Lecture 6: Gravitational-wave search techniques

topics: Decision theory
- type 1 vs. type 2 errors
- detection confidence
  - background estimation (via timeslides)
  - prob. of astrophysical origin

Neyman-Pearson Lemma

maximum likelihood estimators
- matched filter SNR as max-like amplitude
- burst MLE searches

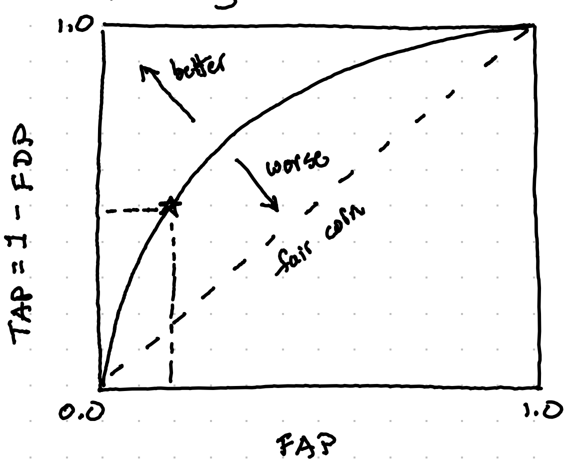
Intro to Decision Theory

Decide what we believe based on observations

Type 1 error: reject null hyp. when it is true (False Alarm)

Type 2 error: fail to reject null hyp. when it is false (False Dismissal)

Receiver Operating Characteristic (ROC) Curves



We compute some statistic of the data (x) and use that to decide whether to reject null hypothesis.

$$TAP = \int_{x > x_{thr}} dx \; p(x \mid \text{signal} + \text{noise})$$

$$FAP = \int_{x > x_{thr}} dx \; p(x \mid \text{noise})$$

additionally, we can compute the prob. of a signal being present

$$p(\text{sig} \mid x) = \frac{p(x \mid \text{sig})\, p(\text{sig})}{p(x \mid \text{sig})\, p(\text{sig}) + p(x \mid \text{noise})\, p(\text{noise})}$$

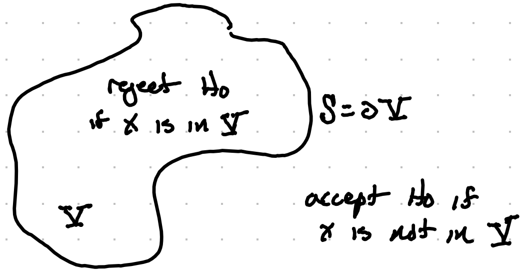
this is often called "p_astro" in the literature

To do this, we need to estimate $p(x \mid \text{sig})$ & $p(x \mid \text{noise})$

Ask me offline how that's done in practice if you're interested

Neymon-Pearson Lemma

"x" space



what decision surface maximizes the TAP at fixed FAP. This turns out to be the likelihood ratio.

$$\max \left\{ \int_V dx\, p(x|sig) + \lambda \left(1 - \int_V dx\, p(x|noise) \right) \right\}$$

w/ respect to the volume V (i.e., choose the best region to max. the ROC curve).

heuristically $\Rightarrow$ $$\underbrace{\lambda = \left. \frac{p(x|sig)}{p(x|noise)} \right|_{S \equiv \partial V}}$$

regions are bounded by surfaces of constant likelihood ratio

$\therefore$ a "likelihood ratio test" (ranks events by $p(x|s)/p(x|n)$) is the optimal detection statistic.

n.b. Bayes factors are likelihood ratios!

but they are often too expensive to be used in low-level processing...

$\Longrightarrow$ we resort to maximization rather than marginalization

Now, let's consider a "realistic scenario" and derive a maximum-likelihood search

Assume stationary, Gaussian <u>additive</u> noise such that

$$x = n + Ah$$

observed data (x), noise (n), signal with known shape h but unknown amplitude A

$$\therefore\ p(x|\text{sig}) = p(x|Ah)$$
$$= P(n = x - Ah) \sim \exp\left(-\frac{1}{2}\left[4\int_0^\infty df\, \frac{|\tilde{x} - A\tilde{h}|^2}{S}\right]\right)$$

$$p(x|\text{noise}) = p(x|A=0)$$
$$= P(n = x) \sim \exp\left(-\frac{1}{2}\left[4\int_0^\infty df\, \frac{|\tilde{x}|^2}{S}\right]\right)$$

construct log-likelihood ratio

$$\ln p(x|s) - \ln p(x|n) \sim -\frac{1}{2}\left[4\int_0^\infty df\, \frac{|\tilde{x} - A\tilde{h}|^2}{S} - \frac{|\tilde{x}|^2}{S}\right]$$

$$\sim -2\int_0^\infty df\, \frac{A^2|\tilde{h}|^2 - 2A\,\mathrm{Re}\{\tilde{x}^*\tilde{h}\}}{S}$$

this would define our decision surface if we knew A. But, since we don't, let's maximize this w/r/t A

i.e., what is the value of A that would make us most likely to reject the null hypothesis?

$$\frac{\partial}{\partial A}(\cdots) = 0 \implies \hat{A} = \left[\frac{4\int_0^\infty df\, \frac{\mathrm{Re}\{\tilde{x}^*\tilde{h}\}}{S}}{4\int_0^\infty df\, \frac{|h|^2}{S}}\right]$$

(it is conventional to define this w/ factors of 4)

We more commonly deal w/ the signal-to-noise ratio

$$\rho = \frac{4\int_0^\infty df\, \mathrm{Re}\{\tilde{x}^*\tilde{h}\}/S}{\sqrt{4\int_0^\infty df\, |h|^2/S}}$$

$\longleftarrow$ "standard deviation" of $\hat{A}$

This is a <u>linear filter</u> and can be computed in either the time or the frequency domain

time: convolution; frequency: multiplication

Note that, if the signal model can additionally be written as " $Ae^{i\phi}\tilde{h}$ "

then it is possible to maximize over both A and ϕ

$$\rho = \sqrt{\rho_R^2 + \rho_I^2}$$

where $$\rho_{(R,I)} = \frac{4\int_0^\infty df\, (R,I)\{\tilde{x}^* h\}/S}{4\int_0^\infty df\, |h|^2/S}$$

Additionally, b/c x is Gaussian, we can analytically compute the behavior of $\rho_R, \rho_I, \,\&\, \rho$

$$\rho_R \sim N(\rho_{opt}, 1) \qquad \text{where } \rho_{opt}^2 = 4\int_0^\infty df\, |h|^2/S$$
$$\rho_I \sim N(0, 1)$$

$\rho \sim$ non-central χ-distributed w/ 2 degrees of freedom & non-centrality parameter ρ_{opt}

Unmodeled Burst Searches

What if we don't have a model for the signal's shape?

maximize likelihood w/ respect to the signal amp & phase in each freq. bin separately!

The mathematics are not exceptionally more complicated, but this is a rather <u>under-determined</u> problem.

$\Rightarrow$ lots of degeneracies and other assumptions are typically needed to make progress