Lecture 5

Bayes' Theorem:

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

Typical problem in experimental physics:

Given model parameters θ_i we can calculate a p.d.f. for observations y_i:

$$f(\vec{y}; \theta_1, \ldots, \theta_m) \qquad [\,P(B|A)\,]$$

What we are interested in is the p.d.f. for parameters θ_i, given the observations y_i.

The likelihood function is the overall p.d.f. evaluated at the observed values y_i:

$$L = f(\vec{y}; \theta_1, \ldots, \theta_m)$$

This is not the same as the probability of observing $\vec{y}$.

For independent measurements,

$$L = \prod_{i=1}^{N} f(y_1^i, \ldots, y_n^i; \theta_1, \ldots, \theta_m)$$

Bayes' Theorem:

$$P(\vec{\theta}\,|\,\vec{y}) = \frac{L(\vec{y}|\vec{\theta})\,\pi(\vec{\theta})}{\int L(\vec{y}|\vec{\theta})\,\pi(\vec{\theta})\,d\theta^m}$$

What is $\pi(\vec{\theta})$? It's the subjective prior probability of parameters $\vec{\theta}$.

If L is sharply peaked about the parameters $\vec{\theta}$ and if $\pi(\vec{\theta})$ is smooth then $P(\vec{\theta}\,|\,\vec{y})$ will not depend strongly on π.

Independent observations:

$$P_1(\vec{\theta}) \propto \pi(\vec{\theta})\, L_1(\vec{y}_1;\vec{\theta})$$

$$P_2(\vec{\theta}) \propto P_1(\vec{\theta})\, L_2(\vec{y}_2;\vec{\theta})$$

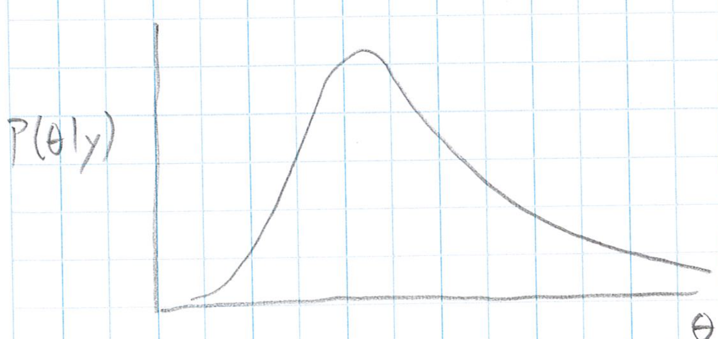
$$P_3(\vec{\theta}) \propto P_2(\vec{\theta})\, L_3(\vec{y}_3;\vec{\theta})$$

So $P_3(\vec{\theta}) \propto \pi(\vec{\theta})\, L_1(\vec{y}_1;\vec{\theta}) \cdot L_2(\vec{y}_2;\vec{\theta}) \cdot L_3(\vec{y}_3;\vec{\theta})$

One dimension:

$$P(\theta; y) = \frac{L(y|\theta)\,\pi(\theta)}{\int L(y|\theta)\,\pi(\theta)\,d\theta}$$

y is an observation. After y is measured we can plot $P(\theta|y)$:

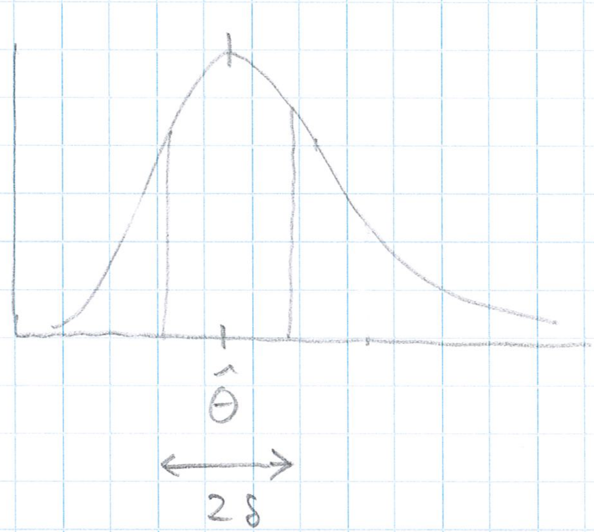


At some point there will be a maximum. This is the most probable value of θ.

If we measured a different y' then we would obtain a different $\hat{\theta}'$ that maximizes L.

The probability that the true value of θ_{true} lies within $\hat{\theta} \pm \delta$ is $1-\alpha = C.L.$ is

$$\int_{\hat{\theta}-\delta}^{\hat{\theta}+\delta} P(\theta|y)\, d\theta = 1-\alpha$$



$$P(\hat{\theta}-\delta < \theta_{true} < \hat{\theta}+\delta) = 1-\alpha = C.L.$$

(Specify C.L. then solve for δ).

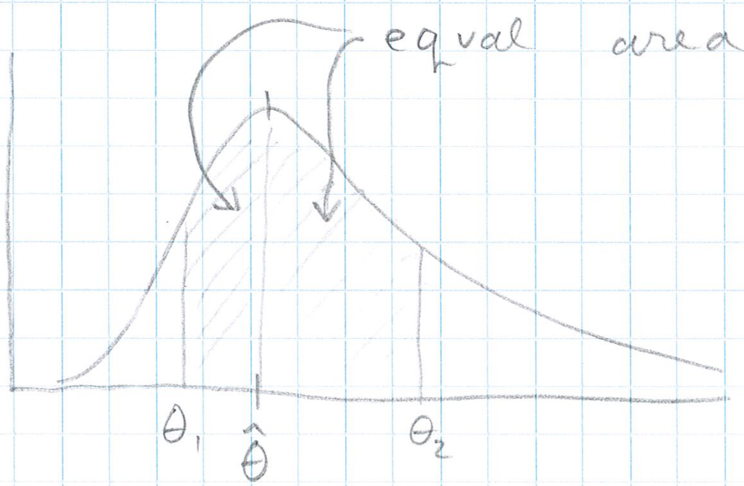
Frequent choices correspond to gaussian $1\sigma = 68.3\%$ but also 90% or 95%.

Typically we write $\theta = \hat{\theta} \pm \delta$.

We can also define asymmetric confidence intervals:

$$\int_{-\infty}^{\theta_1} P(\theta|y)\,d\theta = \frac{\alpha}{2}$$

$$\int_{\theta_2}^{\infty} P(\theta|y)\,d\theta = \frac{\alpha}{2}$$



Typically we write $\theta = \hat{\theta}^{+\delta_1}_{-\delta_2}$

where $\delta_1 = \hat{\theta} - \theta_1$
$\delta_2 = \theta_2 - \hat{\theta}$.

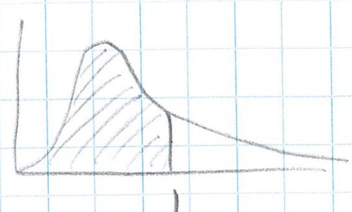
Then $P(\theta_1 < \theta_{true} < \theta_2) = 1 - \alpha$

One-sided confidence intervals.

What is $P(\theta_{true} > \theta_{lower})$? lower limit

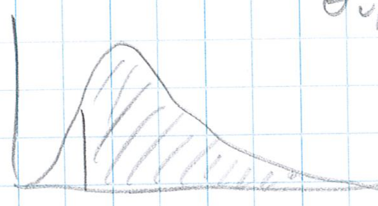
What is $P(\theta_{true} < \theta_{upper})$? upper limit

$$\int_{-\infty}^{\theta_{upper}} P(\theta|y)\,d\theta = 1-\alpha$$



θ_{upper}

$$\int_{\theta_{lower}}^{\infty} P(\theta|y)\,d\theta = 1-\alpha$$



θ_{lower}

Where typically $\alpha = 0.1$ (90% C.L.)
or $\alpha = 0.05$ (95% C.L.)

Example:

Suppose y is gaussian distributed with mean μ and RMS σ.

$$f(y \mid \mu, \sigma_i) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-(y-\mu)^2/2\sigma^2}$$

Furthermore, suppose σ_i is known (eg from measurement uncertainties).

$$L(\vec{y} \mid \mu) = \prod_{i=1}^{n} f(y_i \mid \mu)$$

$$= \frac{1}{(2\pi)^{n/2}} \prod_{i=1}^{n} \frac{e^{-(y_i-\mu)^2/2\sigma_i^2}}{\sigma_i}$$

Rather than maximize L we usually maximize $\log L$.

$$\log L(\vec{y} \mid \mu) = \text{const} + \sum_{i=1}^{n} -\frac{(y_i-\mu)^2}{2\sigma_i^2}$$

$$= \text{const} - \frac{1}{2}\chi^2$$

In this particular example, minimizing χ^2 is equivalent to minimizing $-2\log L$.

χ^2 can be interpreted as a goodness of fit statistic. $\log L$ does not indicate goodness of fit.

Nevertheless,

$$\sigma_\mu^2 \approx \left[\frac{\partial^2}{\partial \mu^2} \left(-2 \log L \right) \right]^{-1}$$

$$= \left[\sum \frac{1}{\sigma_i^2} \right]^{-1}$$

Poisson Counting:

Recall that when s events are expected, the probability of observing n events is

$$P(n;s) = \frac{s^n e^{-s}}{n!}$$

When s is large, $\langle n \rangle = s$

$$\mathrm{Var}(n) = s$$

$$P(n;s) \approx \frac{1}{\sqrt{2\pi s}} e^{-(n-s)^2/2s}$$

Estimate for s is n
Estimate for σ_s is $\sqrt{n}$.

What if s is not large? What if $n=0$?

$$P(n;s) = \frac{s^n e^{-s}}{n!}$$

Then
$$P(s|n) = \frac{\dfrac{s^n e^{-s}}{n!}\,\pi(s)}{\displaystyle\int_0^\infty \frac{s'^n e^{-s'}}{n!}\,\pi(s')\,ds'}$$

What should we use for $\pi(s)$?
Complete lack of knowledge or prior assumptions: uniform on $[0, \infty)$.

In practice, uniform over the region where

$$\frac{s'^{n} e^{-s'}}{n!} > \varepsilon$$ where ε is arbitrarily small.

$$P(s|n) \approx \frac{\frac{s^n e^{-s}}{n!}\, \pi(s)}{\pi(s) \int_0^\infty \frac{s'^n e^{-s'}}{n!}\, ds'}$$

Note that $\int_0^\infty x^n e^{-x}\, dx = \Gamma(n+1) = n!$

So $P(s|n) = \frac{s^n e^{-s}}{n!}$

When $n = 0$, $P(s|0) = e^{-s}$

most probable value is $s = 0$

$$P(s < s_{upper}) = \int_0^{s_{upper}} e^{-s}\, ds = 1-\alpha$$

$$\Rightarrow e^{-s_{upper}} = \alpha$$

When $1-\alpha = 90\%$ c.L.,

$s_{upper} = 2.303$

When $1-\alpha = 95\%$ c.L.,

$s_{upper} = 2.996$