These notes complement the cc-analysis.ipynb file that is available in the course repository.

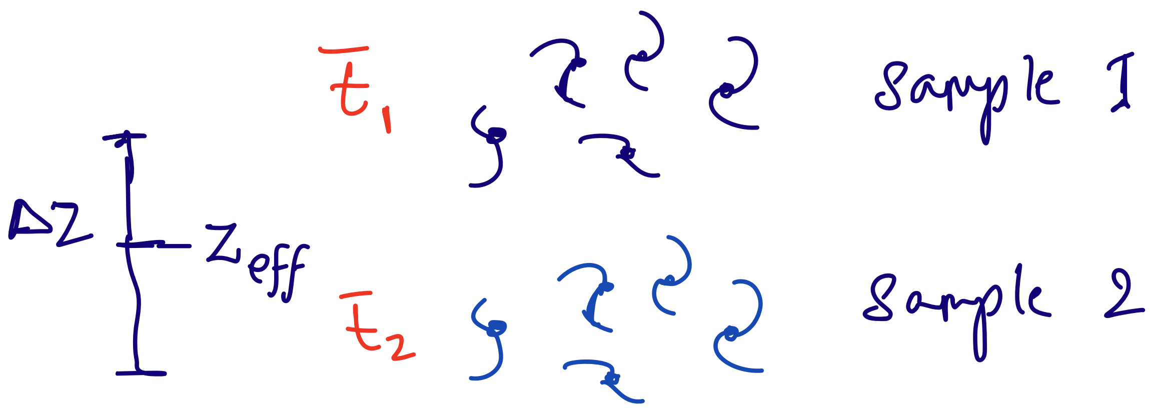
☆ Cosmic chronometer (CC) data:

The Hubble parameter, $H = \frac{1}{a}\frac{da}{dt}$, can be re-written as:

$$H = \frac{-1}{1+z}\frac{dz}{dt} \qquad (z\text{: redshift})$$

Cosmic chronometers (CC) are used to estimate dz and dt, at a given redshift z, to measure H(z) without relying on a specific cosmological model.

Massive, quiescent (no star formation) galaxies are among suitable CC candidates. If two populations of such galaxies are found at a tiny redshift separation (Δz), Spectroscopic analyses of such galaxies can infer their "ages", which is in fact the ages of their dominant stellar populations. These provide estimations of Δt.



$$\Delta t = \bar{t}_2 - \bar{t}_1 \approx dt$$
$$\Delta z \approx dz \longrightarrow H(z_{eff}) = \frac{-1}{1+z_{eff}} \frac{\Delta z}{\Delta t}$$

The galaxies should be as similar as possible (e.g. similar in metalicity, stellar mass, etc.)

Indeed, the inferred values of $H(z)$ are subject to both statistical (e.g. from noises, finite sample sizes, etc.) and systematic (e.g. from poor modeling, incorrect assumptions, instrument errors, etc.) which have been estimated.

In this analysis we use 32 CC data points. Please see the CC-analysis.ipynb file in the repository.

Part 1. chi-squre analysis.

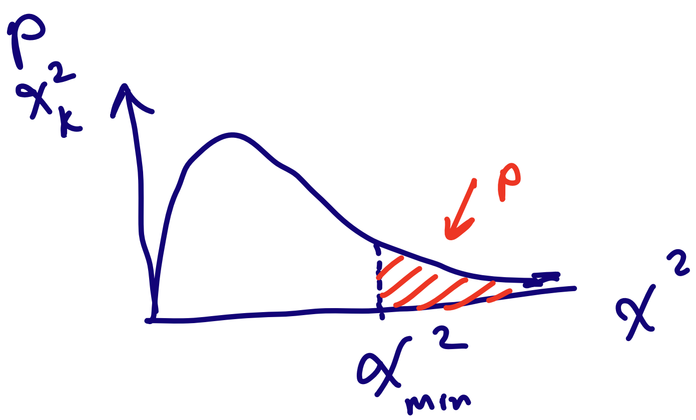
Given the data points, (x_i, D_i), which in our case is $(z_i, H(z_i))$, and a model $y(x, \theta)$ with θ denoting the model parameters, the quantity

$$\chi^2 = \sum_{i,j} \left[D_i - y(x_i|\theta)\right] C^{-1}_{ij} \left[D_j - y(x_j|\theta)\right]$$

with C_{ij} being the covariance matrix, is well described by a chi-squre distribution (in many cases) with $k = n - m$ degrees of freedom (dof). n is the number of data points and m is the number of fitted parameters of the model.

For a given model, one can find model parameters that minimize the χ^2 (best-fit parameters). Then, one can perform "goodness-of-fit" testing, to assess the performance of the model. We can define the "p-value":

$$p = 1 - F_{\chi^2_k}(\chi^2_{min})$$



$F_{\chi^2_k} \rightarrow$ CDF of the chi-squre distribution

If p is too small $\rightarrow$ poor fit

If $p \sim 0.1 - 0.9$ $\rightarrow$ acceptable fit (this may vary in different fields)

If $p \sim 1$ $\rightarrow$ too good to be true! (possible overfit) chi-by-eye: $\chi^2_\nu = \dfrac{\chi^2_{min}}{\nu}$ $\quad \nu \equiv n - m$ (dof)

reduced chi-square $\longrightarrow$

$\chi^2_\nu \simeq 1$ good fit

$\chi^2_\nu \gg 1$ poor fit

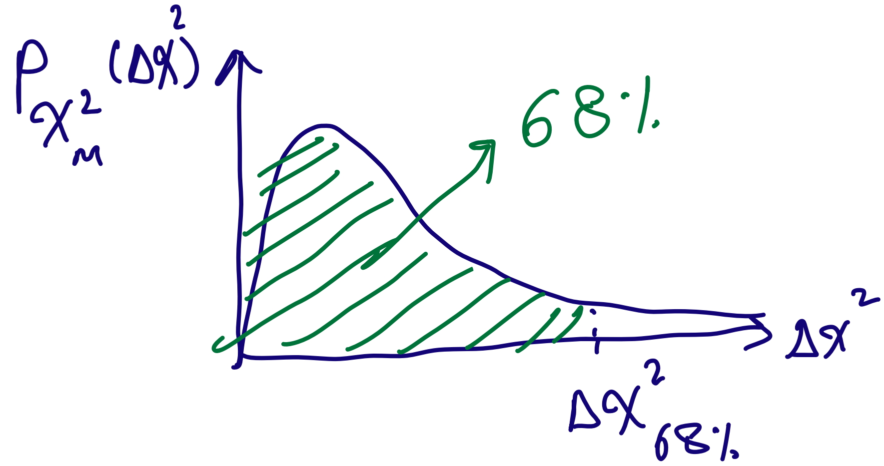
$\chi^2_\nu \ll 1$ too good to be true (overfit)

(over-estimated errors, correlated data, ...)

Confidence regions:

Once χ^2_{min} is found, one can find confidence regions for a give confidence level, using the fact that $\Delta\chi^2 = \chi^2 - \chi^2_{min}$ is shown to follow a χ^2 distribution with m degrees of freedom (m is the number of fitted parameters). This means that we have the probability distribution of $\Delta\chi^2$ and, therefore, we can find the confidence regions from that.

For example, if we are interested in 68% confidence level and our model has m fitted parameters, we have $\rightarrow$



once we find $\Delta\chi^2_{68\%}$,

for $\Delta\chi^2 \leq \Delta\chi^2_{68\%}$

we are in 68% confidence region. In other words,

$$\boxed{\chi^2 - \chi^2_{min} \leq \Delta\chi^2_{68\%}} \rightarrow \text{68\% Confidence region}$$

The values of $\Delta\chi^2_{68\%}$ depend on m, and we can read them from χ^2 tables. Here are some examples for 1σ, 2σ and 3σ confidence levels:

For 1 parameter $(m=1)$:

$1\sigma \rightarrow 68.3\%$ CL, $\Delta\chi^2_{68.3\%} = 1$

$2\sigma \rightarrow 95.4\%$ CL, $\Delta\chi^2_{95.4\%} = 4$

$3\sigma \rightarrow 99.73\%$ CL, $\Delta\chi^2_{99.73\%} = 9$

For 2 parameters $(m=2)$:

$1\sigma \rightarrow 68.3\%$ CL, $\Delta\chi^2_{68.3\%} = 2.3$

$2\sigma \rightarrow 95.4\%$ CL, $\Delta\chi^2_{95.4\%} = 6.18$

$3\sigma \rightarrow 99.73\%$ CL, $\Delta\chi^2_{99.73\%} = 11.8$

Similarly, for any other confidence levels, or number of model parameters the corresponding $\Delta\chi^2_{critical}$ values can be found from χ^2 tables.

Bayesian Inference

Bayes theorem:

For events:

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

For Continuous variables:

$$P(x_1, x_2) = P(x_1|x_2)\,P(x_2)$$
$$P(x_2, x_1) = P(x_2|x_1)\,P(x_1) \Rightarrow P(x_1|x_2) = \frac{P(x_2|x_1)\,P(x_1)}{P(x_2)}$$

Similarly for data, D, and hypothesis, H, we can write:

$$\underbrace{P(H|D)}_{\text{posterior}} = \frac{\overbrace{P(D|H)}^{\text{Likelihood}}\ \overbrace{P(H)}^{\text{prior}}}{P(D)}$$

$$P(D|H) = \mathcal{L} \propto e^{-\chi^2/2}$$

$p(H)$ (prior) can be uniform, Gaussian, etc.

☆ [If a parameter is always positive (like neutrino masses) or always negative, this can be enforced through the prior] ☆

$P(D) \rightarrow$ Normalization (can absorb other constants or unimportant factors)

Having the posterior, one can find credibility regions/intervals for a given credibility level.

people often use "credibility" instead of "confidence" in Bayesian applications.

Having a joint posterior, $P(H|D)$, which is a function of multiple parameters, $P(\theta_1, \theta_2, \cdots, \theta_n)$, we can also find the marginalized posteriors for every parameter, $P(\theta_j)$, which is equivalent to integrating over other parameters, $\theta_{i \neq j}$.

Side note: Error propagation formula:

If $y = f(x)$ and C_x is the covariance matrix of x, then: $C_y \simeq J C_x J^T$

↳ Covariance matrix of y

$J \rightarrow$ Jacobian matrix

$$\text{posterior} = \frac{e^{-\chi^2/2}}{Z} \times \text{prior}$$

$$\log(\text{posterior}) \propto -0.5\,\chi^2 + \log(\text{prior})$$

Model Comparisons

1) Bayes factor: $B_{ij} = \dfrac{P(D|M_i)}{P(D|M_j)}$

fitted parameters of model M_i

$$P(D|M_i) = \int d\theta_i \underbrace{P(D|\theta_i, M_i)}_{\text{likelihood}} \underbrace{P(\theta_i|M_i)}_{\text{prior}}$$

$$\Rightarrow P(D|M_i) = \int d\theta_i \, \mathcal{L}_{M_i}(D, \theta_i) \, \mathrm{Prior}_{M_i}(\theta_i)$$

$$\Rightarrow B_{ij} = \frac{\int d\theta_i \, \mathcal{L}_{M_i}(D, \theta_i) \, \mathrm{Prior}_{M_i}(\theta_i)}{\int d\theta_j \, \mathcal{L}_{M_j}(D, \theta_j) \, \mathrm{Prior}_{M_j}(\theta_j)}$$

If $B_{ij} > 1 \rightarrow i$ is favored over j.

(More specifically, people define intervals to quantify the strength of the preference: $B_{ij} \gg 1 \rightarrow$ strong preference for model i)

2) Akaike Information Criterion (AIC):

$$AIC = \underbrace{-2 \ln \mathcal{L}_{max}}_{\text{Can be replaced with } \chi^2_{min} \text{ for model comparison}} + \overset{\text{number of fitted model parameters}}{2M} \rightarrow \chi^2_{min} + 2M$$

Best model minimizes AIC

3) Bayesian Information Criterion (BIC)

$$BIC = -2\ln\mathcal{L}_{max} + m\ln N \rightarrow \chi^2_{min} + m\ln N$$

number of data points.

Best model minimizes BIC

4) In many cosmology papers, people consider the difference between χ^2_{min} for two models (related to the likelihood ratio test) and analyze the significance of the difference (usually expressed in terms of σ multiples like 2.5σ, 1.1σ, 3σ, ... This helps people find which model is preferred by the data and by what significance.

joint analysis of different data sets

we can combine different data sets and perform statistical analyses. If data sets are not correlated (errors are not correlated), then on can multiply the likelihoods or, equivalently, add up the chi-squares: $\chi^2_{joint} = \chi_1^2 + \chi_2^2 + \dots$. If data sets are correlated, one needs to calculate the corresponding covariance matrix and form the joint χ^2 accordingly.

Tension metrics

when people talk about about the fact that the measured value of parameter X from data A is, for example, 4.2σ different from its measured value from data B, this can be understood from the posteriors (or likelihoods if posteriors are not available).

Such differences can be referred to as "tensions" and there are several ways that a "tension metric" can be defined. One of the simplest ways is to first find the 1σ errors for each measurement ($\sigma_{X,A}$, $\sigma_{X,B}$) and then calculate:

$$\text{Tension} = \frac{|X_A - X_B|}{\sqrt{\sigma_{X,A}^2 + \sigma_{X,B}^2}}$$

1σ error

Measurement A: $X_A \pm \sigma_{X,A}$

Measurement B: $X_B \pm \sigma_{X,B}$

(For example, a 3σ tension means tension = 3)

$\sigma_{X,A}$, $\sigma_{X,B}$ are symmetrical 1σ errors. Most posteriors are very close to Gaussian. Therefore, the errors are almost symmetrical.

* There are other ways to define a tension metric which can be more accurate, especially when

posteriors are not symmetrical, which can be found from online resources.

Also sometimes, the tension is with respect to constant value(s) that are predicted by a model. And sometimes, people use Hypothesis testing to quantify a tension of this kind. So we should see the details of the statistical analyses to fully understand the meaning of such reports.