Lectures 7: Maximum likelihood I.
(nonlinear least square fits)

χ^2 fitting procedure!

example: testing coin making machine

Model for motivating nonlinear least squares fitting ($\chi 2$ fitting)

Manufacturer prints coins noticing that the printing machine produces biased heads/tails with a fixed value of p for heads. p(x) is dependent on the machine temperature x. This p can be measured by tossing n coins from the batch and measuring the binomial probability p of the batch. For some plotting convenience of the analysis 2p - 0.4 is determined by measuring $2n_{head}/n - 0.4$ which turns out to be the function of the temperature where the machine operates (temperature x is recorded for the measurement). The results also depend on five parameters $b_1 \dots b_5$ of the mechanical construction of the printing machine. A smart theorist comes up with a model how the value of p depends on the temperature x and the five parameters $b_1 \dots b_5$:

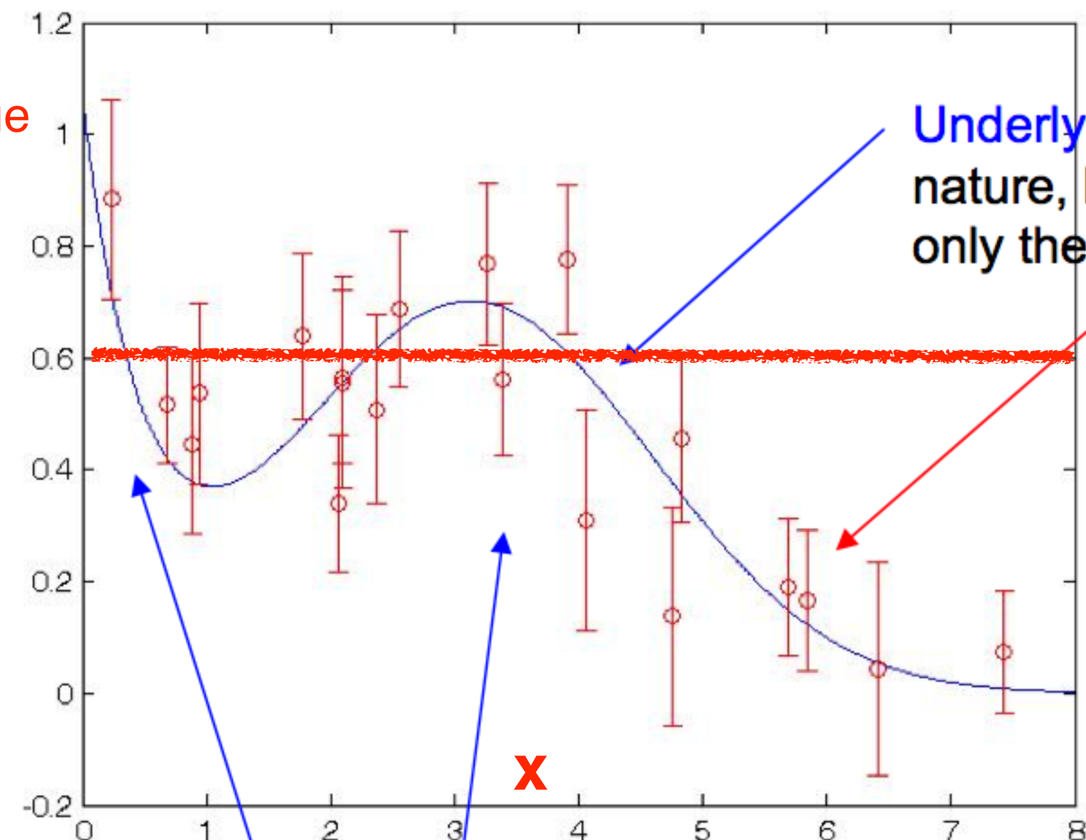
f(x)=2p-0.4 is the measured value of 2p-0.4 as a function of temperature x

$$f(x) = b_1 \exp(-b_2 x) + b_3 \exp\left(-\frac{1}{2}\frac{(x-b_4)^2}{b_5^2}\right)$$

Manufacturer wants to determine the parameters $b_1 \dots b_5$ so that they can operate the machine at the temperature where 2p - 0.4 = 0.6 so that p=0.5 and the coins are unbiased. This will require to fit the five parameters $b_1 \dots b_5$ of the machine based on the available data at many temperatures. **How do we do that?**

$$f(x) = b_1 \exp(-b_2 x) + b_3 \exp\left(-\frac{1}{2}\frac{(x-b_4)^2}{b_5^2}\right)$$

measured value of 2p-0.4 as a function of x



Underlying curve is known to nature, but not to us! We see only the red data points.

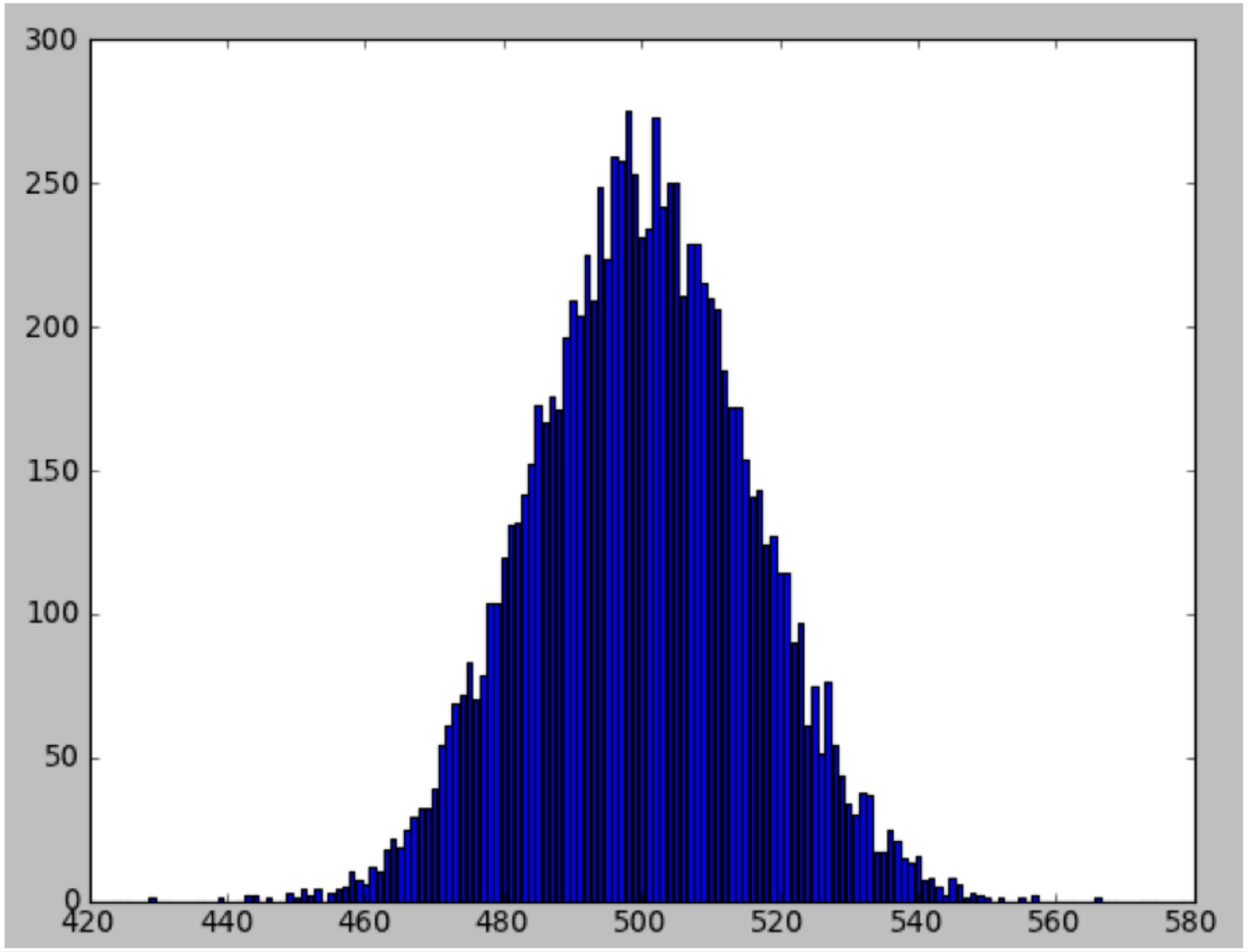
Fit 5 parameters from 20 irregularly space points, with normal errors of known standard deviations.

Can we do it? How well?

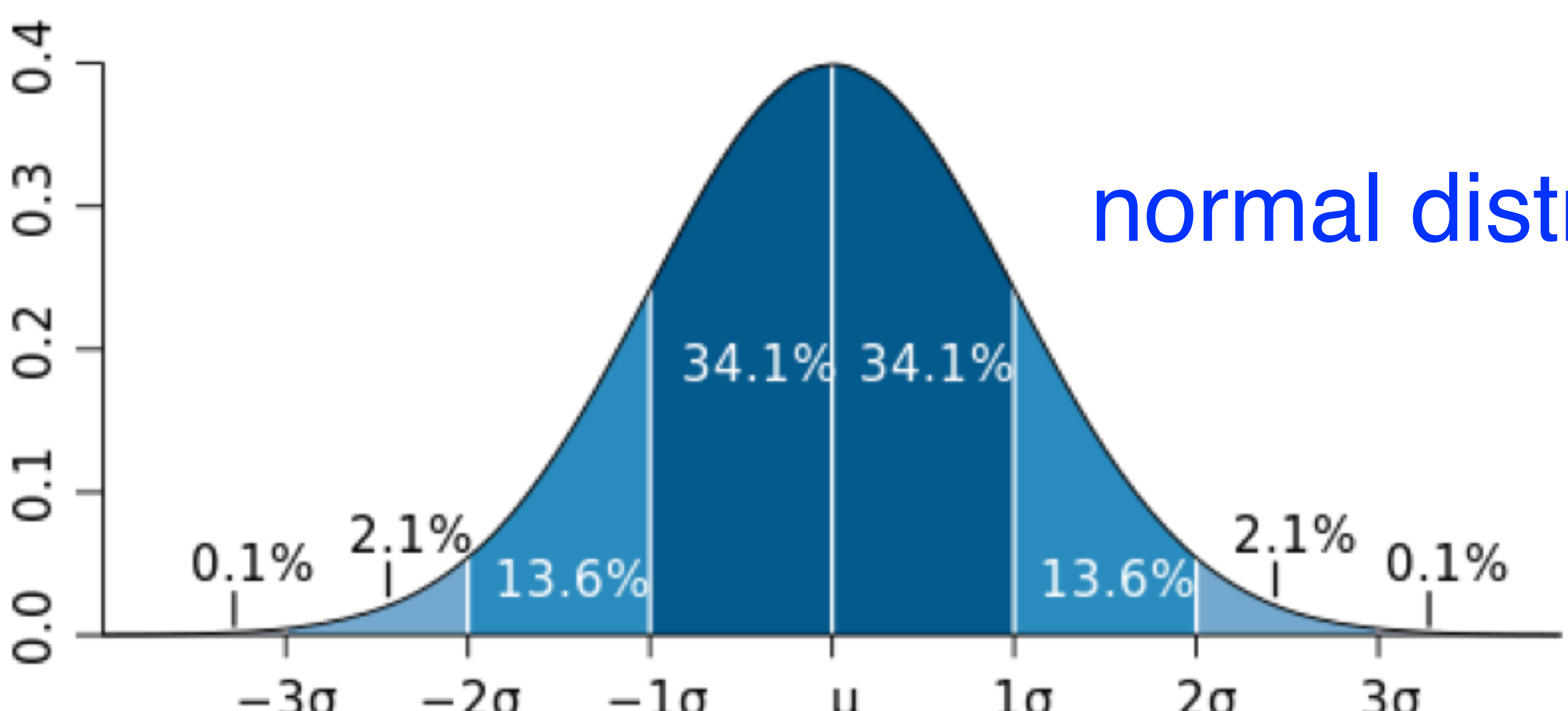
increasing temperature x in some arbitrary units

for example, this rise might be an instrumental or noise effect, while this bump might be what you are really interested in

central limit theorem:



10,000 trials of 1,000 tosses



normal distribution (bell curve)

Data are collected at various temperatures x_i.

At each temperature x_i the value $y_i = 2n^{(i)}_{heads}/n - 0.4$ is measured to approximate $2p - 0.4$ from n coin tosses

But y_i has some error e_i

What is the error?

Weighted Nonlinear Least Squares Fitting
a.k.a. χ^2 Fitting
a.k.a. Maximum Likelihood Estimation of Parameters (MLE)
a.k.a. Bayesian parameter estimation
(with uniform prior and maybe
some other normality assumptions)

these are not all exactly identical,
but they're real close!

$$y_i = y(\mathbf{x}_i|\mathbf{b}) + e_i$$

measured values supposed to be a model, plus an error term

$$e_i \sim N(0, \sigma_i)$$

the errors are Normal, either independently

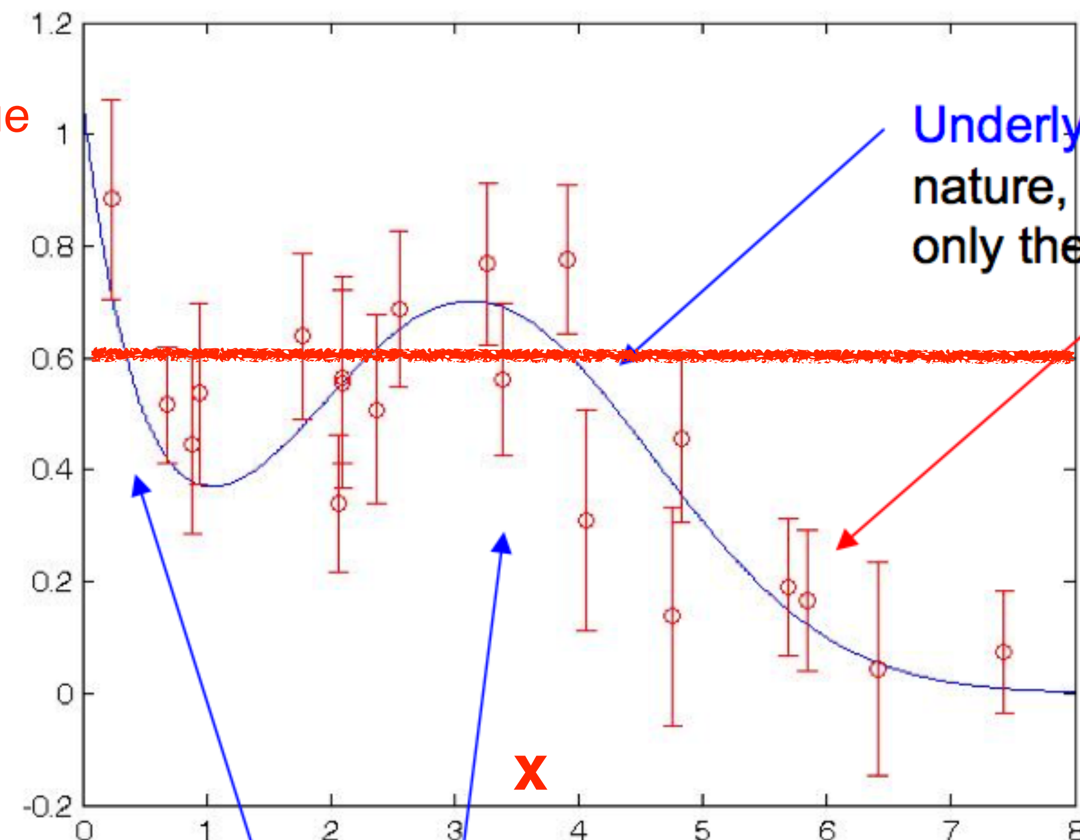
$$\mathbf{e} \sim N(0, \mathbf{\Sigma})$$

or else with errors correlated in some known way (e.g., multivariate Normal)

We want to find the parameters of the model **b** from the data.

$$f(x) = b_1 \exp(-b_2 x) + b_3 \exp\left(-\frac{1}{2}\frac{(x-b_4)^2}{b_5^2}\right)$$

measured value of 2p-0.4 as a function of x



Underlying curve is known to nature, but not to us! We see only the red data points.

increasing temperature x in some arbitrary units

for example, this rise might be an instrumental or noise effect, while this bump might be what you are really interested in

Fit 5 parameters from 20 irregularly space points, with normal errors of known standard deviations.

Can we do it? How well?

Maximum Likelihood discussion

Fitting is usually presented in frequentist, MLE language.
But one can equally well think of it as Bayesian:

$$
\begin{aligned}
P(\mathbf{b}|\{y_i\}) &\propto P(\{y_i\}|\mathbf{b})P(\mathbf{b}) \\
&\propto \prod_i \exp\left[-\tfrac{1}{2}\left(\frac{y_i - y(\mathbf{x}_i|\mathbf{b})}{\sigma_i}\right)^2\right] P(\mathbf{b}) \\
&\propto \exp\left[-\tfrac{1}{2}\sum_i \left(\frac{y_i - y(\mathbf{x}_i|\mathbf{b})}{\sigma_i}\right)^2\right] P(\mathbf{b}) \\
&\propto \exp[-\tfrac{1}{2}\chi^2(\mathbf{b})]P(\mathbf{b})
\end{aligned}
$$

Now the idea is: Find (somehow!) the parameter value $\mathbf{b}_0$ that minimizes χ^2 .

For linear models, you can solve linear "normal equations" or, better, use Singular Value Decomposition. See NR3 section 15.4

In the general nonlinear case, you have a general minimization problem, for which there are various algorithms, none perfect.

Those parameters are the MLE. (So it is Bayes with uniform prior.)

Maximum Likelihood discussion

Nonlinear fits are often easy in MATLAB (or other high-level languages) **if** you can make a reasonable starting guess for the parameters:

$$y(x|\mathbf{b}) = b_1 \exp(-b_2 x) + b_3 \exp\left(-\frac{1}{2}\frac{(x-b_4)^2}{b_5^2}\right)$$

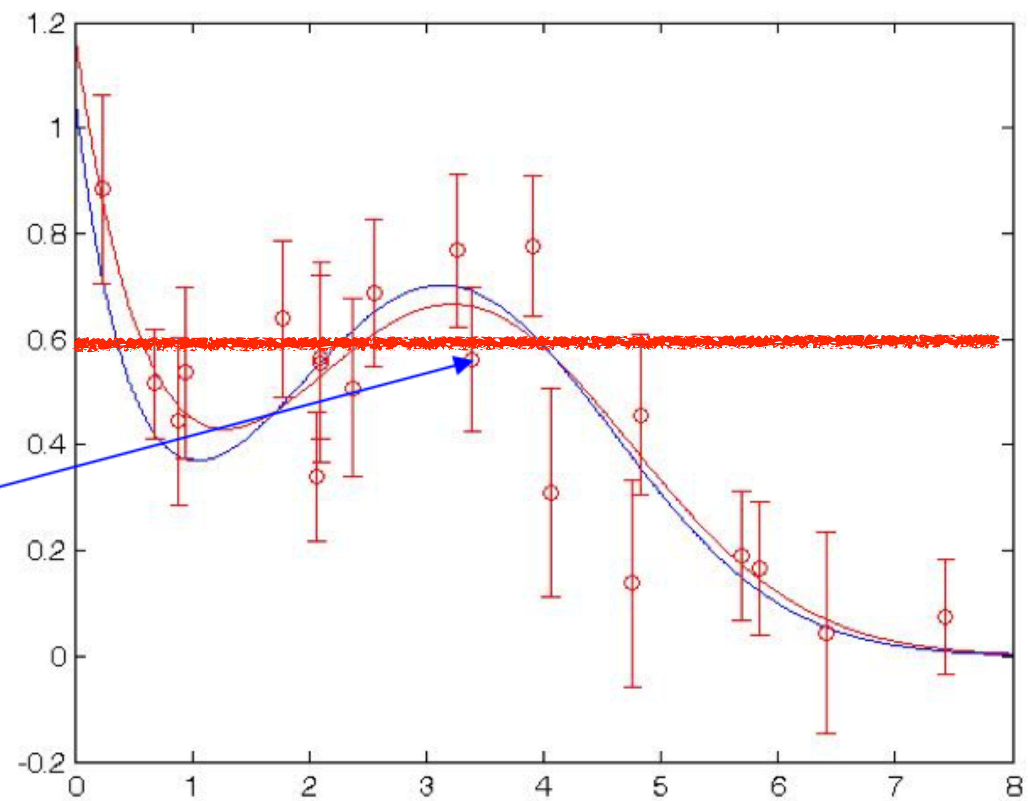
$$\chi^2 = \sum_i \left(\frac{y_i - y(x_i|\mathbf{b})}{\sigma_i}\right)^2$$

```
ymodel = @(x,b) b(1)*exp(-b(2)*x)+b(3)*exp(-(1/2)*((x-b(4))/b(5)).^2)
chisqfun = @(b) sum(((ymodel(x,b)-y)./sig).^2)
```

```
bguess = [1 2 .5 3 1.5]
bfit = fminsearch(chisqfun,bguess)
xfit = (0:0.01:8);
yfit = ymodel(xfit,bfit);
```

```
bfit = 1.1235    1.5210    0.6582
    3.2654    1.4832
```

Suppose that what we really care about is the area of the bump, and that the other parameters are "nuisance parameters".



increasing temperature x in some arbitrary units

Maximum Likelihood parameter errors?

How accurately are the fitted parameters determined?
As Bayesians, we would **instead** say, what is their posterior distribution?

Taylor series:

$$-\tfrac{1}{2}\chi^2(\mathbf{b}) \approx -\tfrac{1}{2}\chi^2_{\min} - \tfrac{1}{2}(\mathbf{b}-\mathbf{b}_0)^T \left[\tfrac{1}{2}\frac{\partial^2\chi^2}{\partial\mathbf{b}\partial\mathbf{b}}\right](\mathbf{b}-\mathbf{b}_0)$$

So, while exploring the χ^2 surface to find its minimum, we must also calculate the Hessian (2nd derivative) matrix at the minimum.

Then

$$P(\mathbf{b}|\{y_i\}) \propto \exp\left[-\tfrac{1}{2}(\mathbf{b}-\mathbf{b}_0)^T\Sigma_b^{-1}(\mathbf{b}-\mathbf{b}_0)\right]P(\mathbf{b})$$

with

$$\Sigma_b = \left[\tfrac{1}{2}\frac{\partial^2\chi^2}{\partial\mathbf{b}\partial\mathbf{b}}\right]^{-1}$$

covariance (or "standard error") matrix of the fitted parameters

Notice that if (i) the Taylor series converges rapidly and (ii) the prior is uniform, then the posterior distribution of the $\mathbf{b}$'s is multivariate Normal, a very useful CLT-ish result!

CLT: confidence level test