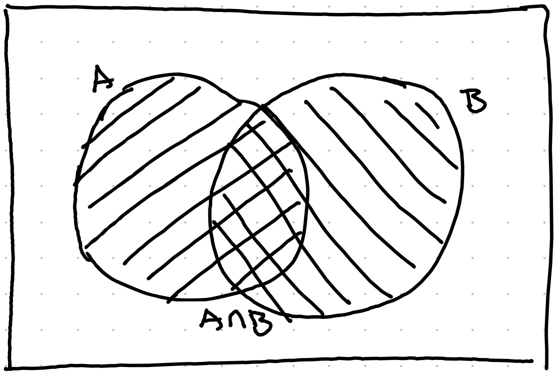
Lecture 5: An introduction to probability & time-series analysis

topics: probability densities
- conditional, joint, & marginal distributions
- Bayes Theorem: prior vs. likelihood vs. posterior vs. evidence
- (hierarchical) graphical models

Bayes Factors, Odds ratios, Prior Odds

time-series as an example of stochastic processes
- Gaussian Processes → Power Spectral Density
- Whittle Likelihood as an approximation
- PSD estimation, DFTs & window functions

Bayesian Probability



$$P(A,B) = P(A|B)P(B)$$
$$= P(B|A)P(A)$$

$$\therefore \; P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes theorem

nomenclature

joint distribution: $p(A,B)$

marginal distribution: $p(A) = \int dB \, p(A,B)$

conditional distribution: $p(A|B) = \dfrac{p(A,B)}{p(B)}$

common terms w/in data analysis

likelihood   prior

$$p(\text{params} \mid \text{data, hypothesis}) = \frac{p(\text{data} \mid \text{params, hypothesis})\; p(\text{params} \mid \text{hypothesis})}{p(\text{data} \mid \text{hypothesis})}$$

posterior

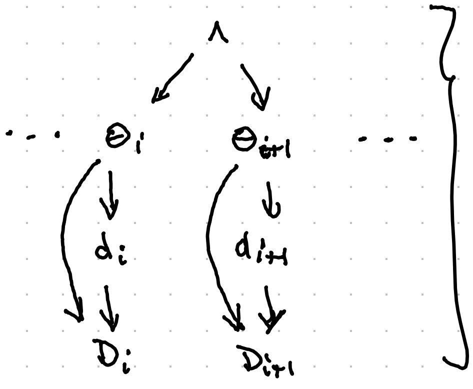
evidence
(marginal likelihood)

Bayes factors: $B^A_B = \dfrac{p(\text{data} \mid A)}{p(\text{data} \mid B)}$

Odds factor: $O^A_B = \dfrac{p(A \mid \text{data})}{p(B \mid \text{data})} = B^A_B \dfrac{p(A)}{p(B)}$

prior odds

Graphical models: ways to express conditional dependencies



$$p(\Lambda, \{\theta_i, d_i, D_i\}) =$$
$$p(\Lambda) \prod_i^N p(\theta_i|\Lambda)\, p(d_i|\theta_i)\, p(D_i|d_i, \theta_i)$$

Prob. mass func: defined over descrete sets

Prob. density: defined over (finite-dimensional) continuous spaces

Prob. process: defined over infinite-dimensional continuous spaces

↑ prob. measure over functions

for many (all?) practical purposes, you can think of a process as a density over a very large-dim. vector space

Gaussian Noise Processes

model the time-series data produced by a detector in the absence of a signal as a stochastic process.

$$n(t) \sim P(n)$$

now, assume this process is Gaussian so that it can be described completely by it's 1st two moments

(often $0 =$) $\langle n(t) \rangle_P$ and $\langle n(t)\, n(t') \rangle_P$

where $\langle x \rangle_P \equiv \int \mathcal{D}n \, P(n)\, x$

If we further specialize to the case of stationary noise, this means

$$\langle n(t)\, n(t+\tau) \rangle_P = f(\tau)$$

autocorrelation depends only on the separation in time (time-translation invariant)

Great, but if I want to evaluate $P(n)$ then I still need to invert a covariance matrix

$$\ln P(n) \sim -\frac{1}{2} n_i \, C^{-1}_{ij} \, n_j$$

where $C_{ij} = \langle n_i n_j \rangle$

this is expensive...

Instead, consider freq. domain

$$\langle \tilde{n}(f)\, \tilde{n}^+(f') \rangle = \frac{1}{2} S(f)\, \delta(f-f')$$

where $\tilde{n}(f) = \int dt\, e^{-2\pi i f t}\, n(t)$ and $+ \Rightarrow$ complex conjugation

and $S(f) = 2 \int dt\, e^{-2\pi i f t}\, f(t)$

is the one-sided Power Spectral Density (PSD)

in the freq domain, then

$$\ln P(\tilde{n}) = -\frac{1}{2} \tilde{n}^{+}_i\, \tilde{C}^{-1}_{ij}\, \tilde{n}_j = -\frac{1}{2}\left[4 \int_0^\infty df \, \frac{|\tilde{n}|^2}{S}\right]$$

if S is constant ⟶ white noise

otherwise ⟶ colored noise

We can completely describe the properties of stationary, Gaussian (zero-mean) noise w/ just the PSD.

Note that we often approximate this with the Whittle Likelihood

take a DFT of discretely sampled $n(t_i)$
approx Gauss. likelihood as

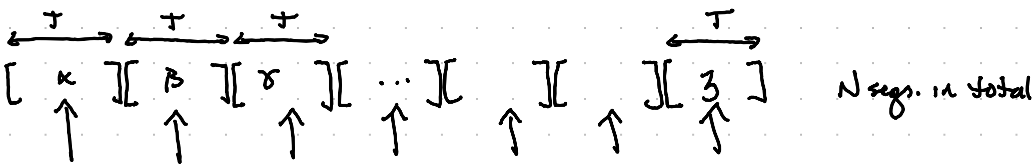
$$\ln P \sim -\frac{1}{2}\left(4 \sum \left[\delta f \, \frac{|DFT(n)|^2}{S}\right]\right)$$

OK, cool. But how do we estimate the PSD?

2 approaches w/in GW literature:

1) parametric model for PSD & fit

2) Welch's Method

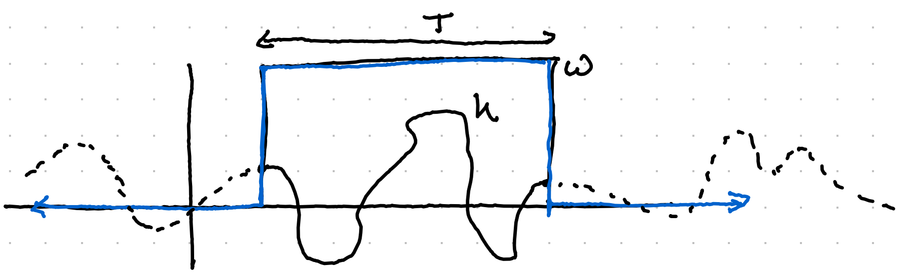


DFT each segment separately
assume each segment has length T

We then estimate the PSD by averaging the power @ each freq.

$$PSD(f_i) \approx \frac{1}{2TN} \sum_{k}^{N} \left[|DFT(h_k)|^2 (f_i) \right]$$

A note about DFTs: windows matter!



a signal that was observed for a finite time window looks like a signal defined over a much wider (infinite) range multiplied by a <u>window</u>

$$\therefore \; DFT \sim \int_{t_0}^{t_0+T} dt\, e^{-2\pi i f t}\, h \sim \int_{-\infty}^{\infty} dt\, e^{-2\pi i f t}\, w(t)\, h(t)$$

$$\sim (\tilde{w} \circ \tilde{h})(f)$$

multiplication in time domain is convolution in the freq. domain.

→ sharp corners have broad side-bands
⇒ smear out the signal

∃ many windowing functions, but they all "roll-off" the signal so that it slowly goes to zero @ the edges of the observation ← <u>avoid sharp corners</u>!