# Statistics for Particle Physics Lecture 3





Taller de Altas EnergíasBenasque, Spain6 September 2021

http://benasque.org/2021tae/



Glen Cowan Physics Department Royal Holloway, University of London g.cowan@rhul.ac.uk www.pp.rhul.ac.uk/~cowan

# Outline

Lecture 1: Introduction, probability, parameter estimation Lecture 2: Hypothesis tests, limits

→ Lecture 3: Systematic uncertainties, experimental sensitivity

## Approximate confidence intervals/regions from the likelihood function

Suppose we test parameter value(s)  $\theta = (\theta_1, ..., \theta_n)$  using the ratio

$$\lambda(\theta) = \frac{L(\theta)}{L(\hat{\theta})} \qquad \qquad 0 \le \lambda(\theta) \le 1$$

Lower  $\lambda(\theta)$  means worse agreement between data and hypothesized  $\theta$ . Equivalently, usually define

$$t_{\theta} = -2\ln\lambda(\theta)$$

so higher  $t_{\theta}$  means worse agreement between  $\theta$  and the data.

*p*-value of  $\theta$  therefore

$$p_{\theta} = \int_{t_{\theta,\text{obs}}}^{\infty} f(t_{\theta}|\theta) \, dt_{\theta}$$
need pdf

G. Cowan / RHUL Physics

#### Confidence region from Wilks' theorem

Wilks' theorem says (in large-sample limit and provided certain conditions hold...)

 $f(t_{\theta}|\theta) \sim \chi_n^2 \qquad \begin{array}{l} \text{chi-square dist. with $\#$ d.o.f. =} \\ \# \text{ of components in $\theta = (\theta_1, ..., \theta_n)$.} \end{array}$ 

Assuming this holds, the *p*-value is

$$p_{m{ heta}} = 1 - F_{\chi^2_n}(t_{m{ heta}}) \quad \leftarrow$$
 set equal to  $lpha$ 

To find boundary of confidence region set  $p_{\theta} = \alpha$  and solve for  $t_{\theta}$ :

$$t_{\theta} = F_{\chi_n^2}^{-1}(1-\alpha)$$

Recall also

$$t_{ heta} = -2\lnrac{L( heta)}{L(\hat{ heta})}$$

G. Cowan / RHUL Physics

Confidence region from Wilks' theorem (cont.)

i.e., boundary of confidence region in  $\theta$  space is where

$$\ln L(\theta) = \ln L(\hat{\theta}) - \frac{1}{2}F_{\chi_n^2}^{-1}(1-\alpha)$$

For example, for  $1 - \alpha = 68.3\%$  and n = 1 parameter,

$$F_{\chi_1^2}^{-1}(0.683) = 1$$

and so the 68.3% confidence level interval is determined by

$$\ln L(\theta) = \ln L(\hat{\theta}) - \frac{1}{2}$$

Same as recipe for finding the estimator's standard deviation, i.e.,

 $[\hat{\theta} - \sigma_{\hat{\theta}}, \hat{\theta} + \sigma_{\hat{\theta}}]$  is a 68.3% CL confidence interval.

#### Example of interval from $\ln L(\theta)$

For n=1 parameter, CL = 0.683,  $Q_{\alpha} = 1$ .



#### Multiparameter case

For increasing number of parameters,  $CL = 1 - \alpha$  decreases for confidence region determined by a given

$$Q_{\alpha} = F_{\chi_n^2}^{-1}(1-\alpha)$$

$Q_{lpha}$	1-lpha					
	n = 1	n = 2	n = 3	n = 4	n = 5	
1.0	0.683	0.393	0.199	0.090	0.037	
2.0	0.843	0.632	0.428	0.264	0.151	
4.0	0.954	0.865	0.739	0.594	0.451	
9.0	0.997	0.989	0.971	0.939	0.891	

#### Multiparameter case (cont.)

Equivalently,  $Q_{\alpha}$  increases with *n* for a given  $CL = 1 - \alpha$ .

$1 - \alpha$	$\widehat{Q}_{lpha}$					
	n = 1	n = 2	n = 3	n = 4	n = 5	
0.683	1.00	2.30	3.53	4.72	5.89	
0.90	2.71	4.61	6.25	7.78	9.24	
0.95	3.84	5.99	7.82	9.49	11.1	
0.99	6.63	9.21	11.3	13.3	15.1	

## Systematic uncertainties and nuisance parameters In general, our model of the data is not perfect:



Can improve model by including additional adjustable parameters.

$$P(x|\mu) \to P(x|\mu, \theta)$$

Nuisance parameter ↔ systematic uncertainty. Some point in the parameter space of the enlarged model should be "true".

Presence of nuisance parameter decreases sensitivity of analysis to the parameter of interest (e.g., increases variance of estimate).

#### **Profile Likelihood**

Suppose we have a likelihood  $L(\mu, \theta) = P(x|\mu, \theta)$  with Nparameters of interest  $\mu = (\mu_1, ..., \mu_N)$  and M nuisance parameters  $\theta = (\theta_1, ..., \theta_M)$ . The "profiled" (or "constrained") values of  $\theta$  are:

$$\hat{\hat{\theta}}(\boldsymbol{\mu}) = \operatorname*{argmax}_{\boldsymbol{\theta}} L(\boldsymbol{\mu}, \boldsymbol{\theta})$$

and the profile likelihood is:  $L_{\mathrm{p}}(\boldsymbol{\mu}) = L(\boldsymbol{\mu}, \hat{\boldsymbol{ heta}})$ 

The profile likelihood depends only on the parameters of interest; the nuisance parameters are replaced by their profiled values.

The profile likelihood can be used to obtain confidence intervals/regions for the parameters of interest in the same way as one would for all of the parameters from the full likelihood.

#### Profile Likelihood Ratio – Wilks theorem

Goal is to test/reject regions of  $\mu$  space (param. of interest).

Rejecting a point  $\mu$  should mean  $p_{\mu} \leq \alpha$  for all possible values of the nuisance parameters  $\theta$ .

Test  $\boldsymbol{\mu}$  using the "profile likelihood ratio":  $\lambda(\boldsymbol{\mu}) = \frac{L(\boldsymbol{\mu}, \hat{\boldsymbol{\theta}})}{L(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\theta}})}$ 

Let  $t_{\mu} = -2 \ln \lambda(\mu)$ . Wilks' theorem says in large-sample limit:  $t_{\mu} \sim \text{chi-square}(N)$ 

where the number of degrees of freedom is the number of parameters of interest (components of  $\mu$ ). So *p*-value for  $\mu$  is

$$p_{\boldsymbol{\mu}} = \int_{t_{\boldsymbol{\mu},\text{obs}}}^{\infty} f(t_{\boldsymbol{\mu}} | \boldsymbol{\mu}, \boldsymbol{\theta}) \, dt_{\boldsymbol{\mu}} = 1 - F_{\chi_N^2}(t_{\boldsymbol{\mu},\text{obs}})$$

G. Cowan / RHUL Physics

#### Profile Likelihood Ratio – Wilks theorem (2)

If we have a large enough data sample to justify use of the asymptotic chi-square pdf, then if  $\mu$  is rejected, it is rejected for any values of the nuisance parameters.

The recipe to get confidence regions/intervals for the parameters of interest at  $CL = 1 - \alpha$  is thus the same as before, simply use the profile likelihood:

$$\ln L_{\rm p}(\boldsymbol{\mu}) = \ln L_{\rm max} - \frac{1}{2} F_{\chi_N^2}^{-1} (1 - \alpha)$$

where the number of degrees of freedom N for the chi-square quantile is equal to the number of parameters of interest.

If the large-sample limit is not justified, then use e.g. Monte Carlo to get distribution of  $t_{\mu}$ .

#### Prototype search analysis

Search for signal in a region of phase space; result is histogram of some variable x giving numbers:

$$\mathbf{n}=(n_1,\ldots,n_N)$$

Assume the  $n_i$  are Poisson distributed with expectation values

$$E[n_i] = \mu s_i + b_i$$
strength parameter

where

G. Cowan / RHUL Physics

## Prototype analysis (II)

Often also have a subsidiary measurement that constrains some of the background and/or shape parameters:

$$\mathbf{m} = (m_1, \ldots, m_M)$$

Assume the  $m_i$  are Poisson distributed with expectation values

$$E[m_i] = u_i(\boldsymbol{\theta})$$
nuisance parameters ( $\boldsymbol{\theta}_{s}, \boldsymbol{\theta}_{b}, b_{tot}$ )

Likelihood function is

$$L(\mu, \theta) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \quad \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k}$$

G. Cowan / RHUL Physics

## The profile likelihood ratio

Base significance test on the profile likelihood ratio:



Define critical region of test of  $\mu$  by the region of data space that gives the lowest values of  $\lambda(\mu)$ .

Important advantage of profile LR is that its distribution becomes independent of nuisance parameters in large sample limit.

G. Cowan / RHUL Physics

#### Test statistic for discovery

Suppose relevant alternative to background-only ( $\mu = 0$ ) is  $\mu \ge 0$ . So take critical region for test of  $\mu = 0$  corresponding to high  $q_0$ and  $\hat{\mu} > 0$  (data characteristic for  $\mu \ge 0$ ).

That is, to test background-only hypothesis define statistic

$$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{\mu} \ge 0\\ 0 & \hat{\mu} < 0 \end{cases}$$

i.e. here only large (positive) observed signal strength is evidence against the background-only hypothesis.

Note that even though here physically  $\mu \ge 0$ , we allow  $\hat{\mu}$  to be negative. In large sample limit its distribution becomes Gaussian, and this will allow us to write down simple expressions for distributions of our test statistics.

G. Cowan / RHUL Physics

Cowan, Cranmer, Gross, Vitells, arXiv:1007.1727, EPJC 71 (2011) 1554

#### Distribution of $q_0$ in large-sample limit

Assuming approximations valid in the large sample (asymptotic) limit, we can write down the full distribution of  $q_0$  as

$$f(q_0|\mu') = \left(1 - \Phi\left(\frac{\mu'}{\sigma}\right)\right)\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}\exp\left[-\frac{1}{2}\left(\sqrt{q_0} - \frac{\mu'}{\sigma}\right)^2\right]$$

The special case  $\mu' = 0$  is a "half chi-square" distribution:

$$f(q_0|0) = \frac{1}{2}\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}e^{-q_0/2}$$

In large sample limit,  $f(q_0|0)$  independent of nuisance parameters;  $f(q_0|\mu')$  depends on nuisance parameters through  $\sigma$ .

G. Cowan / RHUL Physics

### *p*-value for discovery

Large  $q_0$  means increasing incompatibility between the data and hypothesis, therefore *p*-value for an observed  $q_{0,obs}$  is



use e.g. asymptotic formula



From *p*-value get equivalent significance,

 $Z = \Phi^{-1}(1-p)$ 

Cowan, Cranmer, Gross, Vitells, arXiv:1007.1727, EPJC 71 (2011) 1554

## Cumulative distribution of $q_0$ , significance

From the pdf, the cumulative distribution of  $q_0$  is found to be

$$F(q_0|\mu') = \Phi\left(\sqrt{q_0} - \frac{\mu'}{\sigma}\right)$$

The special case  $\mu' = 0$  is

$$F(q_0|0) = \Phi\left(\sqrt{q_0}\right)$$

The *p*-value of the  $\mu = 0$  hypothesis is

$$p_0 = 1 - F(q_0|0)$$

Therefore the discovery significance Z is simply

$$Z = \Phi^{-1}(1 - p_0) = \sqrt{q_0}$$

G. Cowan / RHUL Physics

Cowan, Cranmer, Gross, Vitells, arXiv:1007.1727, EPJC 71 (2011) 1554

#### Monte Carlo test of asymptotic formula

- $n \sim \text{Poisson}(\mu s + b)$
- $m \sim \text{Poisson}(\tau b)$
- $\mu =$  param. of interest
- *b* = nuisance parameter

Here take *s* known,  $\tau = 1$ .

Asymptotic formula is good approximation to  $5\sigma$ level ( $q_0 = 25$ ) already for  $b \sim 20$ .



#### How to read the $p_0$ plot

The "local"  $p_0$  means the *p*-value of the background-only hypothesis obtained from the test of  $\mu = 0$  at each individual  $m_{\rm H}$ , without any correct for the Look-Elsewhere Effect.

The "Expected" (dashed) curve gives the median  $p_0$  under assumption of the SM Higgs ( $\mu$  = 1) at each  $m_{\rm H}$ .



The blue band gives the width of the distribution  $(\pm 1\sigma)$  of significances under assumption of the SM Higgs.

Cowan, Cranmer, Gross, Vitells, arXiv:1007.1727, EPJC 71 (2011) 1554

### Test statistic for upper limits

For purposes of setting an upper limit on  $\mu$  use

$$q_{\mu} = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} \quad \text{where} \quad \lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta})}$$

I.e. when setting an upper limit, an upwards fluctuation of the data is not taken to mean incompatibility with the hypothesized  $\mu$ :

From observed  $q_{\mu}$  find *p*-value:

$$p_{\mu} = \int_{q_{\mu,\text{obs}}}^{\infty} f(q_{\mu}|\mu) \, dq_{\mu}$$

Large sample approximation:

$$p_{\mu} = 1 - \Phi\left(\sqrt{q_{\mu}}\right)$$

To find upper limit at CL =  $1-\alpha$ , set  $p_{\mu} = \alpha$  and solve for  $\mu$ .

G. Cowan / RHUL Physics

#### Monte Carlo test of asymptotic formulae

Consider again  $n \sim \text{Poisson}(\mu s + b)$ ,  $m \sim \text{Poisson}(\tau b)$ Use  $q_{\mu}$  to find *p*-value of hypothesized  $\mu$  values.

E.g.  $f(q_1|1)$  for *p*-value of  $\mu = 1$ . Typically interested in 95% CL, i.e., *p*-value threshold = 0.05, i.e.,  $q_1 = 2.69$  or  $Z_1 = \sqrt{q_1} = 1.64$ . Median[ $q_1|0$ ] gives "exclusion

sensitivity".

Here asymptotic formulae good for s = 6, b = 9.



How to read the green and yellow limit plots For every value of  $m_{\rm H}$ , find the upper limit on  $\mu$ .

Also for each  $m_{\rm H}$ , determine the distribution of upper limits  $\mu_{\rm up}$  one would obtain under the hypothesis of  $\mu$  = 0.

The dashed curve is the median  $\mu_{up}$ , and the green (yellow) bands give the  $\pm 1\sigma$  ( $2\sigma$ ) regions of this distribution.



#### Sensitivity for Poisson counting experiment

Count a number of events  $n \sim \text{Poisson}(s+b)$ , where

- s = expected number of events from signal,
- b = expected number of background events.

To test for discovery of signal compute p-value of s = 0 hypothesis,

$$p = P(n \ge n_{\text{obs}}|b) = \sum_{n=n_{\text{obs}}}^{\infty} \frac{b^n}{n!} e^{-b} = 1 - F_{\chi^2}(2b; 2n_{\text{obs}})$$

Usually convert to equivalent significance:  $Z = \Phi^{-1}(1-p)$ where  $\Phi$  is the standard Gaussian cumulative distribution, e.g., Z > 5 (a 5 sigma effect) means  $p < 2.9 \times 10^{-7}$ .

To characterize sensitivity to discovery, give expected (mean or median) Z under assumption of a given s.

G. Cowan / RHUL Physics

 $s/\sqrt{b}$  for expected discovery significance For large s + b,  $n \to x \sim \text{Gaussian}(\mu, \sigma)$ ,  $\mu = s + b$ ,  $\sigma = \sqrt{(s + b)}$ . For observed value  $x_{\text{obs}}$ , p-value of s = 0 is  $\text{Prob}(x > x_{\text{obs}} | s = 0)$ ,:

$$p_0 = 1 - \Phi\left(\frac{x_{\rm obs} - b}{\sqrt{b}}\right)$$

Significance for rejecting s = 0 is therefore

$$Z_0 = \Phi^{-1}(1 - p_0) = \frac{x_{\text{obs}} - b}{\sqrt{b}}$$

Expected (median) significance assuming signal rate s is

$$\mathrm{median}[Z_0|s+b] = \frac{s}{\sqrt{b}}$$

G. Cowan / RHUL Physics

#### Better approximation for significance

Poisson likelihood for parameter s is

$$L(s) = \frac{(s+b)^n}{n!} e^{-(s+b)}$$
 For now  
no nuisance

To test for discovery use profile likelihood ratio:

$$q_0 = \begin{cases} -2\ln\lambda(0) & \hat{s} \ge 0 \ , \\ 0 & \hat{s} < 0 \ . \end{cases} \qquad \lambda(s) = \frac{L(s, \hat{\hat{\theta}}(s))}{L(\hat{s}, \hat{\theta})}$$

So the likelihood ratio statistic for testing s = 0 is

$$q_0 = -2\ln\frac{L(0)}{L(\hat{s})} = 2\left(n\ln\frac{n}{b} + b - n\right) \quad \text{for } n > b, \ 0 \text{ otherwise}$$

G. Cowan / RHUL Physics

TAE 2021 / Statistics for PP Lecture 3

params.

#### Approximate Poisson significance (continued)

For sufficiently large s + b, (use Wilks' theorem),

$$Z = \sqrt{2\left(n\ln\frac{n}{b} + b - n\right)} \quad \text{for } n > b \text{ and } Z = 0 \text{ otherwise.}$$

To find median[Z|s], let  $n \rightarrow s + b$  (i.e., the Asimov data set):

$$Z_{\rm A} = \sqrt{2\left(\left(s+b\right)\ln\left(1+\frac{s}{b}\right)-s\right)}$$

This reduces to  $s/\sqrt{b}$  for  $s \ll b$ .

G. Cowan / RHUL Physics

 $n \sim \text{Poisson}(s+b)$ , median significance, assuming *s*, of the hypothesis s = 0

CCGV, EPJC 71 (2011) 1554, arXiv:1007.1727



"Exact" values from MC, jumps due to discrete data.

Asimov  $\sqrt{q_{0,A}}$  good approx. for broad range of *s*, *b*.

 $s/\sqrt{b}$  only good for  $s \ll b$ .

G. Cowan / RHUL Physics

## Finally

#### Three lectures only enough for a brief introduction to:

- Parameter estimation, maximum likelihood
- Hypothesis tests, p-values
- Limits (confidence intervals/regions)
- Systematics (nuisance parameters)
- Asymptotic methods (Wilks' theorem)
- No time for many important topics, e.g.,
  - Machine Learning, Bayesian methods, unfolding, MCMC,...

Final thought: once the basic formalism is fixed, most of the work focuses on writing down the likelihood, e.g.,  $P(x|\theta)$ , and including in it enough parameters to adequately describe the data (true for both Bayesian and frequentist approaches).



## Extending $s/\sqrt{b}$ to case where b uncertain

The intuitive explanation of  $s/\sqrt{b}$  is that it compares the signal, s, to the standard deviation of n assuming no signal,  $\sqrt{b}$ .

Now suppose the value of b is uncertain, characterized by a standard deviation  $\sigma_b$ .

A reasonable guess is to replace  $\sqrt{b}$  by the quadratic sum of  $\sqrt{b}$  and  $\sigma_b$ , i.e.,

$$\operatorname{med}[Z|s] = \frac{s}{\sqrt{b + \sigma_b^2}}$$

This has been used to optimize some analyses e.g. where  $\sigma_b$  cannot be neglected.

#### Profile likelihood with *b* uncertain

This is the well studied "on/off" problem: Cranmer 2005; Cousins, Linnemann, and Tucker 2008; Li and Ma 1983,...

Measure two Poisson distributed values:

 $n \sim \text{Poisson}(s+b)$  (primary or "search" measurement)  $m \sim \text{Poisson}(\tau b)$  (control measurement,  $\tau$  known) The likelihood function is

$$L(s,b) = \frac{(s+b)^n}{n!} e^{-(s+b)} \frac{(\tau b)^m}{m!} e^{-\tau b}$$

Use this to construct profile likelihood ratio (*b* is nuisance parameter):  $L(0, \hat{\hat{b}}(0))$ 

$$\lambda(0) = \frac{L(0, b(0))}{L(\hat{s}, \hat{b})}$$

G. Cowan / RHUL Physics

#### Ingredients for profile likelihood ratio

To construct profile likelihood ratio from this need estimators:

$$\hat{s} = n - m/\tau , \hat{b} = m/\tau , \hat{b}(s) = \frac{n + m - (1 + \tau)s + \sqrt{(n + m - (1 + \tau)s)^2 + 4(1 + \tau)sm}}{2(1 + \tau)} .$$

and in particular to test for discovery (s = 0),

$$\hat{\hat{b}}(0) = \frac{n+m}{1+\tau}$$

G. Cowan / RHUL Physics

## Asymptotic significance

Use profile likelihood ratio for  $q_0$ , and then from this get discovery significance using asymptotic approximation (Wilks' theorem):

$$Z = \sqrt{q_0}$$
$$= \left[ -2\left(n\ln\left[\frac{n+m}{(1+\tau)n}\right] + m\ln\left[\frac{\tau(n+m)}{(1+\tau)m}\right]\right) \right]^{1/2}$$

for  $n > \hat{b}$  and Z = 0 otherwise.

#### Essentially same as in:

Robert D. Cousins, James T. Linnemann and Jordan Tucker, NIM A 595 (2008) 480– 501; arXiv:physics/0702156.

Tipei Li and Yuqian Ma, Astrophysical Journal 272 (1983) 317–324.

G. Cowan / RHUL Physics

#### Asimov approximation for median significance

To get median discovery significance, replace *n*, *m* by their expectation values assuming background-plus-signal model:

$$n \to s + b$$
  

$$m \to \tau b$$

$$Z_{\rm A} = \left[ -2\left( (s+b) \ln\left[\frac{s+(1+\tau)b}{(1+\tau)(s+b)}\right] + \tau b \ln\left[1 + \frac{s}{(1+\tau)b}\right] \right) \right]^{1/2}$$
Or use the variance of  $\hat{b} = m/\tau$ ,  $V[\hat{b}] \equiv \sigma_b^2 = \frac{b}{\tau}$ , to eliminate  $\tau$ :
$$Z_{\rm A} = \left[ 2\left( (s+b) \ln\left[\frac{(s+b)(b+\sigma_b^2)}{b^2+(s+b)\sigma_b^2}\right] - \frac{b^2}{\sigma_b^2} \ln\left[1 + \frac{\sigma_b^2 s}{b(b+\sigma_b^2)}\right] \right) \right]^{1/2}$$

#### Limiting cases

Expanding the Asimov formula in powers of s/b and  $\sigma_b^2/b$  (=  $1/\tau$ ) gives

$$Z_{\rm A} = \frac{s}{\sqrt{b + \sigma_b^2}} \left( 1 + \mathcal{O}(s/b) + \mathcal{O}(\sigma_b^2/b) \right)$$

So the "intuitive" formula can be justified as a limiting case of the significance from the profile likelihood ratio test evaluated with the Asimov data set. Testing the formulae: s = 5



G. Cowan / RHUL Physics

#### Using sensitivity to optimize a cut



Figure 1: (a) The expected significance as a function of the cut value  $x_{cut}$ ; (b) the distributions of signal and background with the optimal cut value indicated.

#### G. Cowan / RHUL Physics

#### Summary on discovery sensitivity

Simple formula for expected discovery significance based on profile likelihood ratio test and Asimov approximation:

$$Z_{\rm A} = \left[ 2 \left( (s+b) \ln \left[ \frac{(s+b)(b+\sigma_b^2)}{b^2 + (s+b)\sigma_b^2} \right] - \frac{b^2}{\sigma_b^2} \ln \left[ 1 + \frac{\sigma_b^2 s}{b(b+\sigma_b^2)} \right] \right) \right]^{1/2}$$

For large *b*, all formulae OK.

For small *b*,  $s/\sqrt{b}$  and  $s/\sqrt{(b+\sigma_b^2)}$  overestimate the significance.

Could be important in optimization of searches with low background.

Formula maybe also OK if model is not simple on/off experiment, e.g., several background control measurements (check this).

#### *p*-values in cases with nuisance parameters

Suppose we have a statistic  $q_{\theta}$  that we use to test a hypothesized value of a parameter  $\theta$ , such that the *p*-value of  $\theta$  is

$$p_{\theta} = \int_{q_{\theta,\text{obs}}}^{\infty} f(q_{\theta}|\theta,\nu) \, dq_{\theta}$$

But what values of v to use for  $f(q_{\theta} | \theta, v)$ ?

Fundamentally we want to reject  $\theta$  only if  $p_{\theta} < \alpha$  for all v.

$$\rightarrow$$
 "exact" confidence interval

But in general for finite data samples this is not true; one may be unable to reject some  $\theta$  values if all values of v must be considered (resulting interval for  $\theta$  "overcovers").

## Profile construction ("hybrid resampling")

K. Cranmer, PHYSTAT-LHC Workshop on Statistical Issues for LHC Physics, 2008. oai:cds.cern.ch:1021125, cdsweb.cern.ch/record/1099969.

Approximate procedure is to reject  $\theta$  if  $p_{\theta} \le \alpha$  where the *p*-value is computed assuming the value of the nuisance parameter that best fits the data for the specified  $\theta$ :

^	"double hat" notation means profiled
$\hat{\nu}(\theta)$	value, i.e., parameter that maximizes
	likelihood for the given $\theta$ .

The resulting confidence interval will have the correct coverage for the points  $(\theta, \hat{v}(\theta))$ .

Elsewhere it may under- or overcover, but this is usually as good as we can do (check with MC if crucial or small sample problem).

G. Cowan / RHUL Physics