CHI-SQUARED AND MAXIMUM LIKELIHOOD BASICS

reference slides for CAP Boot Camp tutorials on Interpreting χ^2 and Parameter Estimation by Maximum Likelihood Model Fitting

CHI-SQUARED VALUE VS. DISTRIBUTION

One value of “Chi squared” compares observed minus expected values (residuals from a model) for one set of N data points with uncertainties σ (expected residuals)

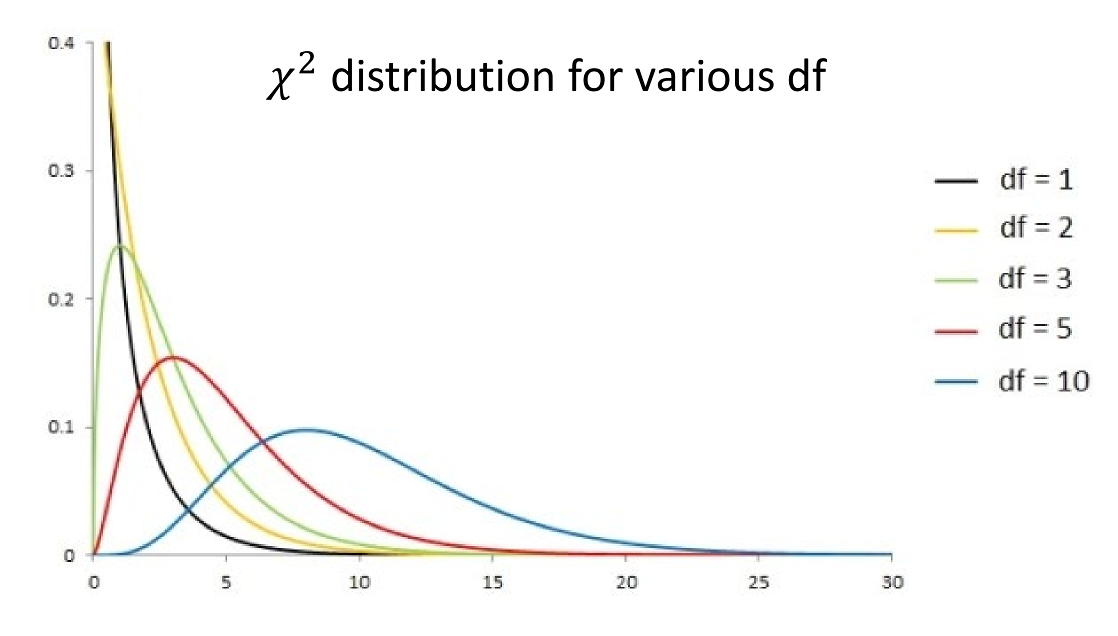
$$\chi^2 = \sum_{i=0,N} \frac{(O_i - E_i)^2}{\sigma_i^2}$$

Different values of χ^2 for different sets of N data points $\{O_i\}$ are drawn from the Chi squared probability distribution, assuming that for a single data point i, the probability of measuring a value O_i is drawn from a normal distribution $\frac{1}{\sqrt{2\pi}\sigma_i} e^{-(O_i - Ei)^2/2\sigma_i^2}$.

CHI-SQUARED AND REDUCED CHI-SQUARED

The equation for the χ^2 distribution is ugly and depends on the # of degrees of freedom:

$df = N - k$ for N data points and k model parameters.



For random χ^2, $\langle\chi^2\rangle \sim df$, so people often define a "reduced" $\widehat{\chi^2} = \chi^2/df$ and expect it to be ~1 for a "good model" (but this is a somewhat risky oversimplification, as you will see in the Interpreting χ^2 tutorial). Notice that for large df, the χ^2 distribution starts to look Gaussian (the central limit theorem!).

CHI-SQUARED AND LIKELIHOOD

Assume that for a given data point i, the probability of measuring a value O_i is $\frac{1}{\sqrt{2\pi}\sigma_i} e^{-(O_i - Ei)^2/2\sigma_i^2}$, and verify for yourself that multiplying the individual data point probability distributions gives an overall probability distribution for the data set $\{O_i\}$ that is proportional to $e^{-\chi^2/2}$ where $\chi^2 = \sum_i \frac{(O_i - E_i)^2}{\sigma_i^2}$.

Based on this, we say that the "likelihood" of a model being correct for a given data set is *proportional to* $e^{-\chi^2/2}$ if the residuals are normally distributed around the data points.

TRADITIONAL MAXIMUM LIKELIHOOD

- seek "best fit" models/parameters
- typically assume likelihood L proportional to $e^{-\chi^2/2}$
- min χ^2 → max likelihood
- "maximum likelihood estimators" (MLEs) of parameters α_i of a model are usually found by $\frac{\partial L}{\partial \alpha_i} = 0$ or equivalently $\frac{\partial \ln(L)}{\partial \alpha_i} = 0$
- assuming all residuals follow same normal distribution (i.e. have same σ), called "ordinary least-squares" (OLS) fitting or "minimizing the rms" (root mean square deviations)

TRADITIONAL MAXIMUM LIKELIHOOD

- Example: model $y = \alpha X + \beta$ with equal Gaussian errors σ
- $\chi^2 = \sum_i \frac{(Y_i - (\alpha X_i + \beta))^2}{\sigma^2}$ ➔ max likelihood
- $\frac{\partial \ln(L)}{\partial \alpha} = 0$ ➔ $\frac{\partial \ln\left(e^{-\frac{\chi^2}{2}}\right)}{\partial \alpha} = 0$ ➔ $\sum_i \frac{\left(Y_i - (\alpha X_i + \beta)\right) X_i}{\sigma^2} = 0$
- $\frac{\partial \ln(L)}{\partial \beta} = 0$ ➔ $\frac{\partial \ln\left(e^{-\frac{\chi^2}{2}}\right)}{\partial \beta} = 0$ ➔ $\sum_i \frac{\left(Y_i - (\alpha X_i + \beta)\right)}{\sigma^2} = 0$
- two eqns, two unknowns – solve to get result in tutorial:

  $\alpha = \frac{\bar{X}\bar{Y} - \overline{XY}}{(\bar{X})^2 - \overline{X^2}}$ and $\beta = \bar{Y} - \bar{X}\alpha$

  (so for this simple case, no numerical χ^2 minimization is needed; but harder for more parameters or different σ_i)

TRADITIONAL MAXIMUM LIKELIHOOD

- uncertainties on MLE params estimated by 1/E(-H) = inverse of expectation of negative "Hessian matrix"
- Hessian matrix example: $y = \alpha X + \beta$

$$\text{Hessian}(\alpha, \beta) = \begin{bmatrix} \frac{\partial^2}{\partial \alpha^2} \log L(\alpha, \beta) & \frac{\partial^2}{\partial \alpha \partial \beta} \log L(\alpha, \beta) \\ \frac{\partial^2}{\partial \alpha \partial \beta} \log L(\alpha, \beta) & \frac{\partial^2}{\partial \beta^2} \log L(\alpha, \beta) \end{bmatrix}$$

note covariance terms!

- complicated to compute Hessians, often done numerically
- fully worked Hessian for least squares case at http://mathworld.wolfram.com/LeastSquaresFitting.html - note errors on parameters generally decrease as $\frac{1}{\sqrt{N}}$