Gravitational Waves Notes 3 Maggiore Vol. 1

Krish Nanavati

November 5, 2024

1 7.4 - Probability and Statistics

1.1 7.4.1 - Frequentist and Bayesian Approaches

- Abstract definition of probability is given by a probability measure satisfying the Kolmogorov axioms for a set S.
- Two main interpretations - frequentist and Bayesian
- Frequentist - A and B are outcomes of repeatable experiments, and $P(A)$ is the frequency of occurrence of A in the limit of infinite experiments
- In this interpretation, probabilities and conditional probabilities are well defined, so $P(data|paramaeters)$ makes sense
- However $P(parameters = \Theta)$ or $P(hypothesis = correct)$ doesn't, since they are not the outcome of experiments
- Bayesian interpretation - allows you to consider the probability that parameters are a certain value or that a hypothesis is correct via Bayes' theorem:
$$P(A|B) = \frac{P(B|A)P(A)}{\sum_i P(B|A_i)P(A_i)}$$
- In this case, A, B do not have to be the outcomes of repeatable experiments, so we are allowed to have $A =$ hypothesis/parameters and $B =$ data to give us
$$P(hypothesis|data) \propto P(data|hypothesis)P(hypothesis)$$
- $P(hypothesis|data) = posterior$, $P(data|hypothesis) = likelihood$, $P(hypothesis) = prior$
- The prior cannot be estimated from trials, so assumptions have to be made to calculate it - 'degree of belief' that the hypothesis is true
- Prior is belief in hypothesis before measurement, posterior is after

- The concepts of confidence intervals and confidence levels vary between the two as well
- Frequentists say that a 90% CL means: say we are measuring x which has a true value x_t. Then we construct $[x_1, x_2]$ such that 90% of all repetitions have x_t in $[x_1, x_2]$
- Say we have a Gaussian centred around x_t and we measure x_0 - we choose x_1, x_2 such that 5% of the area under $P(x|x_1)$ is at $x > x_0$, and similarly for x_2 except now 5% of the area under $P(x|x_2)$ is at $x < x_0$ - Neyman construction, Gaussians centred on probabilistic variables x_1, x_2, x_t fixed
- Bayesian interpretation - x_t has its own probability distribution, not fixed - likelihood function
- Data is that we have observed x_0 and our hypothesis is x_t - denote it as $\Lambda(x_0|x_t)$, which for Gaussian example is $P(x_0|x_t)$
- Now use Bayes to get $P(x_t|x_0)$ - Gaussian centred on x_0 and we set $x_t = x_0$ most probable outcome and a 90% CI by symmetry on the PDF.
- For a Gaussian, the two interpretations give the same result, but not generally true
- Frequentist interpretation is that 90% of CIs generated in the large repetition limit will cover x_t - not true for Bayesian CI, which can have a lower coverage than the stated CL
- We can get differing values of x from the two interpretations when the domain is bounded and observations are near the boundaries
- One example is neutrino mass - cannot be less than 0, but very close to 0, frequentist CL showed upper limit as negative
- The Bayesian approach avoided this by assuming non-negative prior, but then the choice is which prior
- In GW astronomy, we don't have the ability to perform repeated identical trials for parameter estimation, for example, so we need to rely on Bayesian framework

1.2 7.4.2 - Parameter Estimation

- In 7.3, we assumed $h(t)$ has known form, but it relies on a number of free parameters
- For example, temporal width of GW bursts, shape of pulse, distance to sources etc.
- Consider a family of waveforms, or *templates*, given as $h(t; \theta)$ with $\theta = (\theta_1, ..., \theta_n)$

- Correspondingly, family of optimal filters with $\tilde{K}(f;\theta) \propto \frac{\tilde{h}(f;\theta)}{S_n(f)}$
- The question is - given some detection, exceeding a threshold value, how do we reconstruct most probable value of parameters?
- For the likelihood, assume noise is stationary and Gaussian - have the variance from 7.1, so prob. dist. for the noise is

$$p(n_0) = \mathcal{N} exp[-\frac{1}{2}\int_{-\infty}^{\infty} df \frac{|\tilde{n_o}(f)|^2}{(\frac{1}{2})S_n(f)}]$$

- Or in terms of our scalar product notation from last time

$$p(n_0) = \mathcal{N} exp[-\frac{(n_0|n_0)}{2}]$$

- Assuming that the output of the detector is $s(t) = h(t;\theta_t) + n_0(t)$, we get the likelihood as

$$\Lambda(s|\theta_t) = \mathcal{N} exp[-\frac{(s-h(\theta_t)|s-h(\theta_t))}{2}]$$

- Introduce $h_t = h(\theta_t)$ and prior $p^(0)(\theta_t)$ - with some rearranging of the scalar product we get

$$p(\theta_t|s) = \mathcal{N} p^{(0)}(\theta_t) exp[(h_t|s) - \frac{1}{2}(h_t|h_t)]$$

- where we have absorbed a constant, and s is considered constant
- Often this probability space is v high dimensional and complex - binary coalescence is at least 15 parameters f.ex. - want to extract manageable approximations
- To approximate θ_t, we use an estimator - various properties such as consistency, efficiency, bias and robustness to assess estimators - two natural estimators emerge

1.3 Maximum Likelihood Estimation

- The value of θ_t that maximises $\Lambda(s|\theta_t)$, often denoted $\hat{\theta}_{ML}$
- For us this becomes

$$(d_i h_t|s) - (d_i h_t|h_t) = 0$$

- Can show this is the value of θ that maximises signal-to-noise ratio in matched filtering

1.4 Maximum Posterior Estimation

- Sometimes we are better off trying to maximise the full posterior, including the prior term, rather than just the likelihood
- Will no longer give the maximum signal-to-noise ration from matched filtering
- Issue when we consider integrating out some variables - the maximising parameters no longer are the maximising parameters when integrate out some other parameters for non-trivial prior
- Also, neither of these methods gives the estimator with minimal parameter distribution

1.5 Bayes Estimator

- Defined as:
$$\hat{\theta}^i_B(s) = \int d\theta \theta^i p(\theta|s)$$
- The errors matrix is defined as the mean square deviations from the Bayes estimator, where mean refers to integrating over all θ wrt $p(\theta|s)$ so independent of integrating out some parameters
- 'Operational' meaning - averaging out the various possible parameter measurement over the ensemble of signals, and the error matrix is the rms error
- Main drawback - computational cost of multi-dimensional integrals
- In the large signal-to-noise ratio limit ratio (realistic?) all issues become irrelevant, and errors expectation value is related to Fisher Information Matrix

1.6 7.4.3 - Matched Filtering Statistics

- We apply many different templates to the data to get a number of events, and from this we get a list of events - matched filtering
- Applying some parameter estimation we get most probable value of parameters *under hypothesis that a GW signal was present*
- How good is this hypothesis? i.e. statistical significance of a given signal-to-noise ratio?
- Noise can be split into Gaussian 'well-behaved' noise, that falls off rapidly for large values of the argument, as can be eliminated for a sufficiently large threshold value of S/N, and non-Gaussian noise

- Non-Gaussian noise can have long tails at large S/N, decaying as a power law, so small amount of non-Gaussian noise can produce S/N values which are inconceivable by Gaussian
- To veto these, detectors are fitted with equipment to minimise this noise as much as possible, but we can also perform coincidences between detectors to eliminate this as much as possible
- Only consider Gaussian noise - want to study the distribution of the signal to noise ratio, so define

$$\rho = \frac{\hat{s}}{N}$$

- where $\hat{s}$ is the filtered output from 7.3 and N is the rms of $\hat{s}$ when signal absent.
- From the definition of $\hat{s}$, we can get

$$p(\rho|h=0)d\rho = \frac{1}{\sqrt{2\pi}}e^{-\frac{\rho^2}{2}}d\rho$$

- and then we can simply add the case of a signal with S/N $\bar{\rho}$ to get

$$p(\rho|\bar{\rho})d\rho = \frac{1}{\sqrt{2\pi}}e^{-\frac{(\rho-\bar{\rho})^2}{2}}d\rho$$

- ρ is the signal-to-noise ratio in amplitude, and ρ^2 is the signal-to-noise ratio in energy, R.
- Rewriting as the probability distribution for R, we have to remember to include both positive and negative roots for ρ
- Some rearranging gives us

$$P(R|\bar{R})dR = \frac{1}{\sqrt{2\pi R}}e^{\frac{-(R+\bar{R}}{2}}cosh[\sqrt{R\bar{R}}]dR$$

- where $\bar{R}$ is the SNR for the GW signal.
- Looking at expectation values we get

$$\langle R\rangle = \int_{-\infty}^{\infty} dRRP(R|\bar{R}) = 1 + \bar{R}$$

- and defining $T_n = Rk/E$ as the temperature of the noise after matched filtering we get

$$\langle E\rangle = kT_n + \bar{E}$$

- This suggests that there is a threshold value of R which eliminates most of the noise but retains a large fraction of signal distribution, R_t

- We get a false alarm probability of

$$p_{FA} = \int_{R_t}^{\infty} dRP(R|\bar{R} = 0) = 2erfc(\frac{\rho_t}{\sqrt{2}})$$

- and a false dismissal probability of

$$p_{FD} = \int_{0}^{R_t} dRP(R|\bar{R})$$

- R_t determined by our choice of maximum false alarm

- We actually should consider the situation where we have multiple noise outputs, combined in quadrature, so we have $\rho^2 = x_1^2 + ...x_n^2$ and we can recalculate the probability distribution of R as

$$P_n(R|\bar{R}) = \frac{1}{2}(\frac{R}{\bar{R}})^{\frac{(n-2)}{4}} e^{-\frac{(R+\bar{R})}{2}} I_{\frac{n}{2}-1}(\sqrt{R\bar{R}})$$

- where $I_{\frac{n}{2}-1}$ is a modified Bessel function - these are non-central chi-squared densities.

- The average and variance of R with n degrees of freedom is:

$$\langle R \rangle = n + \bar{R} \quad (1)$$
$$\langle E \rangle = n(kT_n) + \bar{E} \quad (2)$$
$$\langle R^2 \rangle - \langle R \rangle^2 = 2n + 4\bar{R} \quad (3)$$