Making sense of data: introduction to statistics for gravitational wave astronomy

Problem Sheet 3: Statistics in Gravitational Wave Astronomy

---

1. (a) As in the question description we denote the two masses by m_1 and m_2, the total mass by $M = m_1 + m_2$, the reduced mass by $\mu = m_1 m_2 / M$, and the chirp mass by

$$\mathcal{M}_c = \frac{m_1^{\frac{3}{5}} m_2^{\frac{3}{5}}}{M^{\frac{1}{5}}}.$$

We will use geometric units throughout, i.e., we set $c = G = 1$ so we don't need to worry about keeping track of these factors.

   i. For a Newtonian binary, the motion is equivalent to that of a body of mass μ orbiting in a fixed Newtonian potential with mass M. Denoting the orbital radius by a (it is also the semi-major axis for a circular binary), the orbital frequency is given by

$$2\pi f = \sqrt{\frac{M}{a^3}}$$

and the total energy of the binary is

$$E = -\frac{M\mu}{2a}.$$

   A. The GW amplitude is determined by the quadrupole moment of the spacetime

$$h \sim \frac{\ddot{I}_{jk}}{D}, \quad I_{jk} = \int \rho x_i x_j \mathrm{d}V.$$

For a binary, the density is only non-zero at the location of the objects. Using the effective-one-body analogy we deduce

$$I \sim \mu a^2 \exp(2\pi i f t)$$

where the frequency is now twice the orbital frequency because we are taking squares of positions, which vary at that frequency. It follows that

$$h \sim \frac{1}{D} f^2 \mu a^2 \sim \frac{1}{D} f^2 \mu \left(\frac{M}{f^2}\right)^{\frac{2}{3}} = \frac{1}{D} f^{\frac{2}{3}} \frac{m_1 m_2}{M^{\frac{1}{3}}} = \frac{1}{D} \mathcal{M}_c^{\frac{5}{3}} f^{\frac{2}{3}}.$$

   B. The GW energy loss is determined by

$$\dot{E}_{\mathrm{GW}} \sim D^2 \dot{h}^2 = \dddot{I}^2 \sim \mu^2 a^4 f^6 \sim \mu^2 f^6 \left(\frac{M}{f^2}\right)^{\frac{4}{3}} = \mu^2 M^{\frac{4}{3}} f^{\frac{10}{3}} = \mathcal{M}_c^{\frac{10}{3}} f^{\frac{10}{3}}.$$

C. The rate of change of frequency is given by

$$\dot{f} \sim \sqrt{\frac{M}{a}} \frac{\mathrm{d}}{\mathrm{dt}} \left(\frac{1}{a} \right) \sim \frac{1}{M\mu} \sqrt{\frac{M}{a}} \dot{E} \sim \mu M^{\frac{1}{3}} (Mf)^{\frac{1}{3}} f^{\frac{10}{3}} = \mu M^{\frac{2}{3}} f^{\frac{11}{3}} = \mathcal{M}_c^{\frac{5}{3}} f^{\frac{11}{3}}.$$

D. The Fourier transform of $h(t)$ is given approximately by

$$\tilde{h} \sim \frac{h}{\sqrt{\dot{f}}} \sim \frac{1}{D} \frac{\mathcal{M}_c^{\frac{5}{3}} f^{\frac{2}{3}}}{\mathcal{M}_c^{\frac{5}{6}} f^{\frac{11}{6}}} = \frac{1}{D} \mathcal{M}_c^{\frac{5}{6}} f^{-\frac{7}{6}}.$$

E. The characteristic strain is given by

$$h_c \sim h \sqrt{\frac{f^2}{\dot{f}}} \sim \frac{1}{D} \mathcal{M}_c^{\frac{5}{3}} f^{\frac{2}{3}} \frac{f}{\mathcal{M}_c^{\frac{5}{6}} f^{\frac{11}{6}}} = \frac{1}{D} \mathcal{M}_c^{\frac{5}{6}} f^{-\frac{1}{6}}.$$

F. The energy density of a GW background generated by a population of these sources is given by

$$\rho_c \Omega_{\mathrm{GW}}(f) = \int_0^\infty \frac{N(z)}{1+z} \left(f_r \frac{\mathrm{d}E}{\mathrm{d}f_r} \right)_{f_r = f(1+z)} \mathrm{d}z.$$

For the inspiraling binaries the previous results give

$$f \frac{\mathrm{d}E}{\mathrm{d}f} \sim f \frac{\dot{E}}{\dot{f}} \sim \mathcal{M}_c^{\frac{5}{3}} f^{\frac{2}{3}}$$

and so we find

$$\Omega_{\mathrm{GW}}(f) \sim \mathcal{M}_c^{\frac{5}{3}} f^{\frac{2}{3}} \int_0^\infty \frac{N(z)}{(1+z)^{\frac{1}{3}}} \mathrm{d}z.$$

ii. The energy of the binary is proportional to $1/a$, hence we have

$$\dot{E}_{\mathrm{hard}} \propto \mu M \frac{\mathrm{d}}{\mathrm{dt}} \left(\frac{1}{a} \right) = k\mu M \frac{\rho_* }{\sigma^3} \frac{m_2}{a} \propto k \frac{\rho_* m_2 \mu}{\sigma^3} (Mf)^{\frac{2}{3}} = k \frac{\rho m_2 \mu}{\sigma^3} M^{\frac{2}{3}} f^{\frac{2}{3}}.$$

iii. The previous derivation of the background energy density assumed that all of the energy loss driving the frequency evolution was due to GW emission. If there are other processes driving energy loss and hence frequency evolution, the background is suppressed because not all of the orbital energy lost is emitted as gravitational waves. In general we have $f = f(E)$ and hence $\dot{f} = (\mathrm{d}f/\mathrm{d}E)\dot{E}$ and therefore

$$\frac{\mathrm{d}E_{\mathrm{GW}}}{\mathrm{d}f} = \frac{\dot{E}_{\mathrm{GW}}}{(\mathrm{d}f/\mathrm{d}E)[\dot{E}_{\mathrm{GW}} + \dot{E}_{\mathrm{other}}]} = \frac{\dot{E}_{\mathrm{GW}}}{\dot{E}_{\mathrm{GW}} + \dot{E}_{\mathrm{other}}} \left(\frac{\mathrm{d}E_{\mathrm{GW}}}{\mathrm{d}f} \right)_{\mathrm{pure \ GW}}.$$

The final bracketed expression denotes the background energy density in the pure GW-driven evolution case. In the case of stellar hardening we therefore find a modified expression for the GW background energy density

$$\rho_c \Omega_{\mathrm{GW}}(f) = \mathcal{M}_c^{\frac{5}{3}} f^{\frac{2}{3}} \int_0^\infty \frac{N(z)}{(1+z)^{\frac{1}{3}}} \frac{\mathcal{M}_c^{\frac{10}{3}}}{\mathcal{M}_c^{\frac{10}{3}} + k(\rho m_2 \mu / \sigma^3) M^{\frac{2}{3}} f^{-\frac{8}{3}} (1+z)^{-\frac{8}{3}}} \mathrm{d}z.$$

This can be simplified a bit more — for example, we notice that the factor $\mu M^{\frac{2}{3}}$ in the hardening term is just $\mathcal{M}_c^{\frac{5}{3}}$ — but the above result is all we need to answer the next few questions.

iv. If the sources are at a common redshift, z_0, we can replace $N(z)$ by a delta function, $\delta(z - z_0)$, and do the integral explicitly. It is then clear that we have

$$\Omega_{\text{GW}}(f) \sim \frac{f^{\frac{2}{3}}}{1 + \lambda f^{-\frac{8}{3}}}$$

where

$$\lambda = k(\rho m_2/\sigma^3)\mathcal{M}_c^{-\frac{5}{3}}(1 + z_0)^{-\frac{8}{3}}.$$

This is a broken power-law, as required. For $f \ll 1$ the term $f^{-\frac{8}{3}}$ dominates in the denominator and we have $\Omega_{\text{GW}} \sim f^{\frac{10}{3}}$. This is the stellar hardening dominated regime. For $f \gg 1$ the constant term dominates in the denominator and we find $\Omega_{\text{GW}} \sim f^{\frac{2}{3}}$. This is the GW dominated regime and this is the standard result for GW backgrounds.

v. If a broken power law background were detected, it tells us about the processes that drive the inspiral of the binary. In this example the power at low frequencies (where hardening dominates) is suppressed relative to that of a pure GW background (see Figure 1). The low frequency slope is characteristic of whatever process drove the early evolution of the binaries — a measurement of this tells you which physical process was important at that time. The high frequency slope tells us about the late evolution of the binary, and in this case the value $f^{\frac{2}{3}}$ is consistent with GW-driven inspiral. The turn over point tells us about the relative efficiencies of the two processes. In this example it occurs where $f \approx \lambda^{\frac{3}{8}}$ and so a measurement of that value tells us about the parameters that go into λ, such as σ, ρ and the typical source redshift, z_0.

vi. (OPTIONAL) No results here. If there is a distribution over masses, then the background energy density involves an integral over the mass distribution as well as the redshift. Try playing around with different choices. Try also including some dependence of ρ and σ on the binary properties. The GW background in the PTA regime may well be suppressed by stellar processes of the type described here. If we see that suppression we will want to be able to interpret it in the context of models of the binary population.

(b) i. The average waveform power is

$$\langle h^2 \rangle = \frac{1}{2T} \int_{-T}^{T} h^2(t)\mathrm{d}t = \frac{1}{2\sqrt{Q}T}\frac{A^2}{D^2} \int_{-\sqrt{Q}T}^{\sqrt{Q}T} \cos^2\left(\frac{2\pi f_0}{\sqrt{Q}}u\right) e^{-u^2} \mathrm{d}u.$$

We see that beyond $\sqrt{Q}T \sim$ few, the waveform is exponentially suppressed. Hence, the duration of the signal is order $\sim 1/\sqrt{Q}$. We take $|\sqrt{Q}T| \lesssim 2$ as a reasonable approximation.
For this choice, we find

$$\langle h^2 \rangle = \frac{A^2}{D^2}\frac{\sqrt{\pi}}{8}\left(\text{erf}(2) + e^{-\left(\frac{2\pi f_0}{\sqrt{Q}}\right)^2}\text{Re}\left[\text{erf}\left(2 + i\frac{2\pi f_0}{\sqrt{Q}}\right)\right]\right) \sim \frac{A^2}{D^2}$$

with a pre-factor that is order 0.few.

ii. Using standard results for Fourier transforms, $\mathcal{F}[g] = \tilde{g}(f)$, including $\mathcal{F}[\exp(-t^2)] = \sqrt{\pi}\exp(-\pi^2 f^2)$, $\mathcal{F}[g(\alpha t)] = \tilde{g}(f/\alpha)/|\alpha|$ and $\mathcal{F}[\exp(2\pi i f_0 t)g(t)] = \tilde{g}(f - f_0)$, we find

$$\tilde{h}(f) = \frac{A}{2D}\sqrt{\frac{\pi}{Q}}\left(e^{-\frac{\pi^2}{Q}(f-f_0)^2} + e^{-\frac{\pi^2}{Q}(f+f_0)^2}\right).$$
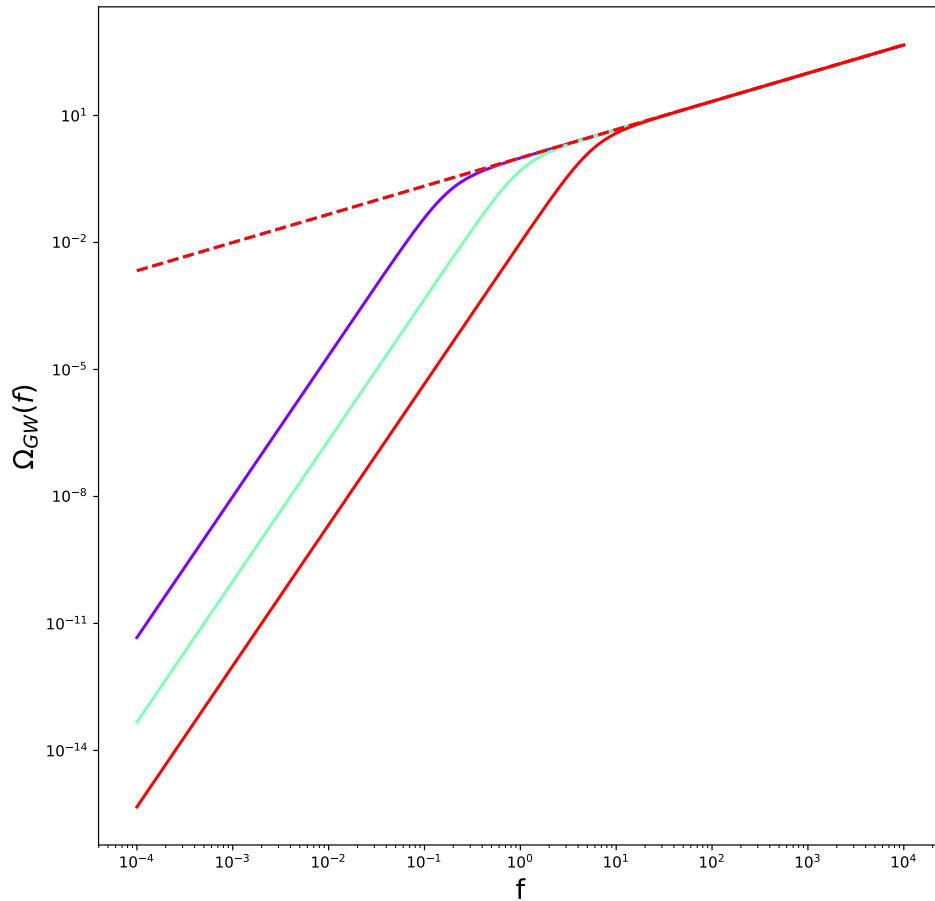
Figure 1: Example backgrounds. We show $\Omega_{\mathrm{GW}}(f)$ as a function of frequency for $\lambda = 0.01$ (purple), $\lambda = 1$ (green) and $\lambda = 100$ (red). Also shown, as a dashed red line, is the background in the absence of stellar hardening.

We can use the fact that the time series is real to wrap onto only positive frequencies and then we have

$$\tilde{h}(f) = \frac{A}{D}\sqrt{\frac{\pi}{Q}}e^{-\frac{\pi^2}{Q}(f-f_0)^2}.$$

We see that the Fourier transform is also proportional to a Gaussian which goes to zero exponentially when $\pi^2(f-f_0)^2/Q \sim$few. Hence the bandwidth is $\Delta f \sim \sqrt{Q}/\pi$.

iii. Using the power ratio formula

$$\left(\frac{S}{N}\right)^2 \approx \frac{\langle h^2 \rangle}{\Delta f S_n(f)}$$

and assuming white noise, $S_n(f) = \sigma^2$, we have

$$\left(\frac{S}{N}\right)^2 \approx k\frac{A^2}{D^2\sqrt{Q}\sigma^2}$$

where k is a constant of order unity. This SNR could be achieved by windowing the data (to the time range $|\sqrt{Q}T| \lesssim$a few) and bandpassing it (to the frequency range $\pi|f - f_0|/\sqrt{Q} \lesssim$a few) and then comparing the signal power to the average off-source noise power.

iv. Using the Fourier transform obtained above, the matched filtering SNR is

$$\left(\frac{S}{N}\right)^2 = 4\int_0^\infty \frac{|\tilde{h}(f)|^2}{S_n(f)}df = \frac{4}{\sigma^2}\frac{A^2\pi}{4D^2Q}e^{-\frac{2\pi^2}{Q}(f-f_0)^2}df \approx \frac{A^2}{2D^2\sigma^2\sqrt{Q}}\int_{-\infty}^\infty e^{-\frac{x^2}{2}}dx$$

which is also equal to $A^2/(D^2\sigma^2\sqrt{Q})$ times a constant of order unity.

We have found that the matched filtering SNR is essentially the same as the burst search SNR, so we are not gaining anything by doing matched filtering. We argued in lectures that matched filtering gained over a burst search by a factor of the square root of the number of cycles spent near a particular frequency. These sine-Gaussian sources are peculiar in that as Q decreases so that the source spends more time near frequency f_0, the bandwidth also decreases so the burst power is increasingly concentrated — we effectively have only '1 cycle' in the vicinity of each relevant frequency. This result does not necessarily mean matched filtering is no better than a burst search — the SNR does not directly translate to a false alarm probability. There may be many instrumental artefacts that could give broadband power in the frequency domain which looks burst like, but those artefacts would look nothing like the specific sine-Gaussian form of the matched filter. Nonetheless, this problem illustrates why excess power searches are quite effective for sources that are burst-like, even if models are available.

v. The energy distribution can be found from

$$\int \frac{\mathrm{d}E}{\mathrm{d}f}\mathrm{d}f = \int_{-\infty}^\infty D^2\dot{h}^2(t)\mathrm{d}t = \int_{-\infty}^\infty D^2 f^2\tilde{h}^2(f)\mathrm{d}f.$$

We find

$$\frac{\mathrm{d}E}{\mathrm{d}f} = A^2\frac{f^2\pi}{2Q}\exp\left(-\frac{\pi^2}{Q}(f - f_0)^2\right).$$

vi. Assuming the number of objects per unit comoving volume with redshift between z and $z + \mathrm{d}z$ and with f_0 between f_0 and $f_0 + \mathrm{d}f_0$ is $N(z)\mathrm{d}f_0\mathrm{d}z$, the background energy density is

$$\rho_c \Omega_{\mathrm{GW}}(f) = \int_0^\infty \int_0^\infty N(z)(1+z)^2 A^2 \frac{f^3 \pi}{2Q} \exp\left(-\frac{\pi^2}{Q}(f(1+z) - f_0)^2\right) f_0^\alpha \mathrm{d}f_0 \mathrm{d}z.$$

vii. The common redshift assumption allows us to replace the integral over z by evaluation of the integrand at z_0 as before. We then have

$$\rho_c \Omega_{\mathrm{GW}}(f) = N_0(1+z_0)^2 A^2 \frac{\pi}{2Q} f^3 \int_0^\infty \exp\left(-\frac{\pi^2}{Q}(f(1+z_0) - f_0)^2\right) f_0^\alpha \mathrm{d}f_0 \mathrm{d}z.$$

The integral over f_0 takes the form

$$\int_0^\infty x^\alpha \exp\left[-(x - \lambda f)^2\right] \mathrm{d}x$$

where $\lambda = \pi(1 + z_0)/\sqrt{Q}$. This integral can be written down as a combination of hypergeometric functions

$$\int_0^\infty x^\alpha \exp\left[-(x - \lambda f)^2\right] \mathrm{d}x = \frac{1}{2}e^{-\lambda^2 f^2}\left[\alpha \lambda f \Gamma\left(\frac{\alpha}{2}\right) {}_1F_1\left(\frac{\alpha}{2} + 1; \frac{3}{2}; \lambda^2 f^2\right)\right.$$
$$\left. + \Gamma\left(\frac{\alpha+1}{2}\right) {}_1F_1\left(\frac{\alpha}{2} + 1; \frac{1}{2}; \lambda^2 f^2\right)\right].$$

The exact background computed from this expression is shown in Figure 2, but we can also find analytic approximations for the low and high frequency behaviour. If $f \ll 1$, then the integral is approximately

$$\int_0^\infty x^\alpha \exp\left[-x^2\right] \mathrm{d}x = \frac{1}{2}\Gamma\left(\frac{\alpha+1}{2}\right)$$

with corrections of order λf. Hence, the dominant behaviour is a constant and $\Omega_{\mathrm{GW}}(f) \sim f^3$ due to the factor out the front of the expression. For $f \gg 1$ we can make a change of variable in the integral

$$\int_0^\infty x^\alpha \exp\left[-(x - \lambda f)^2\right] \mathrm{d}x = \int_{-\lambda f}^\infty (u + \lambda f)^\alpha \exp\left[-u^2\right] \mathrm{d}u$$
$$\approx \lambda^\alpha f^\alpha \int_{-\infty}^\infty \left(1 + \frac{u}{\lambda f}\right)^\alpha \exp\left[-u^2\right] \mathrm{d}u$$
$$= \sqrt{\pi}\lambda^\alpha f^\alpha \left(1 + O\left(\frac{1}{f}\right)\right).$$

So we deduce $\Omega_{\mathrm{GW}} \sim f^{3+\alpha}$.

viii. (OPTIONAL) No results here again, but things to explore would be how the introduction of a redshift distribution modifies things, what happens if the distribution of f_0 is changed, e.g., by introducing a cut-off in the frequency range, what happens if we add a distribution for Q etc.

2. The Fisher Matrix is given by

$$\Gamma_{ij} = \left(\frac{\partial h}{\partial \lambda_i} \middle| \frac{\partial h}{\partial \lambda_j}\right), \qquad \text{where } (a|b) = 2 \int_0^\infty \frac{\tilde{a}^*(f)\tilde{b}(f) + \tilde{a}(f)\tilde{b}^*(f)}{S_h(f)} \mathrm{d}f$$
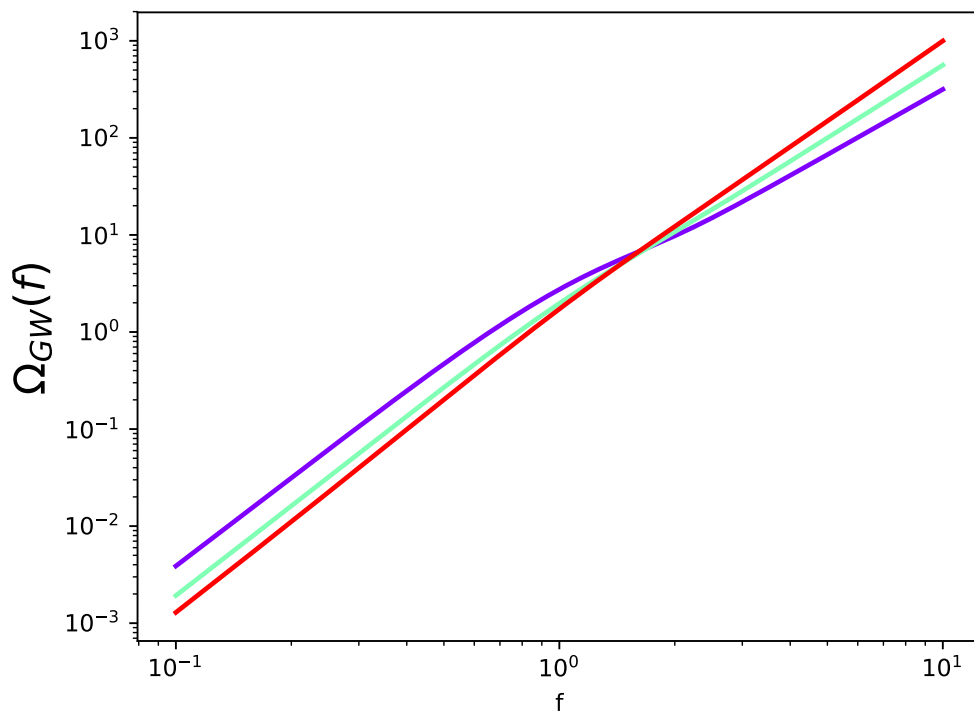
Figure 2: Example backgrounds for the burst population model. We show $\Omega_{\mathrm{GW}}(f)$ as a function of frequency for $\lambda = 1$ and three choices of α: $\alpha = -0.75$ (purple), $\alpha = -0.5$ (green) and $\alpha = -025$ (red).

where $S_h(f)$ is the power spectral density of the detector noise. In this case we are assuming that the source is only observed in the interval $[f_{min}, f_{max}]$ and the PSD is constant in that range and equal to Σ^2. With these assumptions

$$(a|b) = 2\frac{1}{\Sigma^2}\int_{f_{min}}^{f_{max}} \tilde{a}^*(f)\tilde{b}(f) + \tilde{a}(f)\tilde{b}^*(f) \, df.$$

The derivatives of the waveform can be computed as

$$\frac{\partial \tilde{h}}{\partial \mathcal{M}} = \left(\frac{5}{6} - i\left\{\frac{5}{128}(\pi\mathcal{M}f)^{-5/3} + \frac{3}{128}\left(\frac{3715}{756} + \frac{55}{9}\eta\right)\eta^{-\frac{2}{5}}(\pi\mathcal{M}f)^{-2}\right\}\right)\frac{\tilde{h}(f)}{\mathcal{M}}$$

$$\frac{\partial \tilde{h}}{\partial \eta} = i\frac{3}{128}(\pi\mathcal{M}f)^{-1}\left(\frac{11}{3}\eta^{-\frac{2}{5}} - \frac{743}{378}\eta^{-\frac{7}{5}}\right)\tilde{h}(f)$$

$$\frac{\partial \tilde{h}}{\partial \phi_c} = -i\tilde{h}(f)$$

$$\frac{\partial \tilde{h}}{\partial t_c} = 2\pi i f\tilde{h}(f)$$

$$\frac{\partial \tilde{h}}{\partial D_L} = -\frac{1}{D_L}\tilde{h}(f).$$

The key thing to note here is that all of the derivatives are proportional to $\tilde{h}(f)$. When we construct the inner product all terms in the Fisher Matrix are therefore proportional to $|\tilde{h}|^2$, which does not explicitly depend on $\psi(f)$. The Fisher Matrix elements therefore reduce to linear combinations of integrals of the form

$$G(\alpha) = \int_{f_{min}}^{f_{max}} f^{-\alpha} \, df = \frac{1}{\alpha - 1}\left(f_{min}^{\alpha-1} - f_{max}^{\alpha-1}\right).$$

The Fisher Matrix elements are therefore

$$\Gamma_{\mathcal{M}\mathcal{M}} = \frac{4\mathcal{A}^2}{\mathcal{M}^2\Sigma^2}\left(\frac{25}{36}G\left(\frac{7}{3}\right) + \frac{25}{16384}(\pi\mathcal{M})^{-\frac{10}{3}}G\left(\frac{17}{3}\right)\right.$$
$$+ \frac{15}{8192}\left(\frac{3715}{756} + \frac{55}{9}\eta\right)\eta^{-\frac{2}{5}}(\pi\mathcal{M})^{-\frac{11}{3}}G\left(6\right)$$
$$\left.+ \frac{9}{16384}\left(\frac{3715}{756} + \frac{55}{9}\eta\right)^2\eta^{-\frac{4}{5}}G\left(\frac{19}{3}\right)\right)$$

$$\Gamma_{\mathcal{M}\eta} = -\frac{3\mathcal{A}^2}{32\Sigma^2\mathcal{M}}(\pi\mathcal{M})^{-1}\left(\frac{11}{3}\eta^{-\frac{2}{5}} - \frac{743}{378}\eta^{-\frac{7}{5}}\right)\left(\frac{5}{128}(\pi\mathcal{M})^{-\frac{5}{3}}G\left(5\right)\right.$$
$$\left.+ \frac{3}{128}\left(\frac{3715}{756} + \frac{55}{9}\eta\right)\eta^{-\frac{2}{5}}(\pi\mathcal{M})^{-2}G\left(\frac{16}{3}\right)\right)$$

$$\Gamma_{\mathcal{M}\phi_c} = \frac{\mathcal{A}^2}{32\Sigma^2\mathcal{M}}\left(5(\pi\mathcal{M})^{-\frac{5}{3}}G(4) + 3\left(\frac{3715}{756} + \frac{55}{9}\eta\right)\eta^{-\frac{2}{5}}(\pi\mathcal{M})^{-2}G\left(\frac{13}{3}\right)\right)$$

$$\Gamma_{\mathcal{M}t_c} = -\frac{2\pi\mathcal{A}^2}{32\Sigma^2\mathcal{M}}\left(5(\pi\mathcal{M})^{-\frac{5}{3}}G\left(3\right) + 3\left(\frac{3715}{756} + \frac{55}{9}\eta\right)\eta^{-\frac{2}{5}}(\pi\mathcal{M})^{-2}G\left(\frac{10}{3}\right)\right)$$

$$\Gamma_{\mathcal{M}D_L} = \frac{10\mathcal{A}^2}{2\mathcal{M}\Sigma^2 D_L}G\left(\frac{7}{3}\right)$$

$$\Gamma_{\eta\eta} = \frac{9\mathcal{A}^2}{4096\Sigma^2}(\pi\mathcal{M})^{-2}\left(\frac{11}{3}\eta^{-\frac{2}{5}} - \frac{743}{378}\eta^{-\frac{7}{5}}\right)^2 G\left(\frac{13}{3}\right)$$

$$\Gamma_{\eta\phi_c} = -\frac{3\mathcal{A}^2}{32\Sigma^2}(\pi\mathcal{M})^{-1}\left(\frac{11}{3}\eta^{-\frac{2}{5}} - \frac{743}{378}\eta^{-\frac{7}{5}}\right)G\left(\frac{10}{3}\right)$$

$$\Gamma_{\eta t_c} = \frac{3\pi\mathcal{A}^2}{16\Sigma^2}(\pi\mathcal{M})^{-1}\left(\frac{11}{3}\eta^{-\frac{2}{5}} - \frac{743}{378}\eta^{-\frac{7}{5}}\right)G\left(\frac{7}{3}\right)$$

$$\Gamma_{\eta D_L} = 0, \qquad \Gamma_{\phi_c\phi_c} = \frac{4\mathcal{A}^2}{\Sigma^2}G\left(\frac{7}{3}\right), \qquad \Gamma_{\phi_c t_c} = -\frac{8\pi\mathcal{A}^2}{\Sigma^2}G\left(\frac{4}{3}\right), \qquad \Gamma_{\phi_c D_L} = 0$$

$$\Gamma_{t_c t_c} = \frac{16\pi^2\mathcal{A}^2}{\Sigma^2}G\left(\frac{1}{3}\right), \qquad \Gamma_{t_c D_L} = 0, \qquad \Gamma_{D_L D_L} = \frac{4\mathcal{A}^2}{D_L^2\Sigma^2}G\left(\frac{7}{3}\right)$$

The inverse of the Fisher Matrix provides an estimate of parameter estimation precision. We won't attempt toi write down the inverse, but it can be calculated on a case by case basis using the preceding results.

3. We denote the observed data in frequency bin i by s_i. The full data set, denoted D, takes the form

$$s_i = n_i \quad \forall \quad i \in [1, N], \qquad s_{N+1} = n_{N+1} + A, \qquad n_i \sim N(0, \sigma) \quad \forall \quad i \quad (1)$$

with corresponding likelihood

$$p(D|\sigma, A) \propto \frac{1}{\sigma^{N+1}}\left(\prod_{i=1}^{N}\exp\left[-\frac{s_i^2}{2\sigma^2}\right]\right)\exp\left[-\frac{(s_{N+1} - A)^2}{2\sigma^2}\right]. \qquad (2)$$

For flat priors, this is also the posterior, $p(\sigma, A|D)$. We are not so interested in the value of σ, but the value of A, so we can marginalise the posterior over σ. If we write

$$X(A) = (s_{N+1} - A)^2 + \sum_{i=1}^{N} s_i^2 \qquad (3)$$

then we find

$$p(A|D) \quad \propto \quad \frac{1}{\sigma^{N+1}}\exp\left(-X(A)/2\sigma^2\right) \Rightarrow \int_0^\infty p(\sigma, A|D)\mathrm{d}\sigma \propto \left(\frac{2}{X(A)}\right)^{\frac{N}{2}}\Gamma\left(\frac{N}{2}\right)$$

where $\Gamma(x)$ is the gamma function, with $\Gamma(n+1) = n!$. Note that we are assuming a flat prior in σ here, but other priors could be included straightforwardly.

We now recall that the posterior density for the student-t distribution with n degrees of freedom is

$$p_{t,n}(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}.$$

Hence we see that

$$p(A|D) \propto p_{t,N-1}\left(\frac{A - s_{N+1}}{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N} s_i^2}}\right) = p_{t,N-1}\left(\frac{A - s_{N+1}}{\hat{\sigma}}\right)$$

where $\hat{\sigma} = \sqrt{\sum s_i^2/(N-1)}$ is the usual unbiased estimate of the variance. We can compare this to the standard likelihood used in parameter estimation for gravitational wave detectors, which is

$$p(A|D) \propto p_{N(0,1)}\left(\frac{A - s_{N+1}}{\hat{\sigma}}\right)$$

where

$$p_{N(0,1)}(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right]$$

is the pdf of a standard Normal distribution. We see that this marginalisation over the PSD uncertainty is equivalent to replacing the Normal distribution by a student-t distribution. This is the same procedure that we argued could be used for robust regression.

If we take the limit that $N \to \infty$ and write $\sum_{i=1}^{N} s_i^2 = N\sigma_{\text{est}}^2$ we find

$$X(A)^{-\frac{N}{2}} = \left(\frac{1}{N\sigma_{\text{est}}^2}\right)^{\frac{N}{2}} \left(1 + \frac{(s_{N+1} - A)^2}{2(N/2)\sigma_{\text{est}}^2}\right)^{-\frac{N}{2}} \approx B\exp\left(-\frac{(s_{N+1} - A)^2}{2\sigma_{\text{est}}^2}\right) \qquad (4)$$

where the last part follows from the standard result

$$\left(1 + \frac{k}{n}\right)^n \sim e^k \qquad \text{as } n \to \infty \qquad (5)$$

For large N we also expect $\sigma_{\text{est}}^2 = \sigma^2 + O(1/N)$ and so we recover the standard Normal likelihood.

4. (a) The information available before O1 indicates that the rate is uncertain over orders of magnitude. Under these circumstances it is reasonable to suppose that the the log of the rate is uniform in some range. So, we represent the prior as

$$\log_{10}(\lambda) \sim U[-2, 3].$$

This prior has an expectation value of $999.99/\ln(10^5) = 86.858$ and variance of $999999.9999/2\ln(10^5) - 86.86^2 = 35885.13$. The conjugate distribution to a Poisson model is a Gamma distribution, $\Gamma(a, b)$, for which the mean and variance are a/b and a/b^2 respectively. Matching the mean and variance we find $b = 1/413.15$ and $a = 0.210$. We use a conjugate distribution since we then know the posterior will also be in the conjugate family and so it is computationally convenient. [**Note:** any reasonable prior choice is fine, provided it is justified. It must be wide and flat over several decades and make some use of the prior information.]

(b) We note first that all of the observation runs are different lengths. The rate λ was quoted in units of yr^{-1}. Poisson processes are additive, i.e., if the rate in time period T is λ, the rate in time period kT is $\tilde{\lambda} = k\lambda$. If the prior on λ is $\Gamma(a, b)$ then we have

$$p(\tilde{\lambda}) = k^{-1}\frac{b^a}{\Gamma(a)}\left(\frac{\tilde{\lambda}}{k}\right)^{a-1} e^{-b\tilde{\lambda}/k} = \frac{(b/k)^a}{\Gamma(a)}\tilde{\lambda}^{a-1}e^{-(b/k)\tilde{\lambda}},$$

i.e., the prior on $k\lambda$ is $\Gamma(a, b/k)$. As three events are observed in O1, we can write down the posterior distribution on $\tilde{\lambda}$ as $\Gamma(a + 3, b/k + 1)$ and the posterior distribution on λ is $\Gamma(a + 3, b + k)$. In this case using the conjugate prior derived above we have $\Gamma(3.21, 0.252)$. The posterior mean and standard deviation are 12.717 and 7.098 respectively, a 95% symmetric confidence interval is $(2.817, 29.934)$ and the posterior distribution is shown in Figure 3.
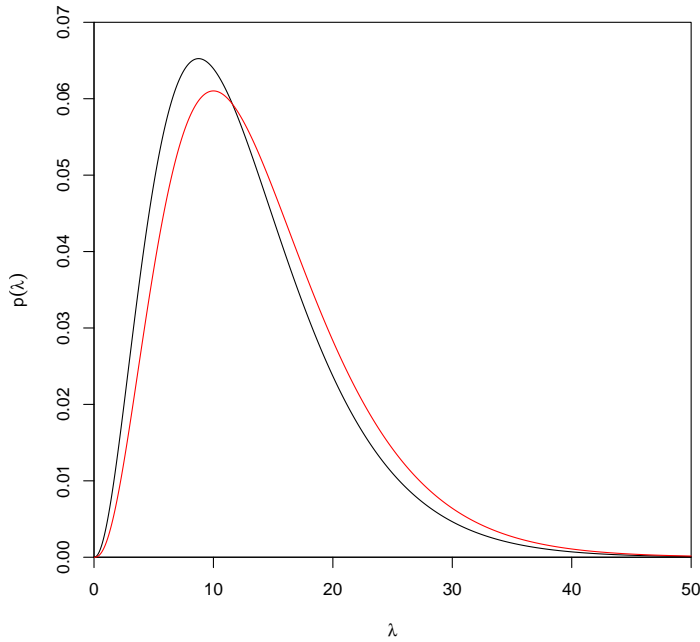
Figure 3: Posterior distribution for the rate per year, λ, after observing the O1 data for the conjugate prior (black line) and the Jeffrey's prior (red line).

(c) The probability that the rate exceeds 15 can be computed from the cumulative density function of the gamma distribution. This is $\gamma(\alpha, \beta x)/\Gamma(\alpha)$, where $\gamma(\alpha, x)$ is the incomplete gamma function. In this case we need $1 - \gamma(3.21, 3.78)/\Gamma(3.21) = 0.312$.

(d) The Jeffrey's prior for the Poisson distribution is the improper prior $p(\lambda) \propto \lambda^{-1/2}$. This can be approximated by a $\Gamma(1/2, \beta)$ distribution with $\beta \to 0$. The posterior on λ from the O1 data with the Jeffrey's prior is therefore $\Gamma(3.5, 0.25)$. The posterior is shown as a red line in Figure 3, the posterior mean and standard deviation are 14 and 7.48 and a 95% symmetric confidence interval is given by $(3.38, 32.0)$. The probability that the rate exceeds 15 is now 0.379. So, the results change a little bit and in particular the Jefffrey's prior favours slightly higher rates than the conjugate prior, but there is not a very big difference between the two.

(e) The second science run, O2, lasts 6 months, so the Poisson rate is 0.5λ. The posterior for O1 therefore gives a prior on the rate in O2 of $\Gamma(3.21, 0.504)$ using the conjugate prior. For a posterior of the form $\Gamma(\alpha, \beta)$, the posterior predictive probability of seeing n events in O2 is therefore

$$p(n|d_{O1}) = \int_0^\infty \frac{\lambda^n e^{-\lambda}}{n!} \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)} d\lambda = \frac{1}{n!} \frac{\beta^\alpha}{(\beta+1)^{\alpha+n}} \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}. \quad (6)$$

This can be recognised as a negative binomial distribution. We compute the probability of seeing 6 or more events in O2 as $1 - \sum_{n=0}^5 p(n|d_{O1}) = 0.504$. The posterior for the rate in the first 5 months of O2 is $\Gamma(3.21, 0.608)$ and the probability of seeing 1 or fewer events in 5 months is given by computing $p(0|d_{O1}) + p(1|d_{O1}) = 0.131$ using this α and β in Eq. (6). The posterior probability for the rate in 1 month is $\Gamma(3.21, 3.024)$, from which we compute

the probability of seeing 5 or more events in one month of O2 as $p_1 = 1 - \sum_{n=0}^4 p(n|d_{O1}) = 0.015$. The probability of seeing 5 or more events in at least one month of O2 is given by $1 - (1 - p_1)^6 = 0.086$. With the Jeffrey's prior the posterior distributions on the rate in 6 months, 5 months and 1 month are $\Gamma(3.5, 0.5)$, $\Gamma(3.5, 0.6)$ and $\Gamma(3.5, 3.0)$ respectively. The probabilities of seeing 6 or more events in O2, 1 or fewer in 5 months of O2, 5 or more in the last month of O2 and seeing 5 or more in at least 1 month of O2 are 0.564, 0.103, 0.019 and 0.110 respectively. The probability that the last month would contain the number of events that were seen is significantly small (at a 2% confidence level). However, there is no reason to single out the last month *a priori* and the probability that one month would be at least this exceptional is only around 10%, which is small but not sufficiently significant to be a cause for concern. The choice of prior does not significantly influence this, indicating that we are data dominated and the conclusion is robust. So, based on O2 we cannot conclude the rate is inhomogeneous in time, but the significance is high enough that we should collect more data and see if the next science run shows any evidence for a time-dependent rate.

(f) In total over O1 and O2 we see 9 events and the total observing time is 0.75 years. Therefore the combined posterior is $\Gamma(9.21, 0.752)$ using the conjugate prior derived in (a) or $\Gamma(9.5, 0.75)$ using the Jeffrey's prior. The posterior predictive distribution for the rate in a given 6 month period of O3 is $\Gamma(9.21, 1.504)$ or $\Gamma(9.5, 1.5)$ respectively. The distribution of the difference $r = |n_1 - n_2|$ of the number of events observed in two independent samples from a Poisson distribution with rate θ is given by the *Skellam distribution* with pmf

$$p(r|\theta) = \begin{cases} e^{-2\theta} I_0(2\theta) & r = 0 \\ 2e^{-2\theta} I_r(2\theta) & r = 1, 2 \dots, \end{cases}$$

where $I_k(x)$ is the modified Bessel function of the first kind. Hence the posterior predictive distribution on r is

$$p(r|d_{1+2}) = \int_0^\infty e^{-2\lambda} I_r(2\lambda) \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)} d\lambda$$

for a $\Gamma(\alpha, \beta)$ posterior distribution on the rate. Note that this can also be written as the difference between two independent negative binomial variables, which follow a *generalised discrete Laplace distribution*, but the expressions that must be evaluated are no easier than this integral.

Table 1 lists the cumulative posterior density for the difference r. We see that there is less than 5% probability of seeing a difference of 7 or more events. Therefore a difference of this size or larger would be significant at a 5% level.

There are a number of other ways in which this question could be addressed. For example, we could look at the number of events in each month and set a threshold, based on the posterior predictive distribution, on the difference between the largest and smallest monthly count. Alternatively, we could model the rates in each one month period as being potentially different, with λ_i denoting the rate in month i. These rates can be connected by a hyperprior, e.g., $\lambda_i \sim \Gamma(\alpha, \beta)$, and the parameters of that hyperprior constrained from the data. Alternatively, the rates can be modelled parametrically, e.g., $\lambda_i = a + bi$, and the parameters of the parametric model constrained from the data. If the

R	$p(r = R|d_{1+2})$	$p(r \leq R|d_{1+2})$	R	$p(r = R|d_{1+2})$	$p(r \leq R|d_{1+2})$
0	0.123	0.123	6	0.048	0.935
1	0.229	0.352	7	0.029	0.965
2	0.194	0.546	8	0.018	0.983
3	0.158	0.704	9	0.009	0.992
4	0.109	0.813	10	0.003	0.995
5	0.075	0.888	11	0.002	0.997

Table 1: Posterior predictive probability of the absolute difference in the number of events detected in the first and second 6 month periods of the O3 science run. The columns give the difference in the number of events, the posterior probability of observing that difference and the cumulative posterior probability of observing a difference less than or equal to that value.

posterior on the slope parameter, b, is inconsistent with 0 there is evidence for an evolving rate. Similarly if the parameters of the hyperprior are inconsistent with a constant rate there is evidence for evolution. The advantage of these kind of approaches is that the results of the analysis give an estimate of the nature and size of the effect, not just the presence of the effect.