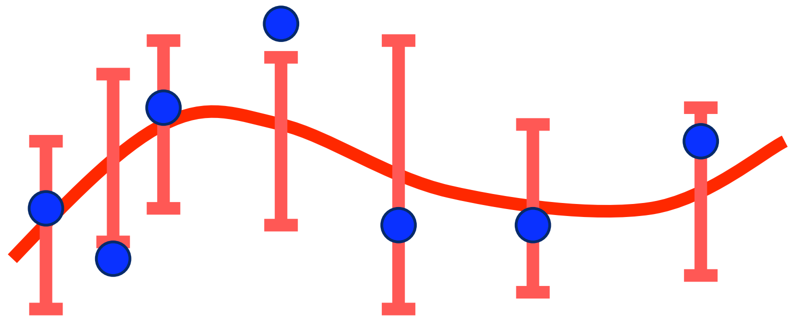
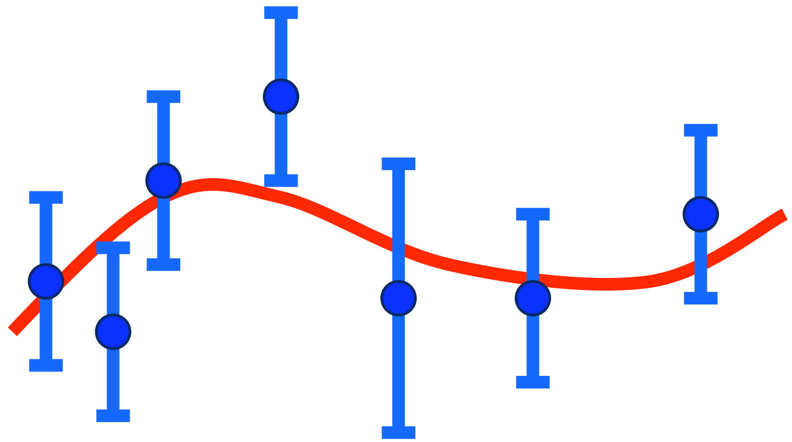
Error Bars live with the Model



Not with the Data

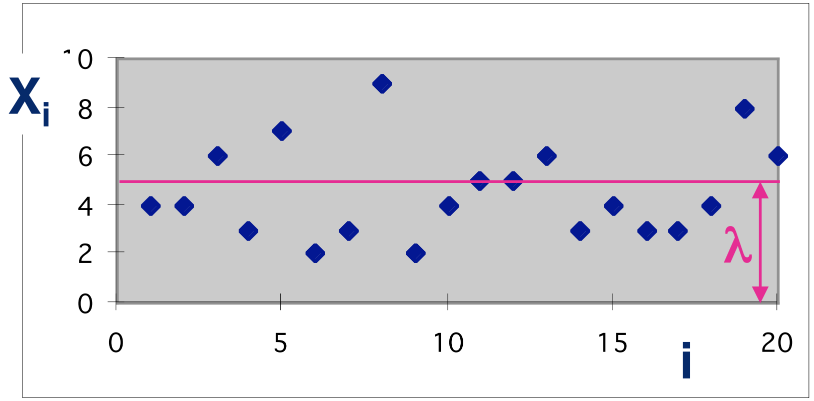


Usually the distinction is unimportant.

But sometimes ***it is important***.

Error bars live with the **model**, not the data!

Example: **Poisson data:**



$$\text{Prob}(x = n \mid \lambda) = \frac{\lambda^n e^{-\lambda}}{n!} \qquad n = 0, 1, 2, \ldots$$

$$\langle X_i \rangle = \lambda, \qquad \sigma^2(X_i) = \lambda$$

How to attach error bars to the data points?

The **wrong way**: If $\sigma(X_i) = \sqrt{X_i}$, then $1/\sigma^2 = \infty$ when $X_i = 0$

$$\text{and} \quad \hat{X} \equiv \frac{\sum_i X_i/\sigma_i^2}{\sum_i 1/\sigma_i^2} = \frac{0 \cdot \infty}{\infty} = 0 \text{ , clearly wrong !}$$

Assigning $\sigma(X_i) = \sqrt{X_i}$ gives a **downward bias**. Points lower than average by chance are given smaller error bars, and hence more weight than they deserve.

The **right way**:

Assign $\sigma = \sqrt{\lambda}$, where λ = mean count rate ***predicted by the model.***

Conditional Probabilities

P(X,Y) = ***joint probability density*** of *X* and *Y*

P(X) = projection of *P(X,Y)* onto *X* axis.

$$\boxed{P(X) = \int P(X,Y)\, \mathrm{d}Y}$$

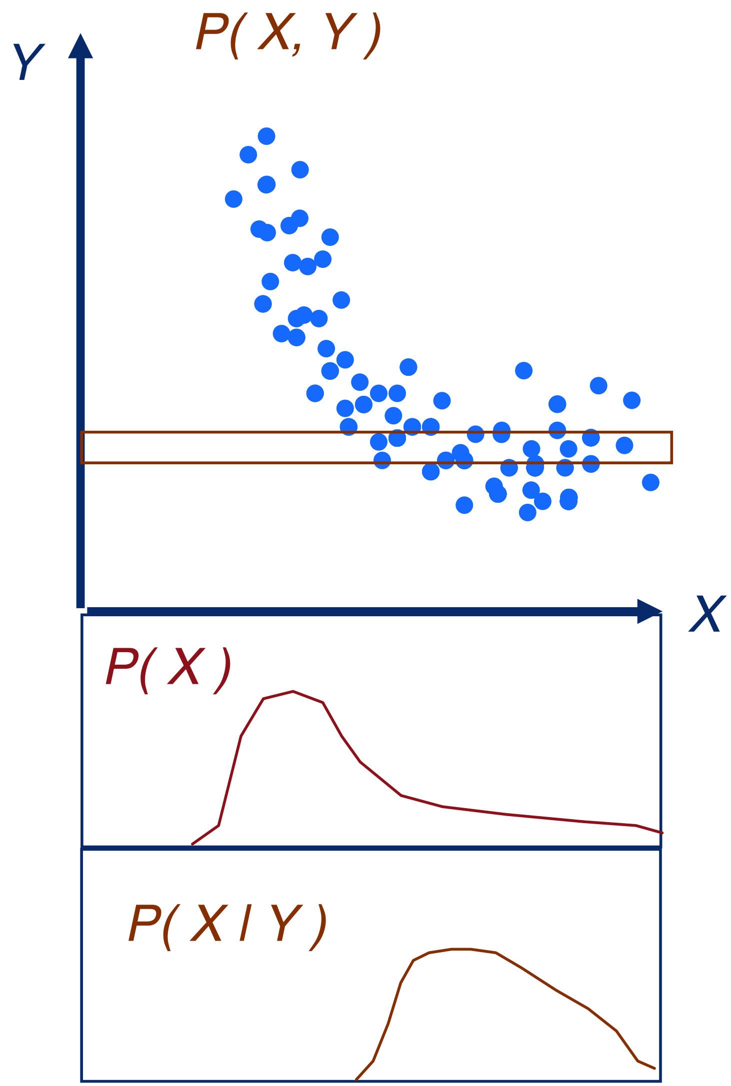
Conditional Probability:

P(X | Y) = "probability of *X* given *Y*"

= "normalised slice" of *P(X,Y)*

at a **fixed value** of *Y*.

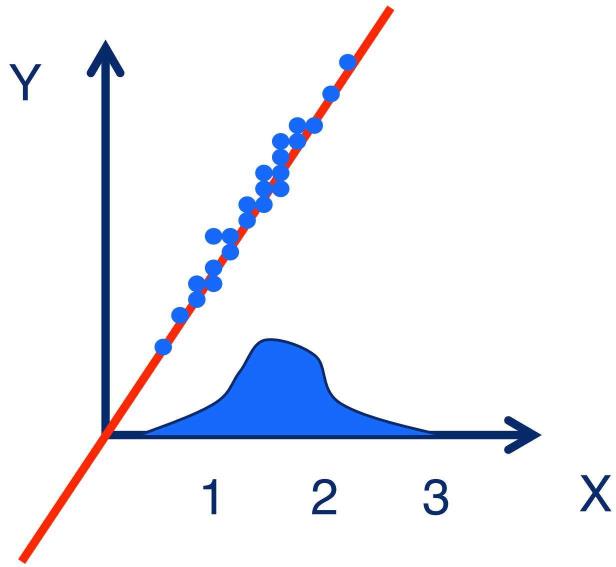
$$\boxed{P(X\,|\,Y) \equiv \frac{P(X,Y)}{P(Y)} = \frac{P(X,Y)}{\int P(X,Y)\, \mathrm{d}X}}$$



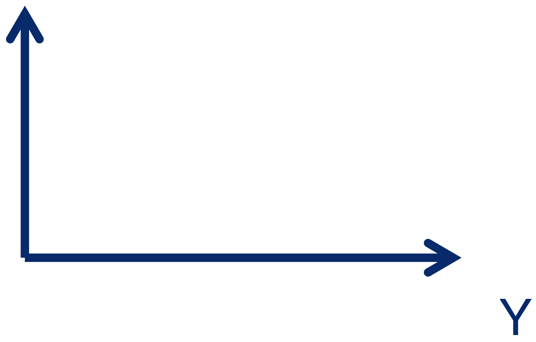
Test Understanding

Y = 3 X

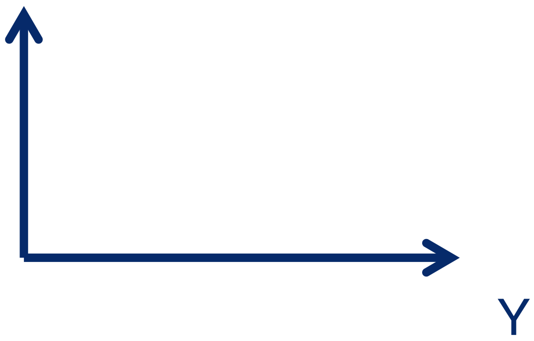
X = Gaussian



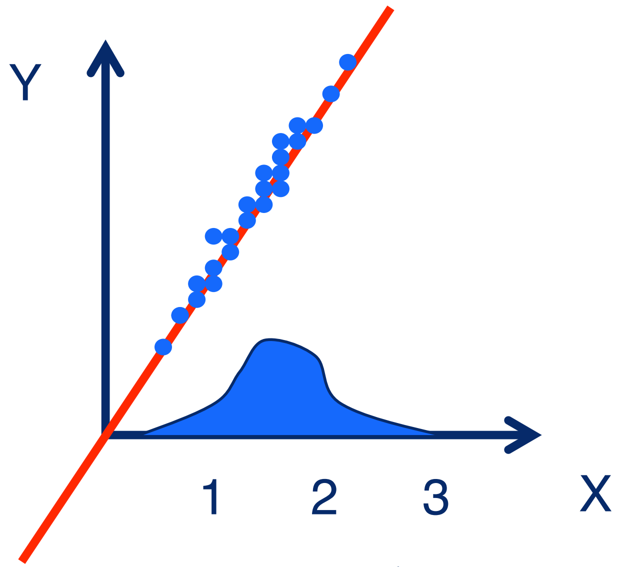
P(Y I X = 2) = ?



P(Y I X > 2) = ?

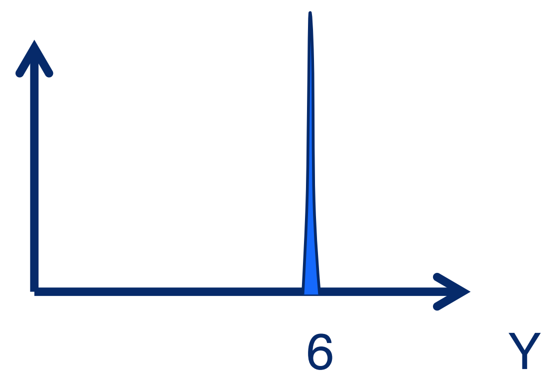


Test Understanding

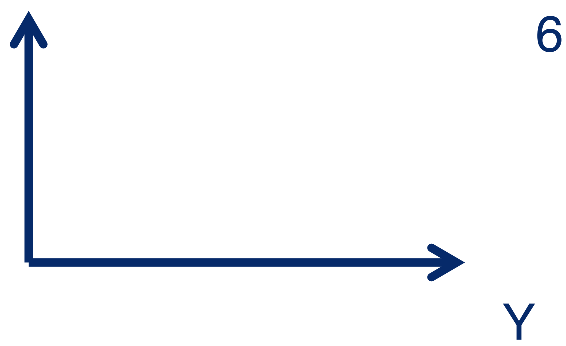


$Y = 3X$

X = Gaussian



$P(Y \mid X = 2) = ?$

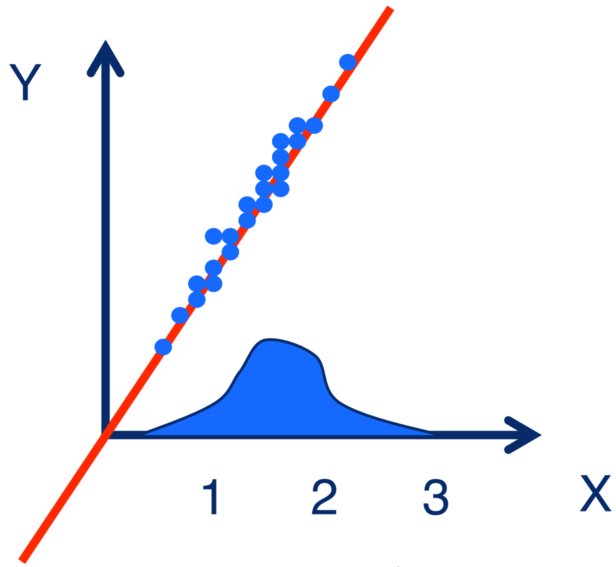


$P(Y \mid X > 2) = ?$

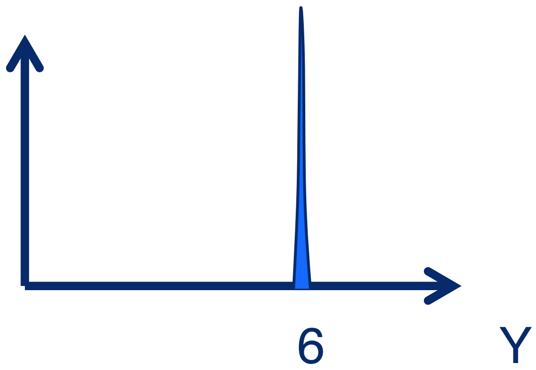
Test Understanding

Y = 3 X

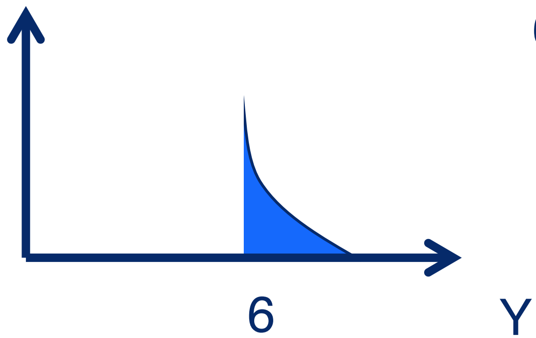
X = Gaussian



P(Y | X = 2) = ?



P(Y | X > 2) = ?

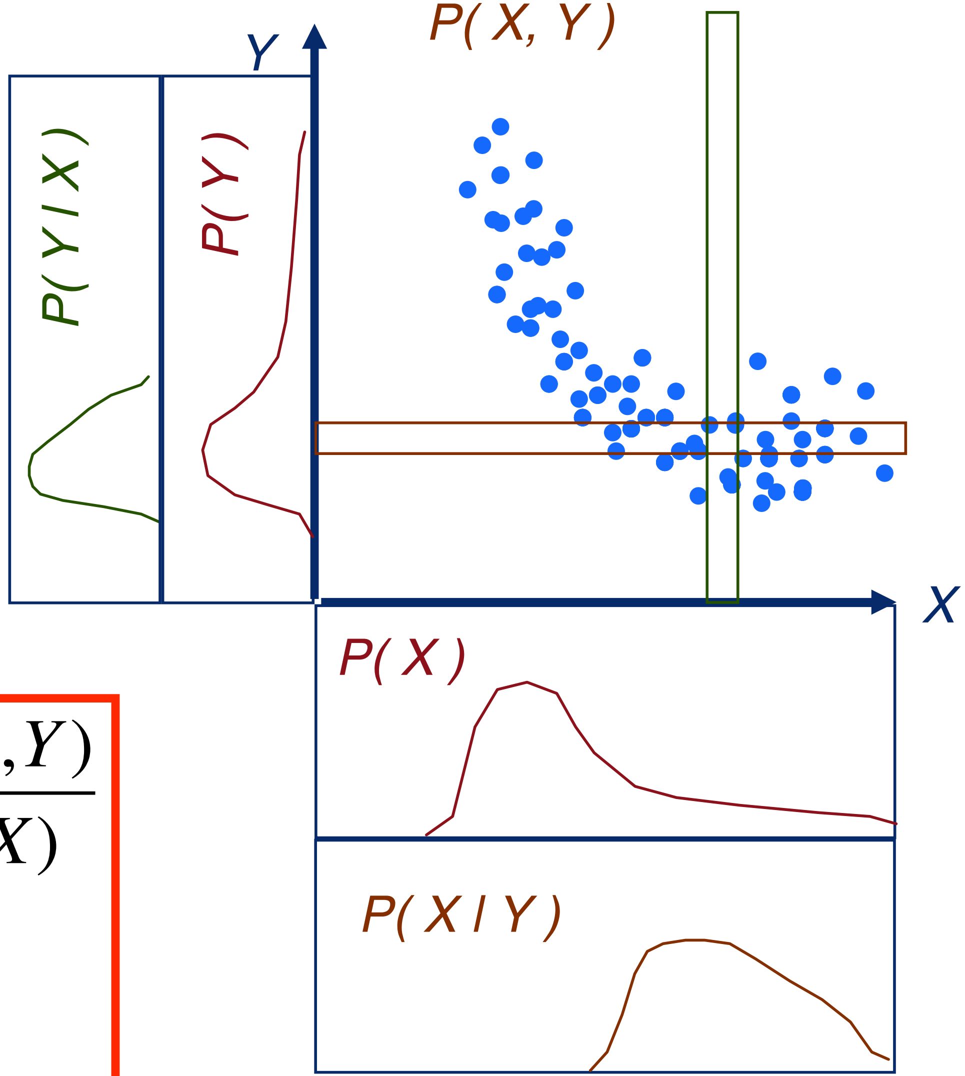


Conditional Probabilities

P(X) = projection onto *X* axis.
P(Y) = projection onto *Y* axis.

$$P(X) = \int P(X,Y)\, \mathrm{d}Y$$
$$P(Y) = \int P(X,Y)\, \mathrm{d}X$$

Conditional Probability:
P(X / Y) = normalised slice at fixed Y
P(Y | X) = normalised slice at fixed X



$$P(X \mid Y) \equiv \frac{P(X,Y)}{P(Y)} \qquad P(Y \mid X) \equiv \frac{P(X,Y)}{P(X)}$$
$$P(X,Y) = P(X \mid Y)\, P(Y)$$
$$= P(Y \mid X)\, P(X)$$

Bayes' Theorem and Bayesian Inference

Bayes' Theorem: $$P(X\,|\,Y) = \frac{P(Y\,|\,X)\,P(X)}{P(Y)}$$

Since $\quad P(X,Y) = P(X\,|\,Y)\,P(Y) = P(Y\,|\,X)\,P(X)$

then $$P(X\,|\,Y) = \frac{P(Y\,|\,X)\,P(X)}{P(Y)} = \frac{P(Y\,|\,X)\,P(X)}{\int P(Y\,|\,X)\,P(X)\,\mathrm{d}X}$$

Bayesian Inference :

$$P(\text{model}\,|\,\text{data}) = \frac{P(\text{data}\,|\,\text{model})\,P(\text{model})}{P(\text{data})}$$

Shows us how to change our probability distribution over various models in light of new data.

Inferences depend on Prior, not just Data

Bayesian inference: (M = model, D = data)

Posterior Probability = (Likelihood × Prior Probability) / Evidence

$$P(M \mid D) = \frac{P(D \mid M)\, P(M)}{P(D)} = \frac{P(D \mid M)\, P(M)}{\int P(D \mid M) P(M) \mathrm{d}M}$$

Relative probability of two models M_1 and M_2 :

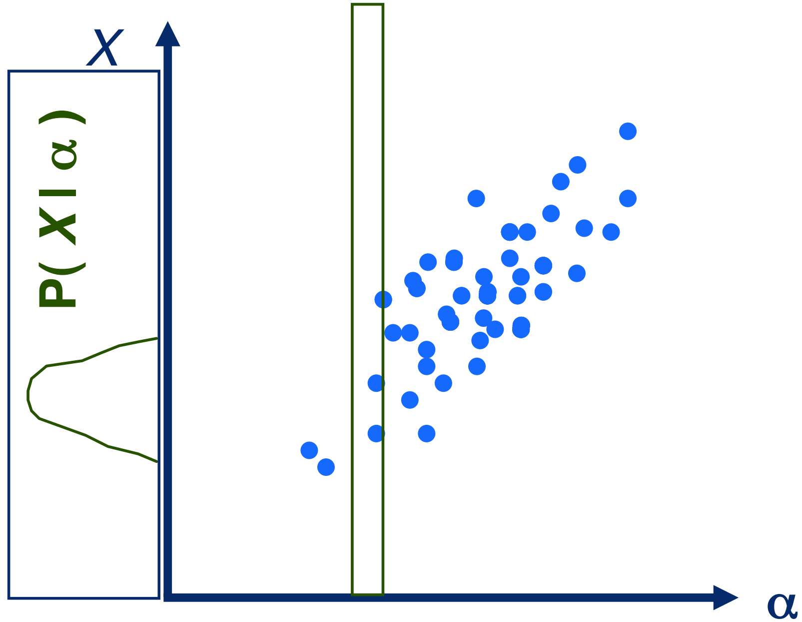
$$\frac{P(M_1 \mid D)}{P(M_2 \mid D)} = \frac{P(D \mid M_1)}{P(D \mid M_2)} \times \frac{P(M_1)}{P(M_2)} \approx \exp\left(\frac{-\Delta\chi^2}{2}\right) \times \frac{P(M_1)}{P(M_2)}$$

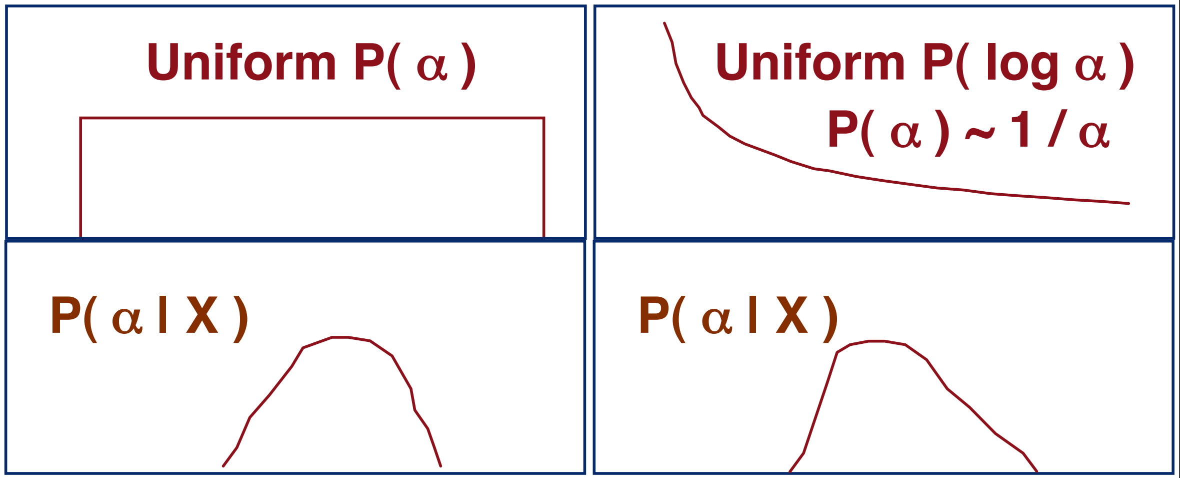
- The ***Likelihood***, P(data I model), is quantified by a ***"badness-of-fit"*** statistic. e.g. P(data I model) ~ $\exp(-\chi^2/2)$
- The ***Prior,*** P(model) expresses your ***prejudice*** (prior knowledge).
- The ***Posterior,*** P(model I data), gives your ***inference***, the relative probabilities of different models (parameters), in light of the data.

No absolute inferences ! New data **changes** your prior knowledge, but **your conclusions always also depend on your prior**.

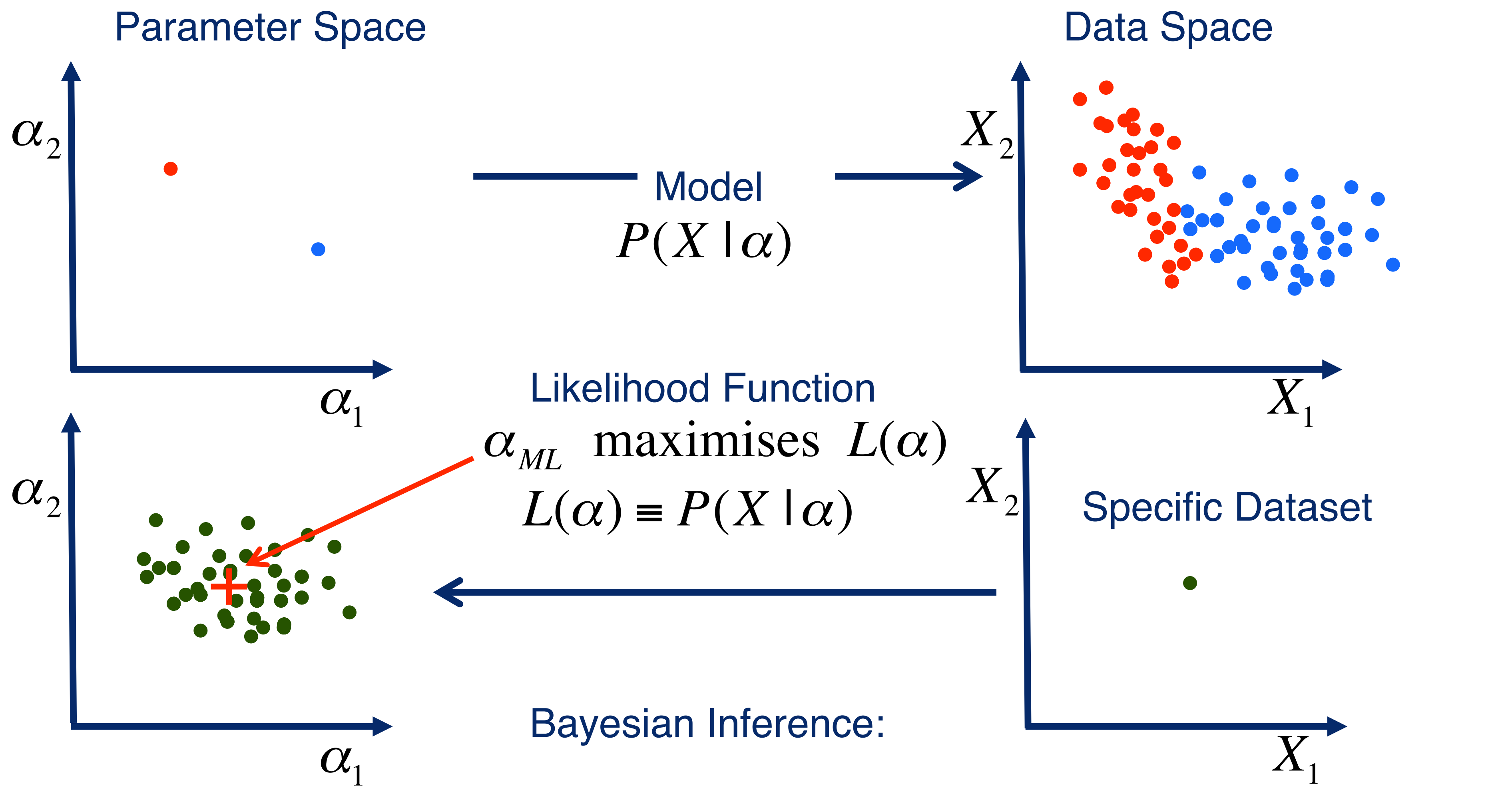
Choice of Prior

- A model for a set of data X depends on a parameter α.
- Our knowledge of α before measuring X is quantified by the **prior** p.d.f. $P(\alpha)$.
- Choice of $P(\alpha)$ is arbitrary subject to common sense!
- After measuring X, Bayes theorem gives **posterior** p.d.f.
- $P(\,\alpha \mid X\,) \sim P(\,X \mid \alpha\,)\,P(\alpha)$
- Different priors $P(\alpha)$ lead to different **inferences** $P(\,\alpha \mid X\,)$





Max Likelihood and Bayesian Inference



$$P(\alpha \mid X) = \frac{P(X \mid \alpha)\, P(\alpha)}{\int P(X \mid \alpha)\, P(\alpha)\, \mathrm{d}\alpha} \propto L(\alpha)\, P(\alpha)$$

Posterior Probability

Likelihood modifies the Prior.

Gaussian Datum with Uniform Prior

$\text{Data}: X \pm \sigma \qquad \text{Model parameter}: \mu$

Likelihood function :

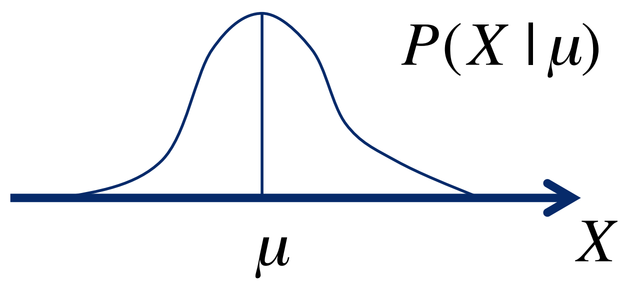
$$L(\mu) \equiv P(X \mid \mu) \quad = \frac{e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}}{\sqrt{2\pi}\ \sigma}$$

$\mu_{ML} = X$ maximises $L(\mu)$.

Posterior probability :

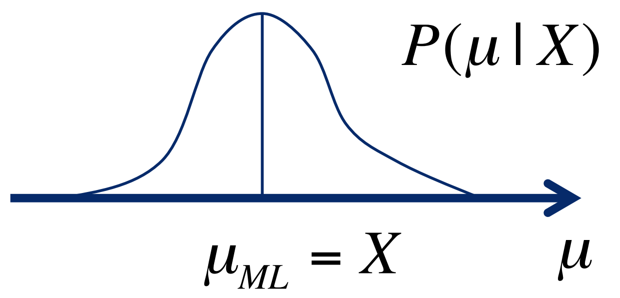
$$P(\mu \mid X) = \frac{P(X \mid \mu)\ P(\mu)}{P(X)}$$

$$P(X) = \int P(X \mid \mu)\ P(\mu)\ \mathrm{d}\mu$$



Uniform prior:

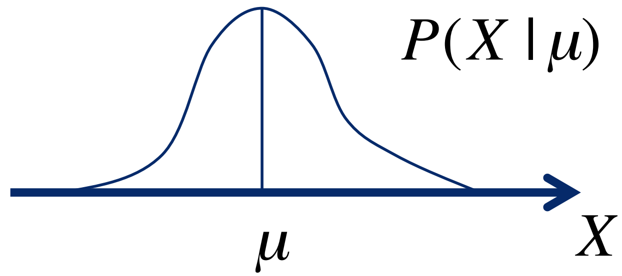
$$P(\mu) = \text{constant}$$



Maximum Likelihood implicitly assumes a Uniform Prior

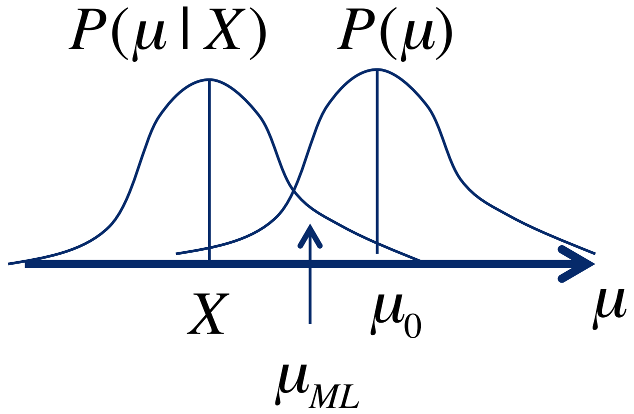
Gaussian Datum with Gaussian Prior

Data : $X \pm \sigma$



Likelihood : $L(\mu) = P(X \mid \mu) \quad = \dfrac{e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}}{\sqrt{2\pi}\,\sigma}$

Likelihood: Gaussian prior:



Prior : $P(\mu) = \dfrac{e^{-\frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2}}{\sqrt{2\pi}\,\sigma}$

Posterior : $P(\mu \mid X) \propto P(X \mid \mu)\, P(\mu)$

$$\propto e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} e^{-\frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2} \quad = e^{-\frac{1}{2}\left(\frac{\mu-\mu_{ML}}{\sigma(\mu_{ML})}\right)^2}$$

$$\mu_{ML} = \frac{\dfrac{X}{\sigma^2} + \dfrac{\mu_0}{\sigma_0^{\,2}}}{\dfrac{1}{\sigma^2} + \dfrac{1}{\sigma_0^{\,2}}} \qquad \sigma^2(\mu_{ML}) = \frac{1}{\dfrac{1}{\sigma^2} + \dfrac{1}{\sigma_0^{\,2}}}$$

Same as Optimal Average !

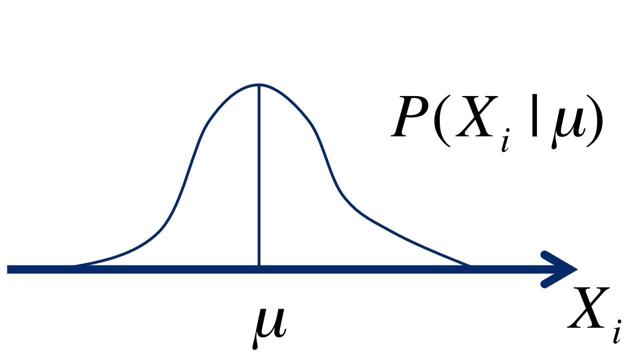
Gaussian prior acts like 1 more data point.

Data pulls the probability away from the prior, and vice-versa.

Verify this result.

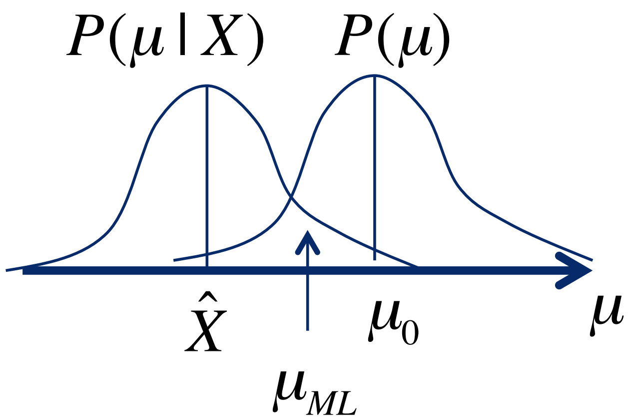
Gaussian Data with Gaussian Prior

Likelihood: $P(X\,|\,\mu) = \dfrac{e^{-\frac{1}{2}\sum_i\left(\frac{X_i-\mu}{\sigma_i}\right)^2}}{(2\pi)^{N/2}\prod_i \sigma_i} = \dfrac{\exp\left(-\dfrac{\chi^2}{2}\right)}{(2\pi)^{N/2}\prod_i \sigma_i}$



Prior: $P(\mu) = \dfrac{e^{-\frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2}}{\sqrt{2\pi}\;\sigma}$

Likelihood: Gaussian prior:



Posterior: $P(\mu\,|\,X) \propto P(X\,|\,\mu)\; P(\mu)$

$$P(\mu\,|\,X) \propto e^{-\chi^2/2}\; e^{-\frac{1}{2}\left(\frac{\mu-\mu_0}{\sigma_0}\right)^2} = e^{-\frac{1}{2}\left(\frac{\mu-\mu_{ML}}{\sigma(\mu_{ML})}\right)^2}$$

$$\mu_{ML} = \frac{\sum_i \dfrac{X_i}{\sigma_i^{\,2}} + \dfrac{\mu_0}{\sigma_0^{\,2}}}{\sum_i \dfrac{1}{\sigma_i^{\,2}} + \dfrac{1}{\sigma_0^{\,2}}} \qquad \sigma^2(\mu_{ML}) = \frac{1}{\sum_i \dfrac{1}{\sigma_i^{\,2}} + \dfrac{1}{\sigma_0^{\,2}}}$$

Same as Optimal Average !

Gaussian prior acts like 1 more data point.

Max Likelihood for Gaussian Data

Likelihood of parameters α for a given dataset:

$$L(\alpha) \equiv P(X \mid \alpha) = P(X_1 \mid \alpha) \times P(X_2 \mid \alpha) \times \ldots \times P(X_N \mid \alpha)$$

$$\equiv \prod_{i=1}^{N} P(X_i \mid \alpha)$$

For Gaussian error distributions:

$$P(X_i \mid \alpha) \quad = \quad \frac{e^{-\frac{1}{2}\left(\frac{X_i - \mu_i(\alpha)}{\sigma_i{}^2}\right)^2}}{\sqrt{2\pi}\,\sigma_i}$$

$$L(\alpha) = e^{-\chi^2/2} \left(\prod_{i=1}^{N} \frac{1}{\sigma_i} \right) (2\pi)^{-N/2}$$

$$-2\ln L = \chi^2 + 2\sum_i \ln\sigma_i + N\ln(2\pi)$$

$$\boxed{\text{To maximise } L(\alpha), \text{ minimise } \chi^2 + 2\sum_i \ln\sigma_i}$$

Maximum Likelihood Parameter Estimation

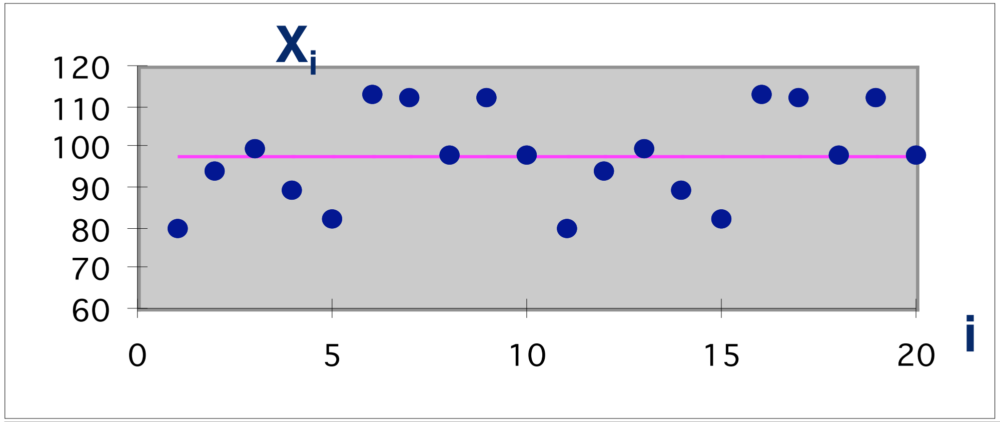
$$\boxed{\begin{aligned} &\alpha_{\mathrm{ML}} \text{ satisfies } 0 = \frac{\partial}{\partial\alpha}[-2\ln L(\alpha)], \\ &\mathrm{Var}[\alpha_{\mathrm{ML}}] \approx \frac{2}{\left(\frac{\partial^2}{\partial\alpha^2}[-2\ln L(\alpha)] \right)_{\alpha=\alpha_{ML}}} \end{aligned}}$$

Generalises χ^2 fitting.

Need ML when Parameters alter Error Bars

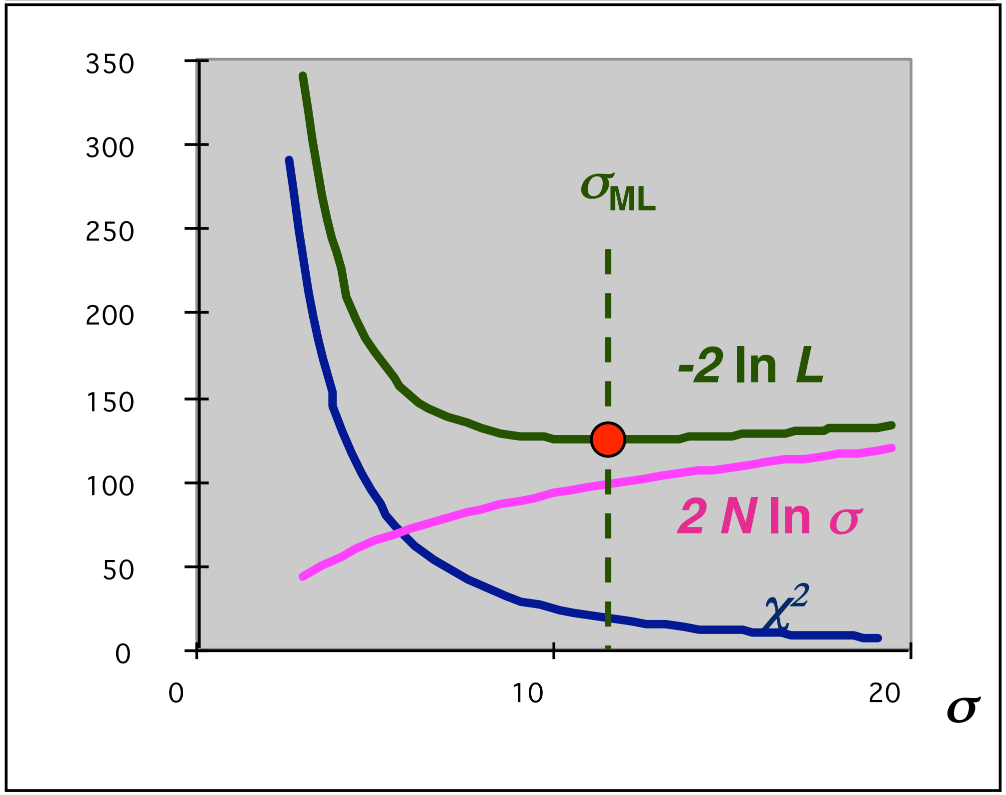
- Data points X_i with no errors:

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{X_i - \mu}{\sigma} \right)^2$$



- To find μ, minimise χ^2.
- To find σ , minimising χ^2 fails!

$$\chi^2 \rightarrow 0 \quad \text{as } \sigma \rightarrow \infty$$



- ML method minimises

$$\boxed{-2 \ln L = \chi^2 + 2\, N \ln \sigma}$$

Need ML to fit low-count Poisson Data

Poisson data X with rate parameter λ :

$$P(X \mid \lambda) = \frac{e^{-\lambda}\lambda^X}{X\,!}$$

Likelihood for N Poisson data points :

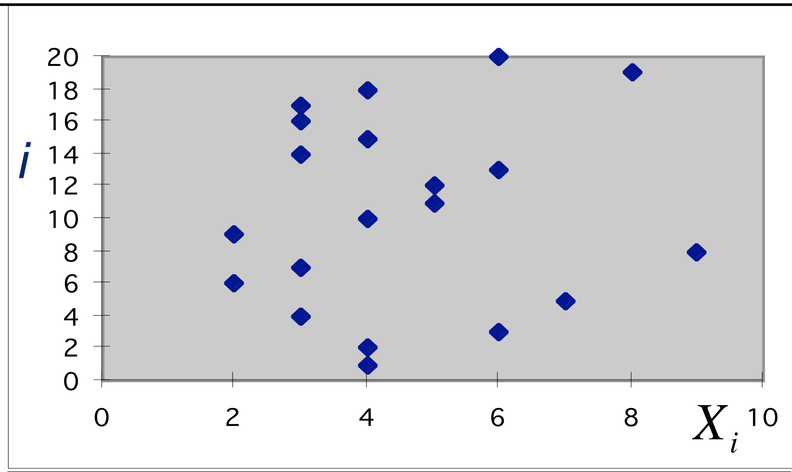
$$L(\lambda) = \prod_{i=1}^{N} P(X_i \mid \lambda) = \prod_{i=1}^{N} \frac{e^{-\lambda}\lambda^{X_i}}{X_i\,!}$$

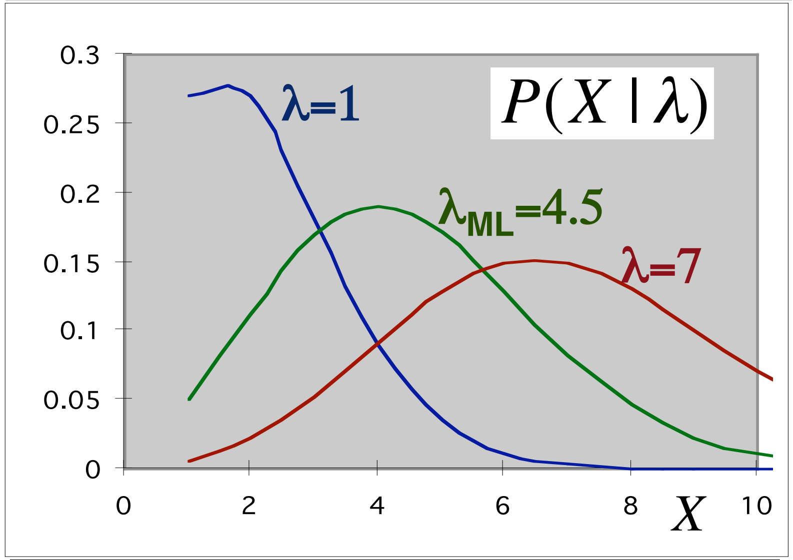
$$\ln L = \sum_i \left(-\lambda + X_i \ln\lambda - \ln X_i\,! \right)$$

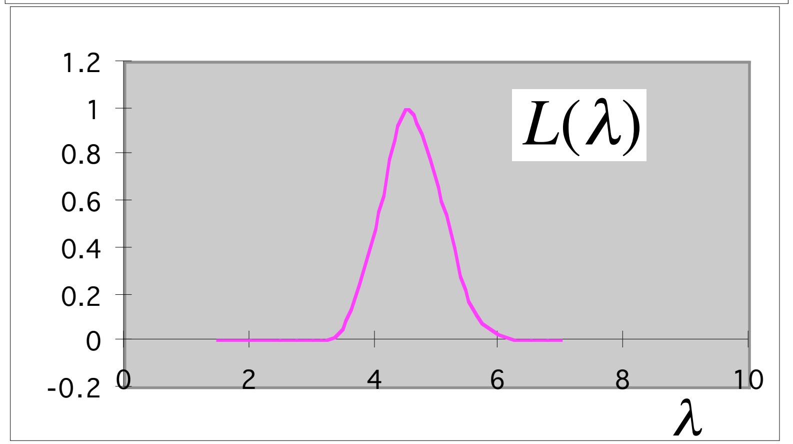
Maximum likelihood estimator of λ :

$$\frac{\partial \ln L}{\partial \lambda} = -N + \frac{1}{\lambda}\sum_i X_i = 0 \quad \text{at} \quad \lambda = \lambda_{ML}$$

$$\therefore \quad \lambda_{ML} = \frac{1}{N}\sum_i X_i\,.$$







Summary:

1. Error bars live with the Model, not with the Data.
2. Bayes Theorem (Bayesian Inference)

$$P(\text{model} \mid \text{data}) = \frac{P(\text{data} \mid \text{model})\, P(\text{model})}{P(\text{data})}$$

3. Maximum Likelihood, $L(\text{model}) \equiv P(\text{data} \mid \text{model})$
   e.g. for Gaussian Data:

$$-2\ln L = \chi^2 + 2\sum_{i=1}^{N} \ln\sigma_i + const$$

4. Use χ^2 if Gaussian errors and known σ_i.
5. Otherwise, use Maximum Likelihood,
   e.g. Error bars not known, or low-count Poisson data.
6. or full Bayesian analysis, incuding the prior:
   e.g. for Gaussian Data:

$$-2\ln P(\text{model} \mid \text{data}) = \chi^2 + 2\sum_{i=1}^{N} \ln\sigma_i - 2\ln P(\text{model}) + const$$