<table>
<thead>
<tr>
<th>Lecture</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Probability, Bayes’ theorem, random variables, pdfs</td>
</tr>
<tr>
<td>2</td>
<td>Functions of r.v.s, expectation values, error propagation</td>
</tr>
<tr>
<td>3</td>
<td>Catalogue of pdfs</td>
</tr>
<tr>
<td>4</td>
<td>The Monte Carlo method</td>
</tr>
<tr>
<td>5</td>
<td>Statistical tests: general concepts</td>
</tr>
<tr>
<td>6</td>
<td>Test statistics, multivariate methods</td>
</tr>
<tr>
<td>7</td>
<td>Significance tests</td>
</tr>
<tr>
<td>8</td>
<td>Parameter estimation, maximum likelihood</td>
</tr>
<tr>
<td>9</td>
<td>More maximum likelihood</td>
</tr>
<tr>
<td>10</td>
<td>Method of least squares</td>
</tr>
<tr>
<td>11</td>
<td>Interval estimation, setting limits</td>
</tr>
<tr>
<td>12</td>
<td>Nuisance parameters, systematic uncertainties</td>
</tr>
<tr>
<td>13</td>
<td>Examples of Bayesian approach</td>
</tr>
<tr>
<td>14</td>
<td>tba</td>
</tr>
</tbody>
</table>
Interval estimation — introduction

In addition to a ‘point estimate’ of a parameter we should report an interval reflecting its statistical uncertainty.

Desirable properties of such an interval may include:

- communicate objectively the result of the experiment;
- have a given probability of containing the true parameter;
- provide information needed to draw conclusions about the parameter possibly incorporating stated prior beliefs.

Often use $+/-$ the estimated standard deviation of the estimator. In some cases, however, this is not adequate:

- estimate near a physical boundary,
- e.g., an observed event rate consistent with zero.

We will look briefly at Frequentist and Bayesian intervals.
Frequentist confidence intervals

Consider an estimator $\hat{\theta}$ for a parameter $\theta$ and an estimate $\hat{\theta}_{\text{obs}}$. We also need for all possible $\theta$ its sampling distribution $g(\hat{\theta}; \theta)$.

Specify upper and lower tail probabilities, e.g., $\alpha = 0.05$, $\beta = 0.05$, then find functions $u_{\alpha}(\theta)$ and $v_{\beta}(\theta)$ such that:

\[
\begin{align*}
\alpha &= P(\hat{\theta} \geq u_{\alpha}(\theta)) \\
 &= \int_{u_{\alpha}(\theta)}^{\infty} g(\hat{\theta}; \theta) \, d\hat{\theta} \\
\beta &= P(\hat{\theta} \leq v_{\beta}(\theta)) \\
 &= \int_{-\infty}^{v_{\beta}(\theta)} g(\hat{\theta}; \theta) \, d\hat{\theta}
\end{align*}
\]
Confidence interval from the confidence belt

The region between $u_\alpha(\theta)$ and $v_\beta(\theta)$ is called the confidence belt.

Find points where observed estimate intersects the confidence belt.

This gives the confidence interval $[a, b]$

Confidence level = $1 - \alpha - \beta = $ probability for the interval to cover true value of the parameter (holds for any possible true $\theta$).
Confidence intervals by inverting a test

Confidence intervals for a parameter $\theta$ can be found by defining a test of the hypothesized value $\theta$ (do this for all $\theta$):

Specify values of the data that are ‘disfavoured’ by $\theta$ (critical region) such that $P(\text{data in critical region}) \leq \gamma$ for a prespecified $\gamma$, e.g., 0.05 or 0.1.

If data observed in the critical region, reject the value $\theta$.

Now invert the test to define a confidence interval as:

set of $\theta$ values that would not be rejected in a test of size $\gamma$ (confidence level is $1 - \gamma$).

The interval will cover the true value of $\theta$ with probability $\geq 1 - \gamma$.

Equivalent to confidence belt construction; confidence belt is acceptance region of a test.
Relation between confidence interval and $p$-value

Equivalently we can consider a significance test for each hypothesized value of $\theta$, resulting in a $p$-value, $p_{\theta}$.

If $p_{\theta} < \gamma$, then we reject $\theta$.

The confidence interval at $\text{CL} = 1 - \gamma$ consists of those values of $\theta$ that are not rejected.

E.g. an upper limit on $\theta$ is the greatest value for which $p_{\theta} \geq \gamma$.

In practice find by setting $p_{\theta} = \gamma$ and solve for $\theta$. 
Confidence intervals in practice

The recipe to find the interval $[a, b]$ boils down to solving

$$
\alpha = \int_{u_{\alpha}(\theta)}^{\infty} g(\hat{\theta}; \theta) \, d\hat{\theta} = \int_{\hat{\theta}_{\text{obs}}}^{\infty} g(\hat{\theta}; a) \, d\hat{\theta}, \\
\beta = \int_{-\infty}^{v_{\beta}(\theta)} g(\hat{\theta}; \theta) \, d\hat{\theta} = \int_{-\infty}^{\hat{\theta}_{\text{obs}}} g(\hat{\theta}; b) \, d\hat{\theta}.
$$

→ $a$ is hypothetical value of $\theta$ such that $P(\hat{\theta} > \hat{\theta}_{\text{obs}}) = \alpha$.
→ $b$ is hypothetical value of $\theta$ such that $P(\hat{\theta} < \hat{\theta}_{\text{obs}}) = \beta$. 
Meaning of a confidence interval

N.B. the interval is random, the true $\theta$ is an unknown constant.

Often report interval $[a, b]$ as $\hat{\theta}^{+d}_{-c}$, i.e. $c = \hat{\theta} - a$, $d = b - \hat{\theta}$.

So what does $\hat{\theta} = 80.25^{+0.31}_{-0.25}$ mean? It does not mean:

$$P(80.00 < \theta < 80.56) = 1 - \alpha - \beta,$$

but rather:

repeat the experiment many times with same sample size,

construct interval according to same prescription each time,

in $1 - \alpha - \beta$ of experiments, interval will cover $\theta$. 
Central vs. one-sided confidence intervals

Sometimes only specify $\alpha$ or $\beta$, $\rightarrow$ one-sided interval (limit)

Often take $\alpha = \beta = \frac{\gamma}{2} \rightarrow$ coverage probability $= 1 - \gamma$

$\rightarrow$ central confidence interval

N.B. ‘central’ confidence interval does not mean the interval is symmetric about $\hat{\theta}$, but only that $\alpha = \beta$.

The HEP error ‘convention’: 68.3% central confidence interval.
Intervals from the likelihood function

In the large sample limit it can be shown for ML estimators:

\[ \tilde{\theta} \sim N(\bar{\theta}, V) \quad (n\text{-dimensional Gaussian, covariance } V) \]

\[ L(\tilde{\theta}) = L_{\text{max}} \exp \left[ -\frac{1}{2} Q(\tilde{\theta}, \bar{\theta}) \right] , \quad Q(\tilde{\theta}, \bar{\theta}) = (\tilde{\theta} - \bar{\theta})^T V^{-1} (\tilde{\theta} - \bar{\theta}) \]

\[ Q(\tilde{\theta}, \bar{\theta}) = Q_\gamma \] defines a hyper-ellipsoidal confidence region,

\[ P(\text{ellipsoid covers true } \bar{\theta}) = P(Q(\tilde{\theta}, \bar{\theta}) \leq Q_\gamma) \]

If \[ \tilde{\theta} \sim N(\bar{\theta}, V) \] then \[ Q(\tilde{\theta}, \bar{\theta}) \sim \text{Chi-square}(n) \]

coverage probability \( \equiv 1 - \gamma = \int_0^{Q_\gamma} f_{\chi^2}(z; n) \, dz = F_{\chi^2}(Q_\gamma; n) \)
Approximate confidence regions from $L(\theta)$

So the recipe to find the confidence region with CL = $1-\gamma$ is:

$$\ln L(\bar{\theta}) = \ln L_{\text{max}} - \frac{Q_\gamma}{2} \quad \text{or} \quad \chi^2(\bar{\theta}) = \chi^2_{\text{min}} + Q_\gamma$$

where $Q_\gamma = F^{-1}_{\chi^2}(1 - \gamma; n)$

<table>
<thead>
<tr>
<th>$Q_\gamma$</th>
<th>$1 - \gamma$</th>
<th>$Q_\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 1$</td>
<td>$n = 2$</td>
</tr>
<tr>
<td>1.0</td>
<td>0.683</td>
<td>0.393</td>
</tr>
<tr>
<td>2.0</td>
<td>0.843</td>
<td>0.632</td>
</tr>
<tr>
<td>4.0</td>
<td>0.954</td>
<td>0.865</td>
</tr>
<tr>
<td>9.0</td>
<td>0.997</td>
<td>0.989</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$1 - \gamma$</th>
<th>$n = 1$</th>
<th>$n = 2$</th>
<th>$n = 3$</th>
<th>$n = 4$</th>
<th>$n = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.683</td>
<td>1.00</td>
<td>2.30</td>
<td>3.53</td>
<td>4.72</td>
<td>5.89</td>
</tr>
<tr>
<td>0.90</td>
<td>2.71</td>
<td>4.61</td>
<td>6.25</td>
<td>7.78</td>
<td>9.24</td>
</tr>
<tr>
<td>0.95</td>
<td>3.84</td>
<td>5.99</td>
<td>7.82</td>
<td>9.49</td>
<td>11.1</td>
</tr>
<tr>
<td>0.99</td>
<td>6.63</td>
<td>9.21</td>
<td>11.3</td>
<td>13.3</td>
<td>15.1</td>
</tr>
</tbody>
</table>

For finite samples, these are approximate confidence regions.

Coverage probability not guaranteed to be equal to $1-\gamma$;
no simple theorem to say by how far off it will be (use MC).

Remember here the interval is random, not the parameter.
Example of interval from $\ln L(\theta)$

For $n=1$ parameter, $CL = 0.683$, $Q_\gamma = 1$.

Our exponential example, now with $n = 5$ observations:

\[
\hat{\tau} = 0.85^{+0.52}_{-0.30}
\]
Setting limits on Poisson parameter

Consider again the case of finding $n = n_s + n_b$ events where

- $n_b$ events from known processes (background)
- $n_s$ events from a new process (signal)

are Poisson r.v.s with means $s$, $b$, and thus $n = n_s + n_b$

is also Poisson with mean $= s + b$. Assume $b$ is known.

Suppose we are searching for evidence of the signal process, but the number of events found is roughly equal to the expected number of background events, e.g., $b = 4.6$ and we observe $n_{obs} = 5$ events.

The evidence for the presence of signal events is not statistically significant,

→ set upper limit on the parameter $s$. 
Upper limit for Poisson parameter

Find the hypothetical value of $s$ such that there is a given small probability, say, $\gamma = 0.05$, to find as few events as we did or less:

$$\gamma = P(n \leq n_{\text{obs}}; s, b) = \sum_{n=0}^{n_{\text{obs}}} \frac{(s + b)^n}{n!} e^{-(s+b)}$$

Solve numerically for $s = s_{\text{up}}$, this gives an upper limit on $s$ at a confidence level of $1-\gamma$.

Example: suppose $b = 0$ and we find $n_{\text{obs}} = 0$. For $1-\gamma = 0.95$,

$$\gamma = P(n = 0; s, b = 0) = e^{-s} \quad \rightarrow \quad s_{\text{up}} = -\ln \gamma \approx 3.00$$
Calculating Poisson parameter limits

To solve for $s_{lo}$, $s_{up}$, can exploit relation to $\chi^2$ distribution:

$$s_{lo} = \frac{1}{2} F_{\chi^2}^{-1}(\alpha; 2n) - b$$

$$s_{up} = \frac{1}{2} F_{\chi^2}^{-1}(1 - \beta; 2(n + 1)) - b$$

For low fluctuation of $n$ this can give negative result for $s_{up}$; i.e. confidence interval is empty.
Limits near a physical boundary

Suppose e.g. $b = 2.5$ and we observe $n = 0$.

If we choose $CL = 0.9$, we find from the formula for $s_{up}$

$$s_{up} = -0.197 \quad (CL = 0.90)$$

Physicist:
We already knew $s \geq 0$ before we started; can’t use negative upper limit to report result of expensive experiment!

Statistician:
The interval is designed to cover the true value only 90% of the time ― this was clearly not one of those times.

Not uncommon dilemma when limit of parameter is close to a physical boundary.
Expected limit for $s = 0$

Physicist: I should have used CL = 0.95 — then $s_{up} = 0.496$

Even better: for CL = 0.917923 we get $s_{up} = 10^{-4}$!

Reality check: with $b = 2.5$, typical Poisson fluctuation in $n$ is at least $\sqrt{2.5} = 1.6$. How can the limit be so low?

Look at the mean limit for the no-signal hypothesis ($s = 0$) (sensitivity).

Distribution of 95% CL limits with $b = 2.5$, $s = 0$. Mean upper limit = 4.44
The Bayesian approach

In Bayesian statistics need to start with ‘prior pdf’ $\pi(\theta)$, this reflects degree of belief about $\theta$ before doing the experiment.

Bayes’ theorem tells how our beliefs should be updated in light of the data $x$:

$$p(\theta|x) = \frac{L(x|\theta)p(\theta)}{\int L(x|\theta')\pi(\theta') d\theta'} \propto L(x|\theta)\pi(\theta)$$

Integrate posterior pdf $p(\theta | x)$ to give interval with any desired probability content.

For e.g. Poisson parameter 95% CL upper limit from

$$0.95 = \int_{-\infty}^{sup} p(s|n) ds$$
Bayesian prior for Poisson parameter

Include knowledge that $s \geq 0$ by setting prior $\pi(s) = 0$ for $s < 0$.

Often try to reflect ‘prior ignorance’ with e.g.

$$\pi(s) = \begin{cases} 
1 & s \geq 0 \\
0 & \text{otherwise}
\end{cases}$$

Not normalized but this is OK as long as $L(s)$ dies off for large $s$.

Not invariant under change of parameter — if we had used instead a flat prior for, say, the mass of the Higgs boson, this would imply a non-flat prior for the expected number of Higgs events.

Doesn’t really reflect a reasonable degree of belief, but often used as a point of reference;

or viewed as a recipe for producing an interval whose frequentist properties can be studied (coverage will depend on true $s$).
Bayesian interval with flat prior for $s$

Solve numerically to find limit $s_{\text{up}}$.

For special case $b = 0$, Bayesian upper limit with flat prior numerically same as classical case (‘coincidence’).

Otherwise Bayesian limit is everywhere greater than classical (‘conservative’).

Never goes negative.

Doesn’t depend on $b$ if $n = 0$.
Likelihood ratio limits (Feldman-Cousins)

Define likelihood ratio for hypothesized parameter value $s$:

$$l(s) = \frac{L(n|s, b)}{L(n|\hat{s}, b)} \quad \text{where} \quad \hat{s} = \begin{cases} n - b & n \geq b, \\ 0 & \text{otherwise} \end{cases}$$

Here $\hat{s}$ is the ML estimator, note $0 \leq l(s) \leq 1$.

Critical region defined by low values of likelihood ratio.

Resulting intervals can be one- or two-sided (depending on $n$).


See also Cowan, Cranmer, Gross & Vitells, arXiv:1007.1727 for details on including systematic errors and on asymptotic sampling distribution of likelihood ratio statistic.
Wrapping up lecture 12

In large sample limit and away from physical boundaries, +/- 1 standard deviation is all you need for 68% CL.

Frequentist confidence intervals

  Complicated! Random interval that contains true parameter with fixed probability.

  Can be obtained by inversion of a test; freedom left as to choice of test.

  Log-likelihood can be used to determine approximate confidence intervals (or regions)

Bayesian intervals

  Conceptually easy — just integrate posterior pdf.

  Requires choice of prior.
Extra slides
Distance between estimated and true $\theta$

**Fig. 9.7** (a) A contour of constant $g(\hat{\theta}; \theta_{\text{true}})$ (i.e. constant $Q(\hat{\theta}, \theta_{\text{true}})$) in $\hat{\theta}$-space. (b) A contour of constant $L(\theta)$ corresponding to constant $Q(\hat{\theta}_{\text{obs}}, \theta)$ in $\theta$-space. The values $\theta_{\text{true}}$ and $\hat{\theta}_{\text{obs}}$ represent particular constant values of $\theta$ and $\hat{\theta}$, respectively.
More on intervals from LR test (Feldman-Cousins)

Caveat with coverage: suppose we find $n >> b$. Usually one then quotes a measurement: $\hat{s} = n - b$, $\hat{\sigma}_s = \sqrt{n}$

If, however, $n$ isn’t large enough to claim discovery, one sets a limit on $s$.

FC pointed out that if this decision is made based on $n$, then the actual coverage probability of the interval can be less than the stated confidence level (‘flip-flopping’).

FC intervals remove this, providing a smooth transition from 1- to 2-sided intervals, depending on $n$.

But, suppose FC gives e.g. $0.1 < s < 5$ at 90% CL, $p$-value of $s=0$ still substantial. Part of upper-limit ‘wasted’?
Properties of upper limits

Example: take $b = 5.0$, $1 - \gamma = 0.95$

Upper limit $s_{up}$ vs. $n$

Mean upper limit vs. $s$
Upper limit versus $b$

If $n = 0$ observed, should upper limit depend on $b$?
- Classical: yes
- Bayesian: no
- FC: yes

Feldman & Cousins, PRD 57 (1998) 3873
Coverage probability of intervals

Because of discreteness of Poisson data, probability for interval to include true value in general > confidence level (‘over-coverage’)

![Coverage probability graphs]

Bayes $\pi(s) = \text{const.}$

Total and upper bounds for confidence level.

Classical and Likelihood methods.

G. Cowan
Lectures on Statistical Data Analysis

28