Conditional Probabilities, Bayes Theorem, and Stationary Gaussian Noise

24 Sept. 2021
Reed Essick

Learning Objectives:

- Construct a joint distribution over signal parameters, noise realizations, and observed data
- Marginalize over (unobserved) noise to obtain likelihood
- Use Bayes theorem to construct a posterior
- Define stationary Gaussian noise
  - → define the Power Spectral Density in terms of the noise autocorrelation
- Write the standard GW likelihood for an arbitrary detector network and arbitrary signal model

We model noise as a RANDOM PROCESS. That is, we do not know exactly what noise is present in the detectors at any given time, but we believe certain realizations are more common than others. That is, we assume

$$n \sim p(n) \qquad (1)$$

where $p(n)$ describes the noise characteristics within the detector. Furthermore, we often assume ADDITIVE NOISE so that the observed data is the linear combination of noise and signal

$$h = n + s \qquad (2)$$

h: observed data
n: noise
s: signal

Now, let's construct a probabilistic model for the observed data and the (latent) noise and signal. This is simply

$$p(h,n,s) = \delta(h-(n+s))\,p(n)\,p(s) \qquad (3)$$

where $p(s)$ describes our (prior) beliefs about the relative frequency of different signals.

Let us now convert this expression into other useful probability densities:

$$\text{LIKELIHOOD}: \quad p(h|s) = p(n=h-s) \qquad (4)$$

we can obtain this trivially by noting

$$\begin{aligned} p(h|s) &= \int dn\, p(h,n|s) \\ &= \int dn\, \frac{p(h,n,s)}{p(s)} \\ &= \int dn\, \delta(h-(n+s))\,p(n) \\ &= \int dn\, p(n)\,\delta(n-(h-s)) \\ &= p(n=h-s) \end{aligned}$$

That is, the likelihood of obtaining the observed data given a signal is just the probability of the corresponding noise realization.

POSTERIOR: $$p(s|h) = \frac{p(h|s)\,p(s)}{p(h)} \quad (5)$$

This is obtained from our likelihood and prior beliefs about signals via Bayes Theorem

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

$$\therefore \; P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (6)$$

in (6), the following names are conventional

$P(A|B)$: posterior
$P(B|A)$: likelihood
$P(A)$: prior
$P(B)$: evidence

Many applications only require knowledge of a function proportional to the posterior, and the evidence is often neglected. One will often see only

$$P(A|B) \propto P(B|A)P(A)$$

However, the evidence plays a key role in model selection.

Now, let us specialize to common assumptions for the noise model w/in GW data analysis:

Stationarity: the characteristics of the noise process do not change over time

Gaussianity: the noise distribution can be described by a mean and covariance (autocorrelation) ⇒ is Gaussian.

Specifically, we consider the noise in the detector as a function of time: $n(t)$ and assume

$$\langle n \rangle = \int dn\, p(n)\, n = 0 \quad \forall\, t \tag{7}$$

$$\langle n(t)\, n(t+\tau) \rangle = \int dn(t)\, dn(t+\tau)\; p(n(t), n(t+\tau))\; n(t)\, n(t+\tau)$$

$$= f(\tau) \tag{8}$$

the noise autocorrelation only depends on the separation: τ

With these definitions and the assumption of Gaussianity we obtain

$$p(n) \propto \exp\left(-\frac{1}{2}\int dt_1 \int dt_2\, n(t_1)\, n(t_2)\, K^{-1}(|t_1 - t_2|)\right) \tag{9}$$

where K^{-1} is the "inverse covariance kernel" corresponding to the noise autocorrelation function. In general, K^{-1} can be <u>very</u> "non-diagonal". However, we can simplify these expressions in the frequency domain.

first, define

$$\tilde{n}(f) = \int dt\, e^{-2\pi i f t}\, n(t) \qquad (10)$$

then

$$\langle \tilde{n}(f)\, \tilde{n}^{+}(f') \rangle = \int d\tilde{n}\, d\tilde{n}^{+}\, p(\tilde{n}(f), \tilde{n}^{+}(f'))\, \tilde{n}(f)\, \tilde{n}^{+}(f')$$

$$= \int d\tilde{n}\, d\tilde{n}^{+}\, p(\tilde{n}, \tilde{n}^{+}) \int dt\, e^{-2\pi i f t}\, n(t) \int dt'\, e^{+2\pi i f' t'}\, n(t')$$

$$= \int dt\, e^{-2\pi i f t} \int dt'\, e^{+2\pi i f' t'} \int d\tilde{n}\, d\tilde{n}^{+}\, p(\tilde{n}, \tilde{n}^{+})\, n(t)\, n(t')$$

now,

$$d\tilde{n}(f)\, d\tilde{n}^{+}(f')\, p(\tilde{n}(f), \tilde{n}^{+}(f')) = dn(t)\, dn(t')\, p(n(t), n(t'))$$

so that

$$\int d\tilde{n}\, d\tilde{n}^{+}\, p(\tilde{n}, \tilde{n}^{+})\, n(t)\, n(t') = \int dn(t)\, dn(t')\, p(n(t), n(t'))\, n(t)\, n(t')$$

$$= \langle n(t)\, n(t') \rangle = f(t'-t)$$

with a further change of variables, we obtain

$$\langle \tilde{n}(f)\, \tilde{n}^{+}(f') \rangle = \int dt\, e^{-2\pi i (f-f')t} \int d\tau\, e^{2\pi i f' \tau}\, f(\tau)$$

$$= \delta(f-f')\, S_n(f)$$

where $S_n(f) \equiv \int d\tau\, e^{2\pi i f \tau}\, f(\tau)$ is the POWER SPECTRAL DENSITY, defined as the Fourier Transform of the noise autocorrelation function.

We also note that the covariance in the frequency domain is diagonal, so the noise distribution simplifies to

$$p(\tilde{x}) \propto \exp\left(-\frac{1}{2}\int df \, \frac{\tilde{x}(f)\,\tilde{x}^*(f)}{S_n(f)}\right) \qquad (11)$$

This is the standard noise distrib assumed w/in GW data analysis. While it is convenient to work in the frequency domain because the integrals simplify, we note that one can at times construct a more natural inference in the time domain (see further reading).

=

The last complication is the fact that the detectors respond differently to astrophysical signals that arrive from different directions &/or with different polarization content. The ANTENNA PATTERNS or "detector response" is the transfer function from astrophysical strain in polarization "p" (h_p) to the signal observed by a detector. We often write this in the freq. domain

$$\tilde{s} = \sum_p F_p \tilde{h}_p \qquad (12)$$

where F_p are the Antenna functions. While this expression always holds in the freq. domain (linearity), it may not be as simple in the time domain: F_p can depend on frequency (see further reading).

Exercise:

Derive the joint likelihood for a network of N detectors, each with independent stationary Gaussian noise and separate antenna patterns, that observe the same astrophysical signal.

Advanced Exercise:

Derive the time-domain likelihood for discretely sampled data under the assumption of stationary Gaussian noise. Include the proper normalization and an explicit expression for the covariance matrix between discrete data points.

Advanced Exercise:

Justify why we can exchange

$$d\tilde{n}(f)\, d\tilde{n}^*(f')\; p(\tilde{n}(f), \tilde{n}^*(f'))$$

for

$$dn(t)\, dn(t')\; p(n(t), n(t'))$$

when computing $\langle \tilde{n}(f)\, \tilde{n}^*(f') \rangle$.