Statistics III

Atreyee Sinha

Topics today

- Error propagation
- Parameter estimation:
  - We saw many different distributions for random variables, and all of those distributions had parameters. How to obtain those parameters from the data
  - Maximum likelihood estimation
  - Maxima à Posteriori (MAP)

Parameter estimations

Maximum likelihood Estimator

- For which parameter values (θ_i) does the observed data (x_i) have the biggest probability?
  - Single value for the unknown parameter
  - Maximise $L(\theta) = \Pi_i^n f(X_i | \theta)$
  - If X_i are not independent, L is the joint probability.
  - For different values, likelihood of the data is different. For the "correct" value, likelihood is maximum.
    - $\hat{\theta} = argmax(L(\theta))$
  - Often simpler to compute the log of L
    - $LL(0) = logL(\theta) = log\Pi_{i=1}^N f(X_i | \theta) = \Sigma_{i=1}^N logf(X_i | \theta)$
    - Instead of maximising LL, sometimes -LL (NLL) is minimised

Bernoulli MLE

- Our unknown parameter here is ***p***

$$f(X_i \mid p) = p^{X_i}(1-p)^{1-X_i}$$

$$L(p) = \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i} = p^{\sum X_i}(1-p)^{n-\sum X_i}$$

$$\boxed{\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i}$$

$$\ln L(p) = \left(\sum_{i=1}^{n} X_i\right)\ln(p) + \left(n - \sum_{i=1}^{n} X_i\right)\ln(1-p)$$

$$\frac{d}{dp}[\ln L(p)] = \frac{\sum X_i}{p} - \frac{n - \sum X_i}{1-p} = 0$$

Poisson MLE

- Parameter to find: Sample mean

$$f(X_i \mid \lambda) = \frac{e^{-\lambda}\lambda^{X_i}}{X_i!}$$

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\ln L(\lambda) = -n\lambda + \left(\sum_{i=1}^{n} X_i\right)\ln(\lambda) - \sum_{i=1}^{n}\ln(X_i!)$$

$$\frac{d}{d\lambda}[\ln L(\lambda)] = -n + \frac{\sum X_i}{\lambda} = 0$$

Gaussian MLE

- Here, we need to find two parameters, mean and variance

$$f(X_i \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_i - \mu)^2}{2\sigma^2} \right)$$

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$\ln L(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2$$

$$\frac{\partial}{\partial \mu}[\ln L(\mu, \sigma^2)] = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu) = 0 \Rightarrow \text{for the mean}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu})^2$$

$$\frac{\partial}{\partial \sigma^2}[\ln L(\mu, \sigma^2)] = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} (X_i - \mu)^2 = 0$$

Properties of a good estimator

- Unbiased:
  - expected value equals the true population parameter $\mathbb{E}[\hat{\theta}] = \theta$
  - Not true for variance at small N!
  - Asymptotically converges to the true value
- Asymptotically minimal variance
- Behaves well under transformations:
  - That is, if ˆp is the MLE for p and g is a one-to-one function, then g(ˆp) is the MLE for g(p). For example, if ˆσ is the MLE for the standard deviation σ then (ˆσ)2 is the MLE for the variance σ2

Mean Square Error

$$\begin{aligned}\mathrm{MSE}(\hat{\theta}) &= \langle(\hat{\theta}-\theta)^2\rangle \\ &= \langle\hat{\theta}^2+\theta^2-2\hat{\theta}\theta\rangle \\ &= \langle\hat{\theta}^2\rangle+\theta^2-2\langle\hat{\theta}\rangle\theta \\ &= \langle(\hat{\theta}-\langle\hat{\theta}\rangle)^2\rangle+(\langle\hat{\theta}\rangle-\theta)^2 \\ &= V[\hat{\theta}]+b^2\end{aligned}$$

- MSE: Sum of variance and bias
- Optimal: bias = 0

Variance of ML estimators

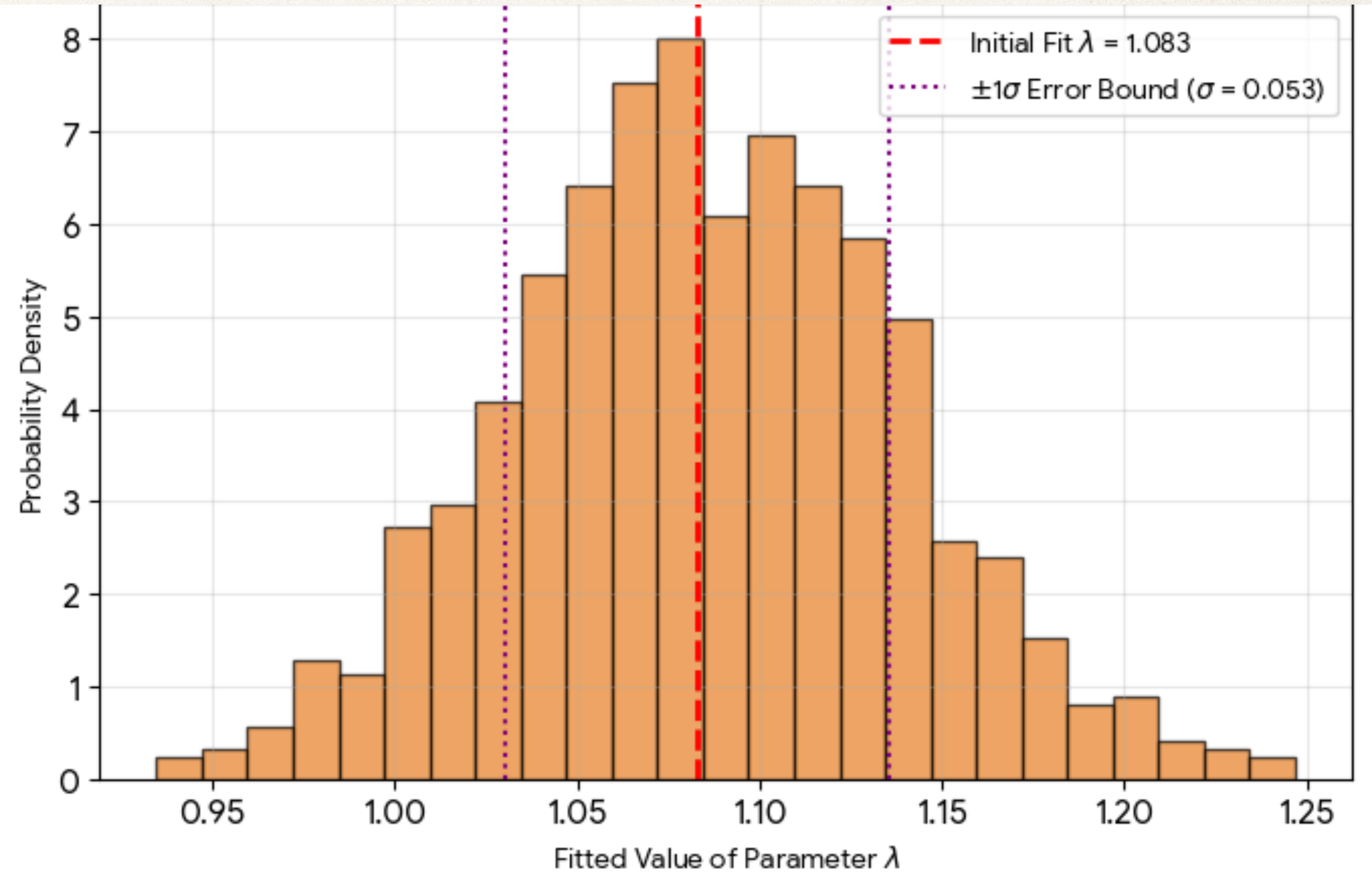
- We have found the optimal value of the parameters.
  - What is the statistical uncertainty on the estimates?
- Different ways:
  - Analytically
  - From the covariance matrix
  - Using Monte Carlo
  - From the likelihood profiles

Analytically

- For certain pdf, one can calculate the variance directly
- $\sigma^2 = E[X^2] - E[X]^2$
- Error on mean = $\sigma/\sqrt{n}$
- For a uniform distribution
    - $\mu = (b - a)/2$
    - $\sigma^2 = (b - a)/12$
    - $\delta\mu = (b - a)/\sqrt{12n}$

With Monte Carlo

- Computationally intensive
- Simulate a large number of experiments and look at the distribution of ML estimates from MC experiments.
  - Lets say you have a dataset from radioactive decay
  - You fit it with a Poisson and get $\tau = 1.08$ -> assume this as true
  - Now, simulate ~500 distributions with $\tau = 1.08$ and fit it again
  - Plot distribution of fit parameters —> get the error



$$\lambda = 1.08$$

$$\sigma_\lambda = 0.053$$

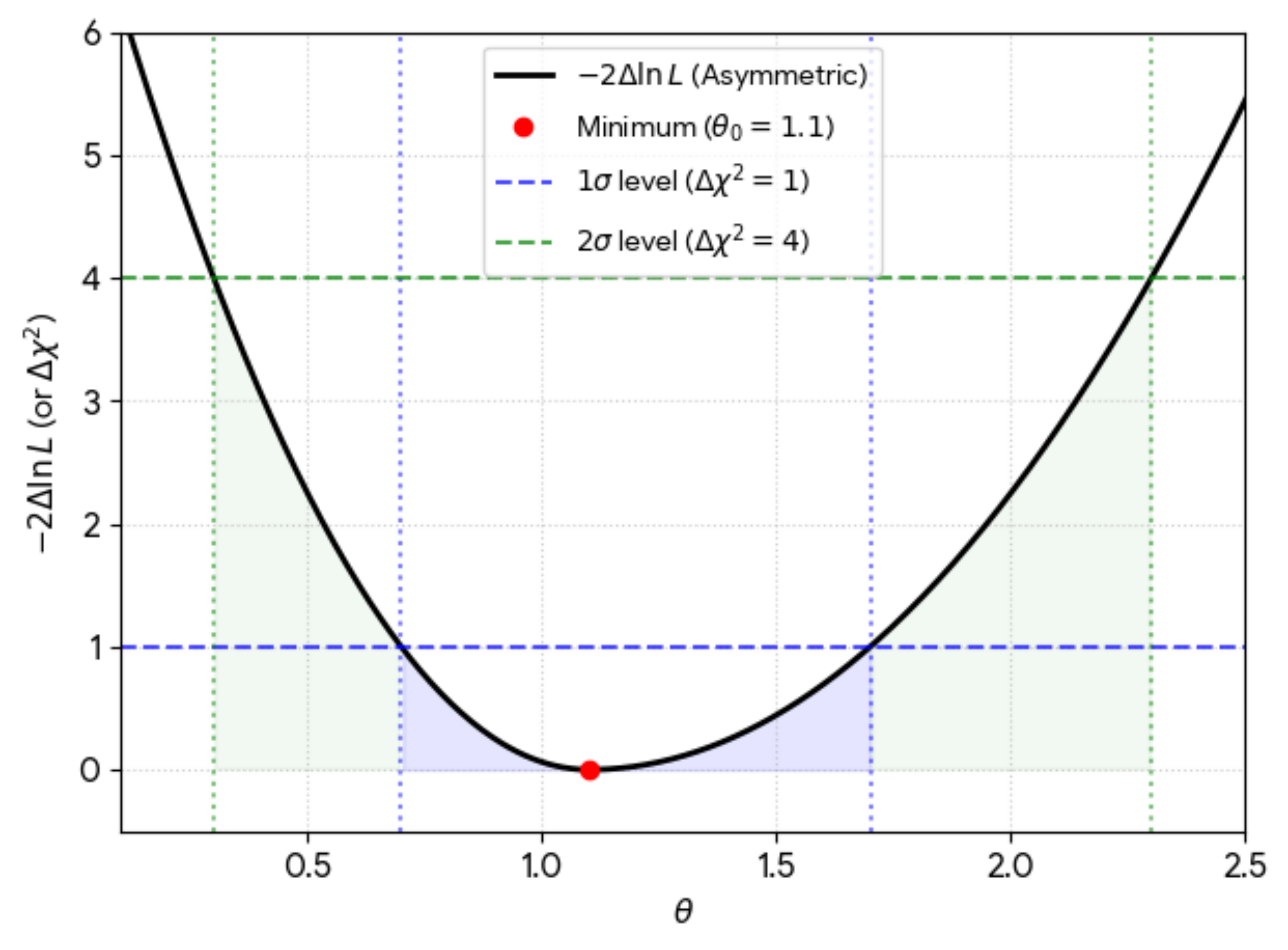
$$\sqrt{\lambda/N} = 0.033$$

From the covariance matrix

- This is the error that you will get, eg, from minuit
- $I(\theta) = -E[\frac{\partial^2}{\partial\theta^2} lnL(\theta)]$ => Hessian matrix
- Cov matrix: Inverse of the Hessian
- Square root of diagonal terms -> errors
- Off diagonal terms - > correlations between parameters
- $SE(\hat{\theta}) = \sqrt{I(\hat{\theta})^{-}1}$

From the likelihood profiles

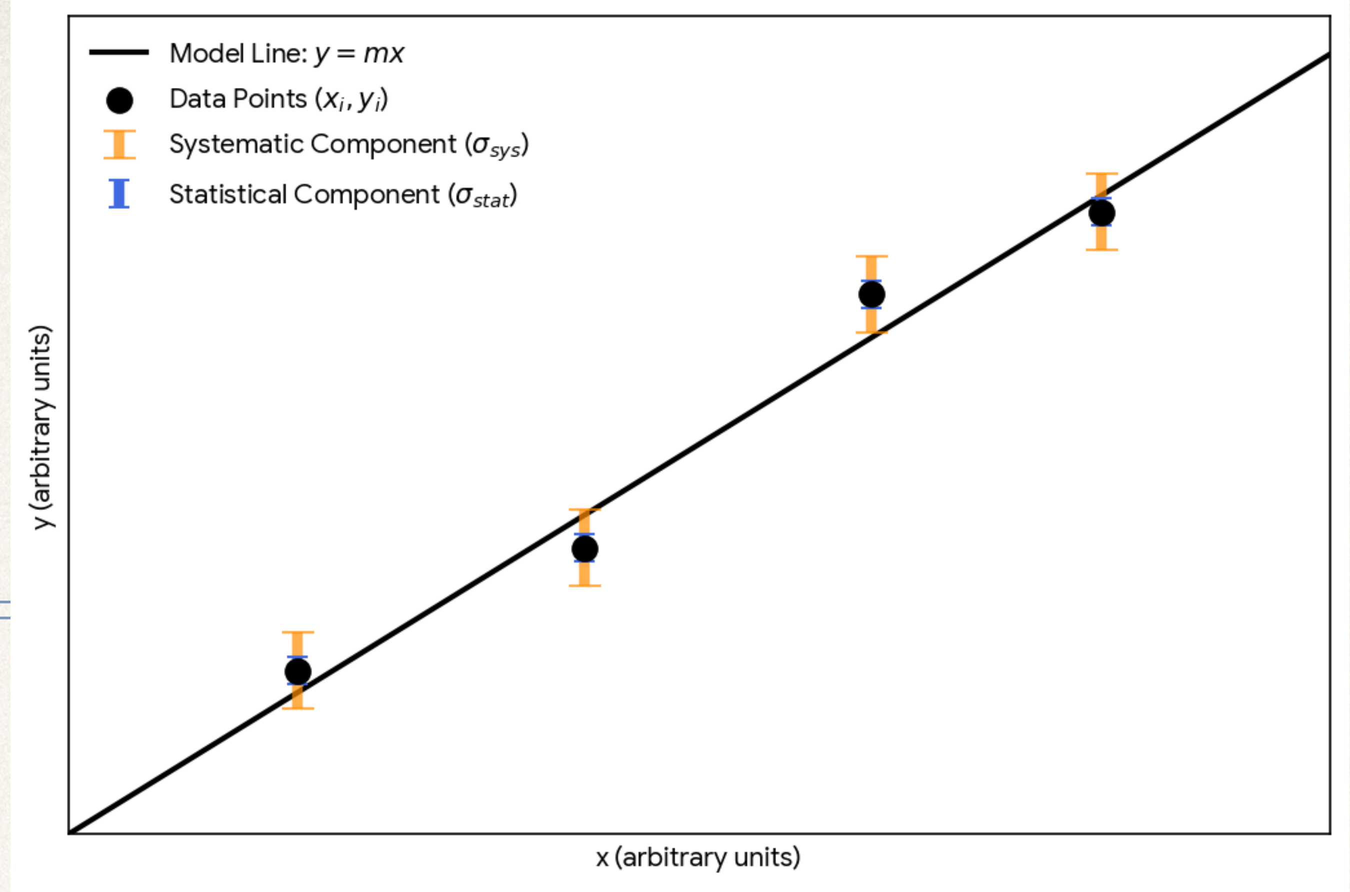
- Robust method
- Plot the likelihood profile along the parameter of interest.
- Compute intervals from this



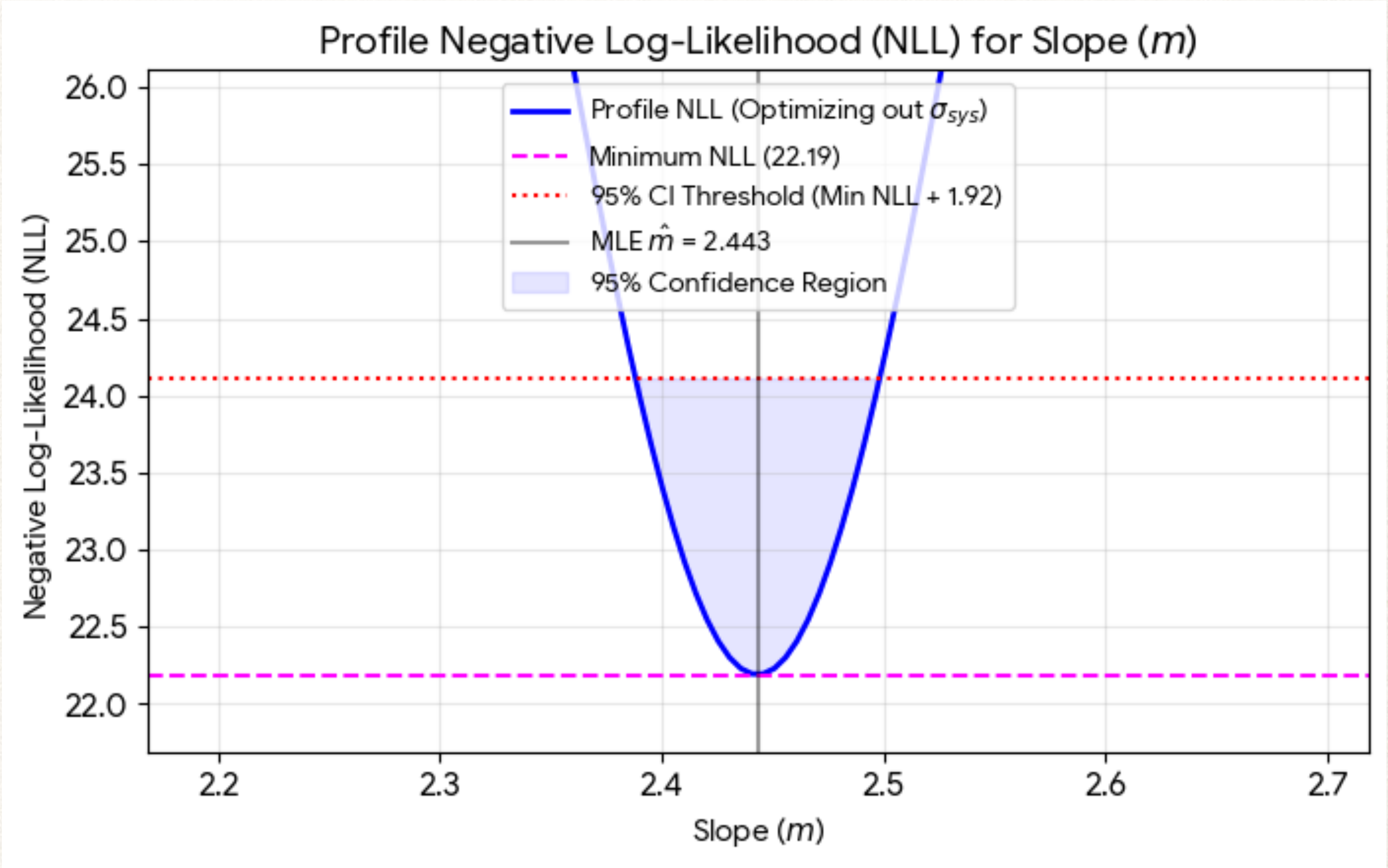
Hypothesis testing

- MLE does not give a direct way to characterise the goodness of fit
- We rather check if an alternate hypothesis H_1 is preferred over the null hypothesis H_0
- $\lambda = \dfrac{max\ L(X|H_1)}{max\ L(X|H_0)}$
- From Wilk's theorem: $2log\lambda \sim \chi^2$ with n d.o.f
- Converted into a p-value

Profile likelihood



- Statistical method used to evaluate a specific parameter of interest in a complex model by "maximizing out" all other unnecessary parameters (nuisance parameters)
  - Common nuisance parameters: systematic uncertainties
  - Great if can be done analytically (WSTAT)



- Eg: Let's say I am fitting $y = mx + c$ to some data. The data have an unknown systematic uncertainty. For simplicity, lets say $c = 0$
  - Freeze 'm' at various points in a grid. Profile our the unknown systematic. Compute the likelihood at different values of m

$$\ell(m, \sigma_{\text{sys}}) = -\frac{1}{2}\sum_{i=1}^{n} \ln\left(2\pi\left(\sigma^2_{\text{stat},i} + \sigma^2_{\text{sys}}\right)\right) - \frac{1}{2}\sum_{i=1}^{n} \frac{\left(y_i - mx_i\right)^2}{\sigma^2_{\text{stat},i} + \sigma^2_{\text{sys}}}$$

Error propagation

Beyond quadrature addition

- Simple quadrature formulas assume independent variables.
- Uncertainties are available in the covariance matrix

$$\mathbf{\Sigma}_x = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots \\ \sigma_{21} & \sigma_2^2 & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

$$\boxed{\mathbf{\Sigma}_f = \mathbf{J}\,\mathbf{\Sigma}_x\,\mathbf{J}^T}$$

- For error propagation to $f(x)$ use the Jacobian

$$\mathbf{J} = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \dots \end{pmatrix}$$

Eg: Measurement of g

You have all measure acc due to gravity, and computed the error in the final result. Let's say for some reason (air drag?) L and T have a weak correlation.

$$g(L,T) = \frac{4\pi^2 L}{T^2}$$

$$\mathbf{J} = \begin{pmatrix} \frac{\partial g}{\partial L} & \frac{\partial g}{\partial T} \end{pmatrix} = \begin{pmatrix} \frac{4\pi^2}{T^2} & -\frac{8\pi^2 L}{T^3} \end{pmatrix}$$

$$\sigma_g^2 = \begin{pmatrix} \frac{4\pi^2}{T^2} & -\frac{8\pi^2 L}{T^3} \end{pmatrix} \begin{pmatrix} \sigma_L^2 & \sigma_{LT} \\ \sigma_{LT} & \sigma_T^2 \end{pmatrix} \begin{pmatrix} \frac{4\pi^2}{T^2} \\ -\frac{8\pi^2 L}{T^3} \end{pmatrix}$$

$$\mathbf{\Sigma}_x = \begin{pmatrix} \sigma_L^2 & \sigma_{LT} \\ \sigma_{LT} & \sigma_T^2 \end{pmatrix}$$

$$\sigma_g^2 = \left(\frac{4\pi^2}{T^2} \right)^2 \sigma_L^2 + \left(-\frac{8\pi^2 L}{T^3} \right)^2 \sigma_T^2 - 2 \left(\frac{32\pi^4 L}{T^5} \right) \sigma_{LT}$$