$ for this POI_i in 95% of all toys $q \geq q_{\mathrm{obs}}$.

- Assume our POI is μ: does the 90% CL upper limit on μ correspond to a higher or a lower value?

(keep ≤ 5%)

Signal looks more H_0-like (→excluded by mistake)

In ≥ 95% of all cases/toys the signal would look more signal like than observed ($q \geq q_{\mathrm{obs}}$).

95% CL Upper Limits

- Our *pdf*'s usually depend on another parameter, which is the actual *POI* (μ in SM, $\tan\beta$ in MSSM case).
- Traditionally we set 95% CL upper limits on this POI.

95% CL Upper Limit:

- POI_i is the value at which in case that H_1 is the true hypothesis the chance that q looks more H_1-like than observed is 95% (i.e. we should have seen it – we haven't → exclude).
- Still there is a chance of 5% that q looks more H_0-like.

art from

arate the
ore H_0 &
uishable.

- Find POI_i for which:

$df = 0.05$

n 95% of all

- Assume our POI is μ: does the 90% CL upper limit on μ correspond to a higher or a lower value ? → It's lower!

$\mu_{99\%}$ → 1%
$\mu_{95\%}$ → 5%
$\mu_{90\%}$ → 10%

} probability of q to be "more background like" than q_{obs}.

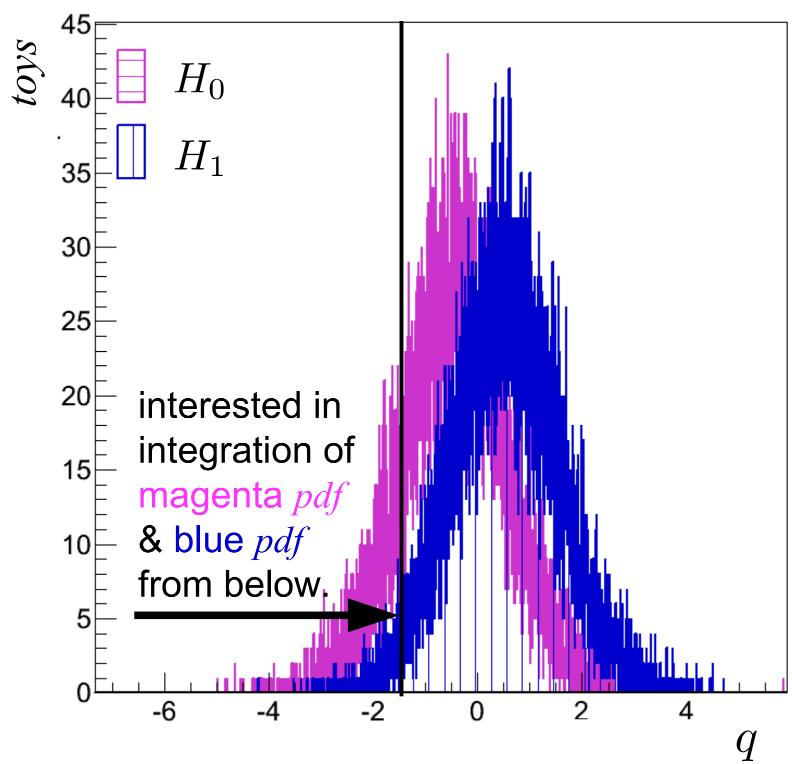
(keep ≤ 5%)

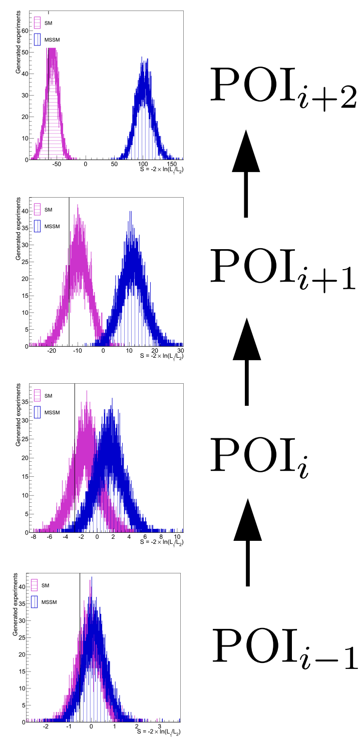
Signal looks more H_0-like (→excluded by mistake)

In ≥ 95% of all cases/toys the signal would look more signal like than observed ($q \geq q_{\mathrm{obs}}$).

CLs Limits

- In particle physics we set more conservative limits than this, following the CLs method:
- Assume H_1 to be signal+background and H_0 to be background only hypothesis.





$$\mathrm{CL(S+B)} = \int_{-\infty}^{q_{\mathrm{obs}}} pdf_{H_1}$$

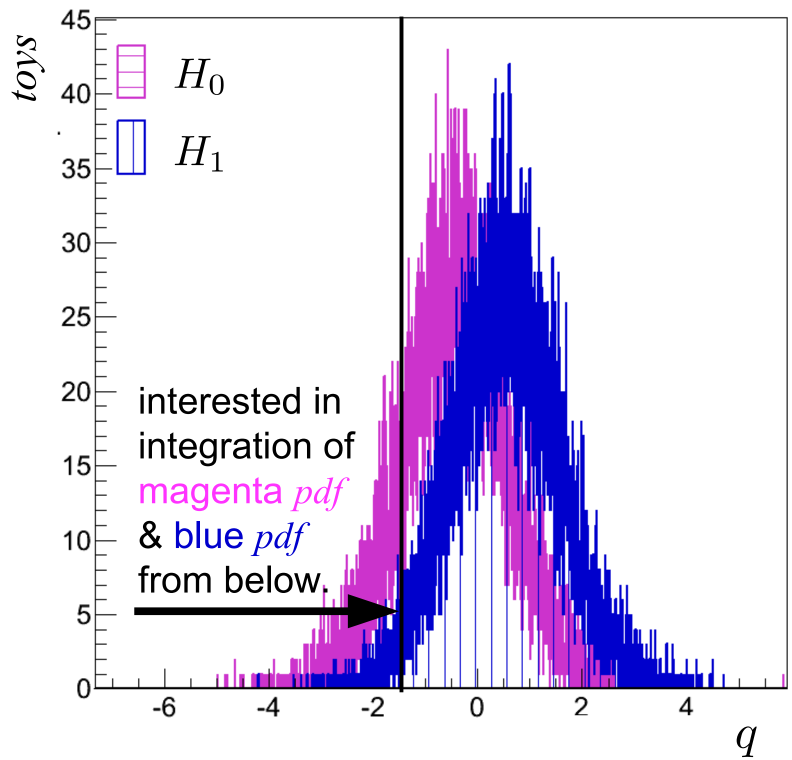
$$\mathrm{CL(B)} = \int_{-\infty}^{q_{\mathrm{obs}}} pdf_{H_0}$$

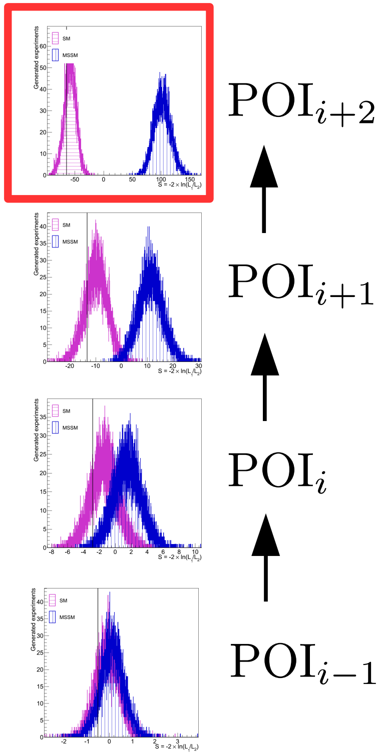
- Find POI_i for which:

$$\mathrm{CL_S} = \frac{\mathrm{CL(S+B)}}{\mathrm{CL(B)}} = 0.05$$

CLs Limits

- In particle physics we set more conservative limits than this, following the *CLs* method:
- Assume H_1 to be signal+background and H_0 to be background only hypothesis.





$$\mathrm{CL(S+B)} = \int_{-\infty}^{q_{\mathrm{obs}}} pdf_{H_1}$$

$$\mathrm{CL(B)} = \int_{-\infty}^{q_{\mathrm{obs}}} pdf_{H_0}$$

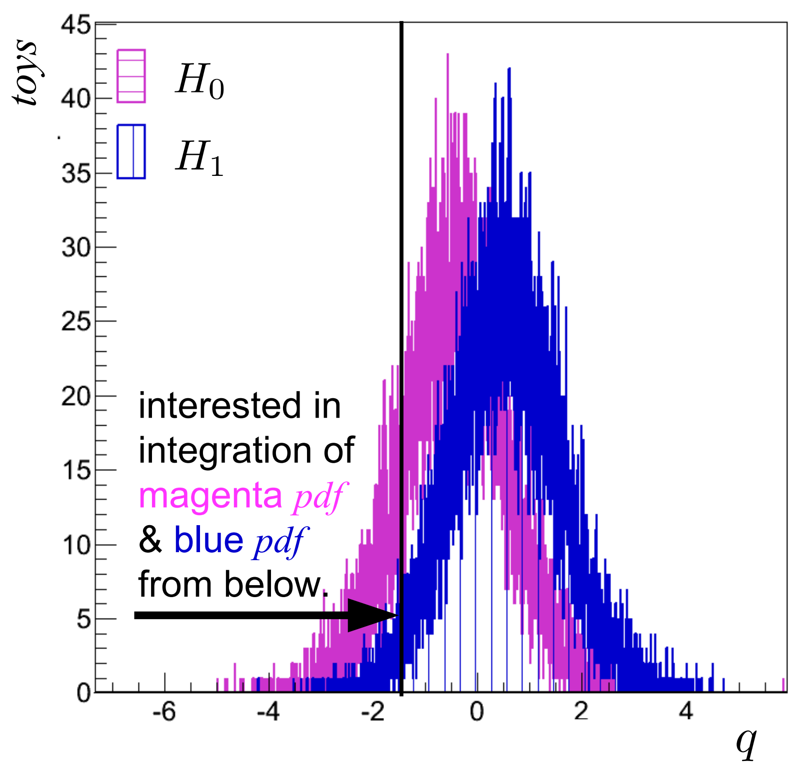
- Find POI_i for which:

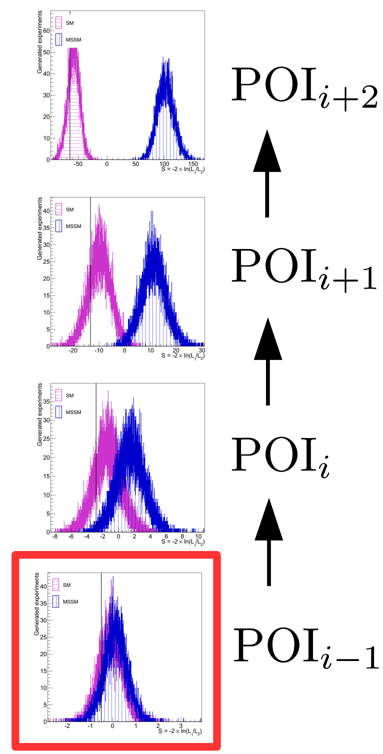
$$\mathrm{CL_S} = \frac{\mathrm{CL(S+B)}}{\mathrm{CL(B)}} = 0.05$$

- If H_0 & H_1 are clearly distinguishable $\mathrm{CL_S} \rightarrow \mathrm{CL(S+B)}$.

CLs Limits

- In particle physics we set more conservative limits than this, following the CLs method:
- Assume H_1 to be signal+background and H_0 to be background only hypothesis.





$\mathrm{CL(S+B)} = \int_{-\infty}^{q_{\mathrm{obs}}} pdf_{H_1}$

$\mathrm{CL(B)} \quad = \int_{-\infty}^{q_{\mathrm{obs}}} pdf_{H_0}$

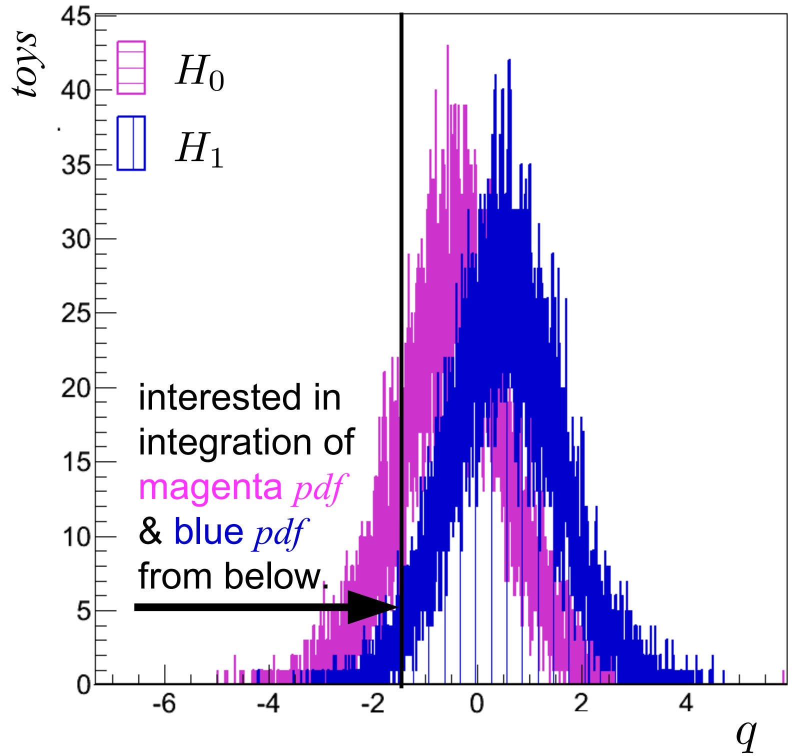
- Find POI_i for which:

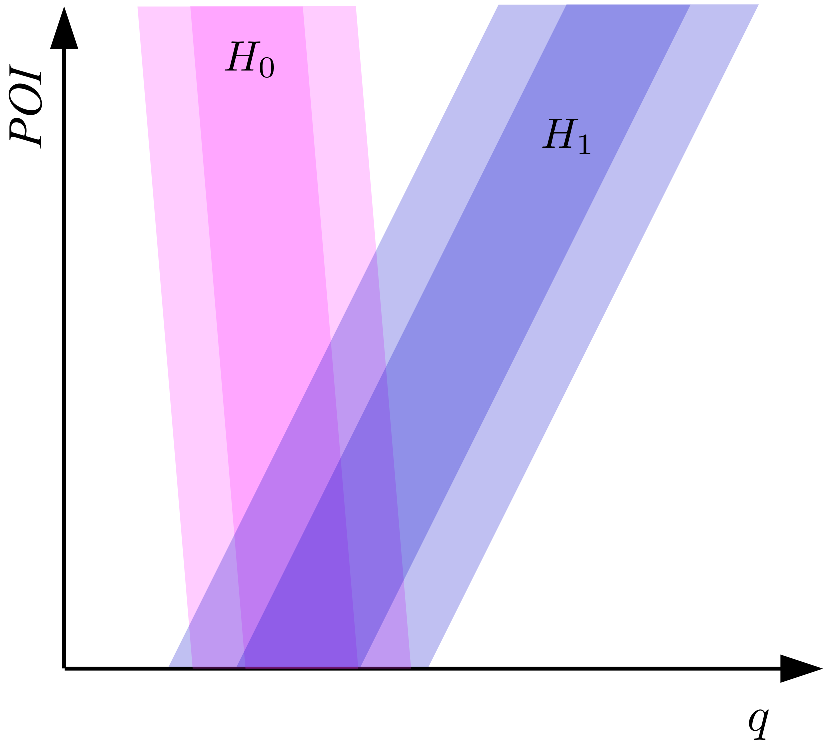
$$\mathrm{CL_S} = \frac{\mathrm{CL(S+B)}}{\mathrm{CL(B)}} = 0.05$$

- If H_0 & H_1 are clearly distinguishable $\mathrm{CL_S} \rightarrow \mathrm{CL(S+B)}$.
- If they cannot be distinguished $\mathrm{CL_S} > \mathrm{CL(S+B)}$.

CLs Limits (more schematic)

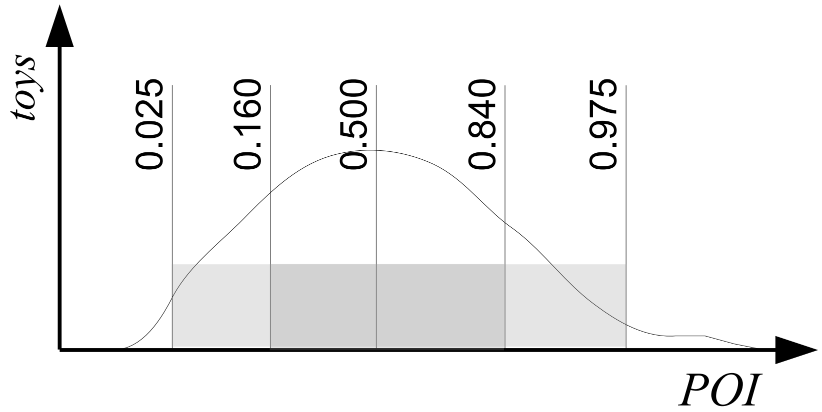
- In particle physics we set more conservative limits than this, following the CLs method:
- Assume H_1 to be signal+background and H_0 to be background only hypothesis.



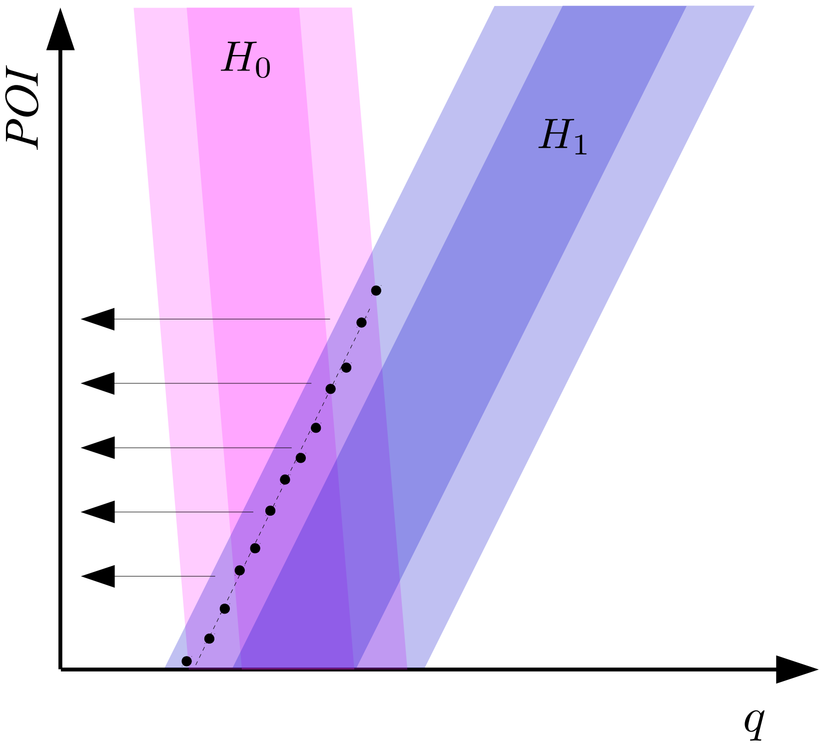


Expected Limit (canonical approach)

- To obtain the expected limit mimic calculation of observed, but base it on toy experiments.
- Make use of the fact that the *pdf*'s do not depend on toys (i.e. schematic plot on the left does not change).
- Throw number of toys under the BG only hypothesis (H_0) determine distribution of 95% CL limits on POI.



- Obtain quantiles for expected limit from this distribution. Usually expected limit = median of this distribution.



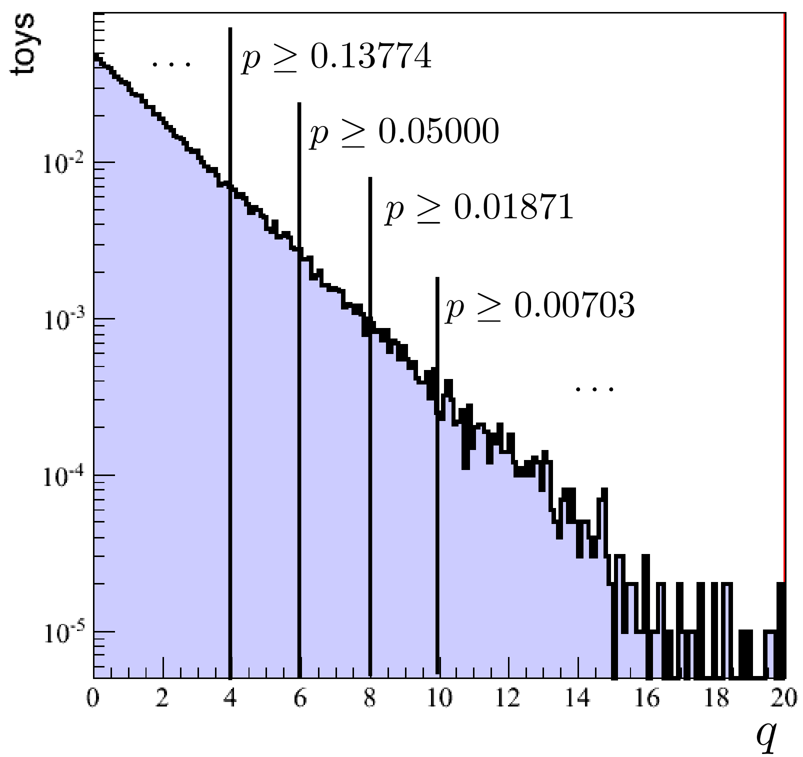
And if the signal appears...

p-Value

- How do we know whether what we see is not just a background fluctuation?
- The p-value is the probability $\mathcal{P}(q \geq q_{\mathrm{obs}}|H_0)$ to observe values of q larger than q_{obs} under the assumption that the background only hypothesis H_0 is the true hypothesis.
- Think of...

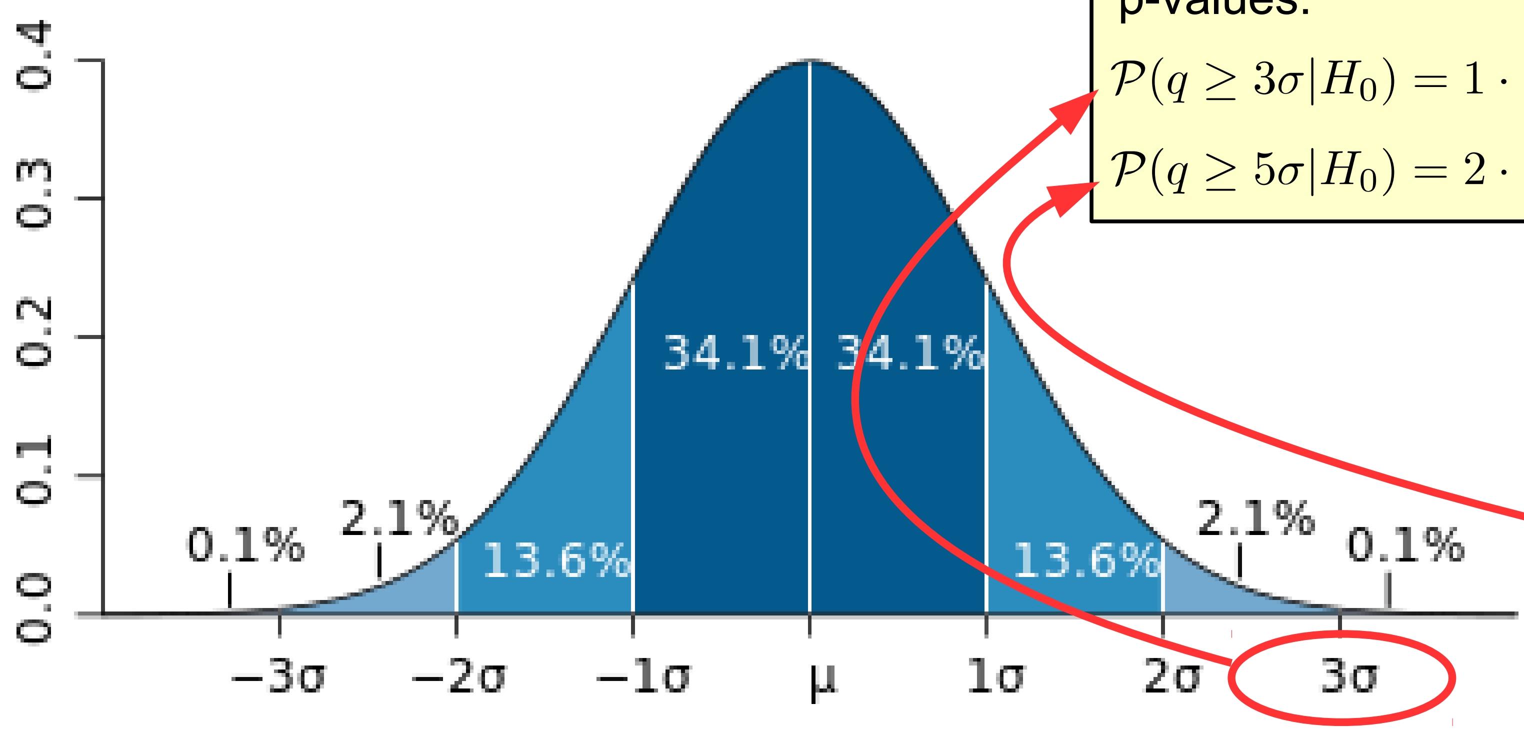
  … the limit as a way to falsify the signal plus background hypothesis (H_1).

  … the p-value as a way to falsify the background only hypothesis (H_0).



Significance

- If the measurement is normal distributed q is distributed according to a χ^2 distribution.
- The χ^2 probability can then be interpreted as a Gaussian confidence interval.



p-values:

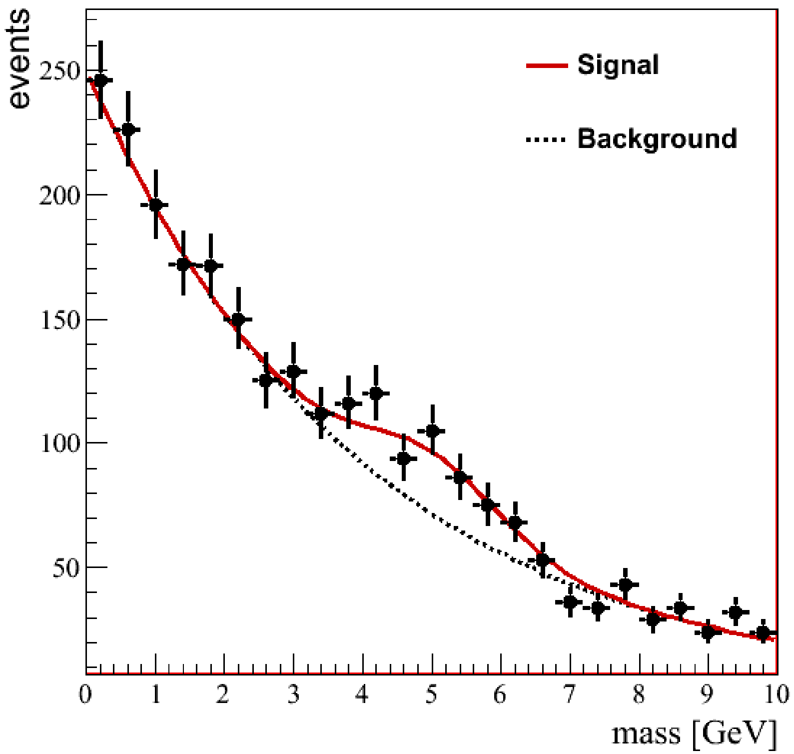
$$\mathcal{P}(q \geq 3\sigma | H_0) = 1 \cdot 10^{-3}$$

$$\mathcal{P}(q \geq 5\sigma | H_0) = 2 \cdot 10^{-5}$$

Significance (in practice)

- If the measurement is normal distributed q is distributed according to a χ^2 distribution.
- The χ^2 probability can then be interpreted as a Gaussian confidence interval.
- Usual approximation in practice is to estimate significances by:

$$\mathcal{S} = \frac{n_{\mathrm{obs}} - n_b}{\sqrt{n_b}}$$

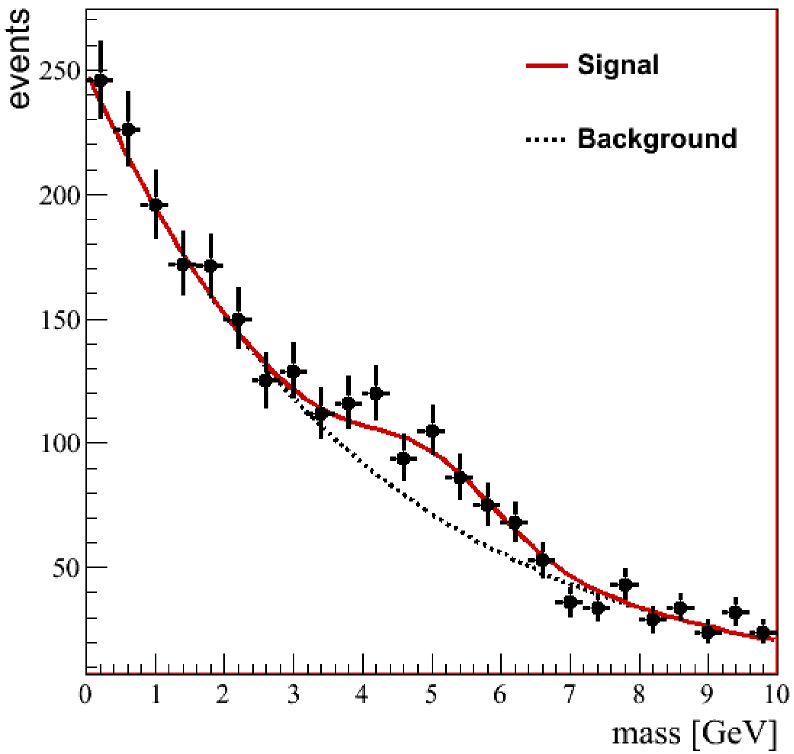


Significance (in practice)

- If the measurement is normal distributed q is distributed according to a χ^2 distribution.
- The χ^2 probability can then be interpreted as a Gaussian confidence interval.
- Usual approximation in practice is to estimate significances by:

expected signal events

$$\mathcal{S} = \frac{n_{\mathrm{obs}} - n_b}{\sqrt{n_b}}$$



Significance (in practice)

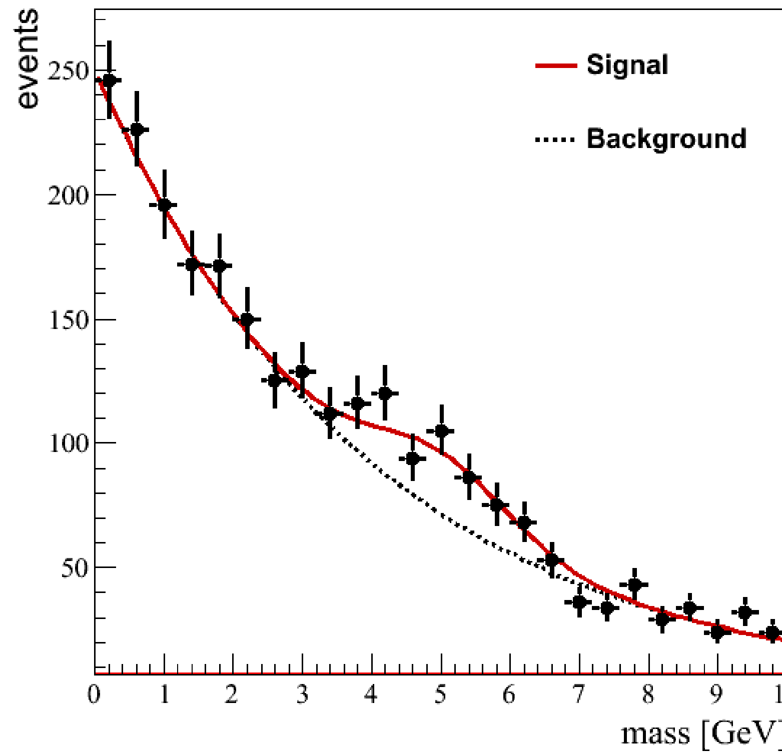
- If the measurement is normal distributed q is distributed according to a χ^2 distribution.
- The χ^2 probability can then be interpreted as a Gaussian confidence interval.

- Usual approximation in practice is to estimate significances by:

expected signal events

$$\mathcal{S} = \frac{n_{\mathrm{obs}} - n_b}{\sqrt{n_b}}$$

Poisson uncertainty on expected background events.



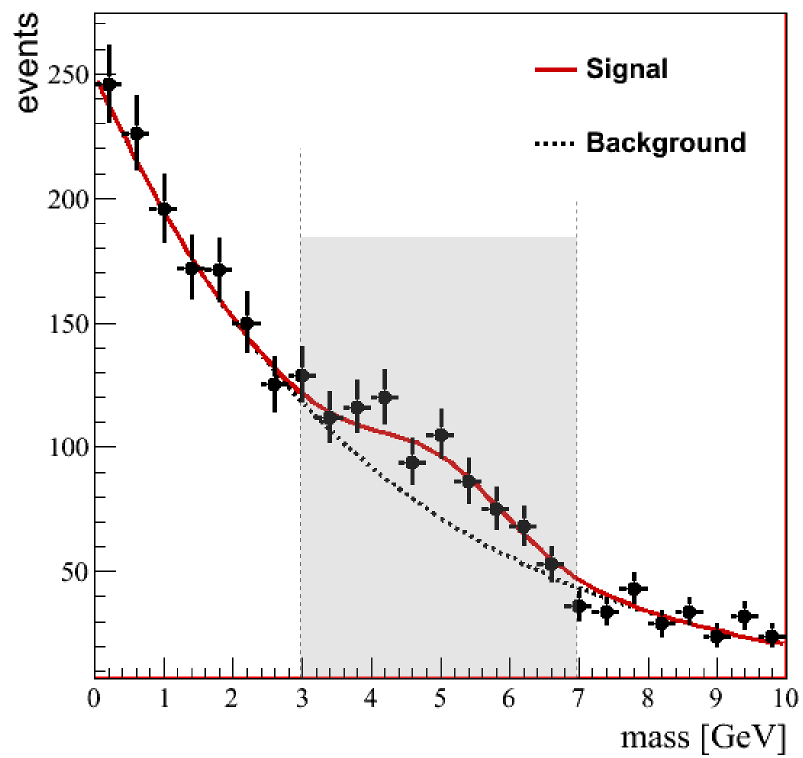
Significance (in practice)

- If the measurement is normal distributed q is distributed according to a χ^2 distribution.
- The χ^2 probability can then be interpreted as a Gaussian confidence interval.
- Usual approximation in practice is to estimate significances by:

expected signal events

$$\mathcal{S} = \frac{n_{\text{obs}} - n_b}{\sqrt{n_b}}$$

Poisson uncertainty on expected background events.



Concluding Remarks

- Reviewed all statistical tools necessary to search for the Higgs boson signal (→ as a small signal above a known background):
  - Probability distributions, likelihood functions, limits, p-values, ...
- Limits are a usual way to 'exclude' the signal hypothesis (H_1).
- p-values are a usual way to 'exclude' the background hypothesis (H_0).
- Under the assumption that the test statistic q is χ^2 distributed p-values can be translated into Gaussian confidence intervals σ.
- In particle physics we call an observation with $\geq 3\sigma$ an evidence.
- We call an observation with $\geq 5\sigma$ a discovery.

Concluding Remarks

- Reviewed all statistical tools necessary to search for the Higgs boson signal ($\rightarrow$ as a small signal above a known background):
  - Probability distributions, likelihood functions, limits, p-values, ...
- Limits are a usual way to 'exclude' the signal hypothesis (H_1).
- p-values are a usual way to 'exclude' the background hypothesis (H_0).
- Under the assumption that the test statistic q is χ^2 distributed p-values can be translated into Gaussian confidence intervals σ.
- In particle physics we call an observation with $\geq 3\sigma$ an evidence.
- We call an observation with $\geq 5\sigma$ a discovery.
- Once a measurement is established the search is over! Measurements of properties are new and different world!

Sneak Preview for Next Week

- Review indirect estimates of the Higgs mass and searches for the Higgs boson that have been made before 2012:
- Estimates of m_t and m_H from high precision measurements at the Z-pole mass at LEP.
- Direct searches for the Higgs boson at LEP.
- Direct searches for the Higgs boson at the Tevatron.
- For the remaining lectures we then will turn towards the discovery of the Higgs boson at the LHC.

During the next lectures we will see 1:1 life examples of all methods that have been presented here.

Backup & Homework Solutions