

Abilita` Informatiche e Telematiche 2022/2023 - P. Monaco

Per confrontare delle misure con un modello bisogna conoscere l'errore sui dati; se i dati sono correlati bisogna conoscerne la covarianza. In un approccio di maximum likelihood, una funzione di likelihood Gaussiana sara`:

$$\mathcal{L} \propto \exp \left(-\frac{1}{2} (D - M)^T \mathbf{C} (D - M) \right)$$

Qui D e` un vettore di dati, M il modello che li descrive e \mathbf{C} la matrice di covarianza, la cui diagonale da` l'errore quadratico sulla misura, mentre gli elementi fuori diagonale danno la correlazione dei dati.

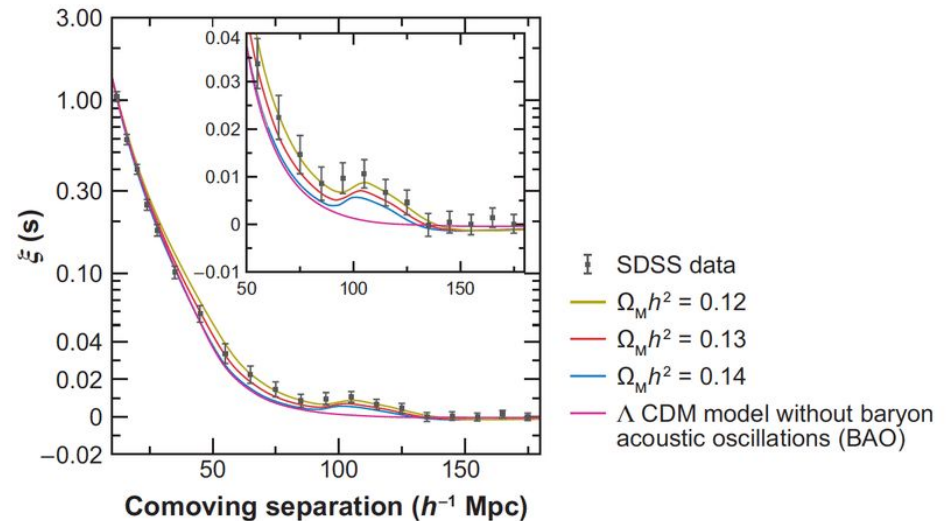
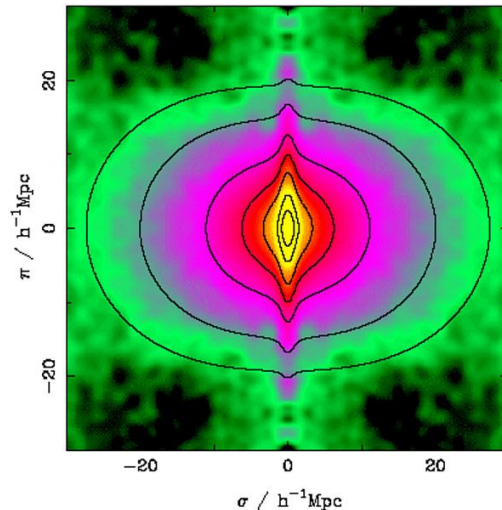
In diversi casi (come in cosmologia) la covarianza di un set di misure sperimentali va calcolata utilizzando un grande numero di simulazioni, misurandone una “summary statistics” (che costituisce D) e calcolandone la covarianza con metodi standard:

$$C_{ij} = \frac{1}{N_{\text{sim}} - 1} \sum_{n=1}^{N_{\text{sim}}} (D_i^{(n)} - \bar{D}_i)(D_j^{(n)} - \bar{D}_j)$$

Abilita` Informatiche e Telematiche 2022/2023 - P. Monaco

Abbiamo generato tre insiemi di 10,000 **misure** che rappresentano la *funzione di correlazione a due punti* (2PCF) di un campione di galassie.

Ogni misura consiste in 200 valori per 200 bin di distanza. La 2PCF e` misurata in due dimensioni (lungo la linea di vista e trasversale alla stessa), la misura 2D e` compressa in cinque *multipoli* della 2PCF. I multipoli pari sono nulli per simmetria.



Abilita` Informatiche e Telematiche 2022/2023 - P. Monaco

Per validare un codice che calcola la covarianza numerica, bisogna generare un set di misure che hanno covarianza nota. Questo e` possibile utilizzando una tecnica nota come decomposizione di Cholesky, che permette di creare set di dati con covarianza (Gaussiana) nota. L'autocorrelazione per un singolo multipolo sara`:

$$C^{\text{th}}(D_{\ell}(r_i), D_{\ell}(r_j)) = \sigma_{\ell}^2 \exp\left(\frac{-(r_i - r_j)^2}{2h_{\ell}^2}\right)$$

I multipoli sono correlati, la loro correlazione mista sara`:

$$C^{\text{th}}(D_{\ell}(r_i), D_{\ell'}(r_j)) = \sigma_{\ell}\sigma_{\ell'}\sqrt{\frac{2h_{\ell}h_{\ell'}}{h_{\ell}^2 + h_{\ell'}^2}} \exp\left(\frac{-(r_i - r_j)^2}{h_{\ell}^2 + h_{\ell'}^2}\right)$$

Abilita` Informatiche e Telematiche 2022/2023 - P. Monaco

Risulta utile definire la matrice di correlazione, che e` la covarianza normalizzata ad uno sulla diagonale:

$$R_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}$$

Date N_{sim} misure simulate, ci aspettiamo che i residui tra covarianza teorica e misura, normalizzati alla loro varianza, siano distribuiti come una Gaussiana a deviazione standard unitaria:

$$\text{Res}_{ij} = (C_{ij}^{\text{th}} - C_{ij}^{\text{meas}}) \sqrt{\frac{N_{\text{sim}} - 1}{(1 + R_{ij})C_{ii}^{\text{th}}C_{jj}^{\text{th}}}}$$

Abilita` Informatiche e Telematiche 2022/2023 - P. Monaco

L'esercizio consiste in:

- scaricare i tre set di misure simulate da <https://adlibitum.oats.inaf.it/monaco/etc/perAbInf.tgz>

- scrivere uno script bash per copiare i file in una directory nota, con un nome noto

- leggere un singolo multipolo da un set limitato di dati

- calcolarne la covarianza numerica

$$(\sigma_0=\sigma_2=\sigma_4=0.02, h_0=25, h_2=50, h_4=75, \rho_0=\rho_2=\rho_4=1.0)$$

- calcolarne la covarianza teorica

- confrontarle e calcolare la differenza quadratica media

$$(\sigma_0=0.02, \sigma_2=0.01, \sigma_4=0.005, h_0=h_2=h_4=50, \rho_0=\rho_2=\rho_4=1.0)$$

dei residui normalizzati, verificando che sia ~ 1 (validazione)

$$(\sigma_0=0.02, \sigma_2=0.01, \sigma_4=0.005, h_0=h_2=h_4=5, \rho_0=\rho_2=\rho_4=1.0)$$

- fare grafici di queste matrici di covarianza

- estendere la procedura a tre multipoli pari (0, 2 e 4) includendo

le cross-correlazioni

- validare il calcolo della covarianza numerica sui tre set di misure

- mettere tutto su github e consegnare