Max Planck Institute for Gravitational Physics
IMPRS Lecture Series

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

Problem Sheet 1: Frequentist Statistics and Stochastic Processes

---

Solutions to questions on Frequentist statistics

1. The pdf of the Beta(a,b) distribution is

$$p(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

where

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,\mathrm{d}x$$

is the Beta function. The mode is found by setting the derivative of the pdf to zero

$$(a-1)x^{a-2}(1-x)^{b-1} - (b-1)x^{a-1}(1-x)^{b-2} = 0 \quad \Rightarrow \quad (a+b-2)x = (a-1).$$

So the mode is $(a-1)/(a+b-2)$ unless $a+b=2$, in which case the mode is $x=1$.

To derive the other quantities we need to compute moments of the distribution. This is most easily done using the identity

$$\text{Beta}(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

which you can use without proof [1]. Using this identity we can prove

$$\mathbb{E}(x^n) = \frac{\text{Beta}(a+n,b)}{\text{Beta}(a,b)} = \frac{\Gamma(a+n)\Gamma(a+b)}{\Gamma(a+b+n)\Gamma(a)} = \frac{(a+n-1)\ldots a}{(a+b+n-1)\ldots(a+b)}.$$

The mean is found by setting $n=1$ in the above, giving $a/(a+b)$. The variance can be found using

$$\text{var}(X) = \mathbb{E}(x^2) - \mathbb{E}(x)^2 = \frac{a(a+1)}{(a+b+1)(a+b)} - \frac{a^2}{(a+b)^2} = \frac{ab}{(a+b)^2(a+b+1)}$$

and so on. The skewness is

$$\gamma_1(X) = \frac{2(b-a)\sqrt{a+b+1}}{(a+b+2)\sqrt{ab}}$$

and the excess kurtosis is

$$\text{Ex. Kurt}(X) = \frac{6[(a-b)^2(a+b+1) - ab(a+b+2)]}{ab(a+b+2)(a+b+3)}$$

Solutions that explained how to compute the results and quoted the final results (or derived them using computer algebra packages) were acceptable.

---

[1] The proof involves writing $\Gamma(m)\Gamma(n) = \int_0^\infty \int_0^\infty \mathrm{e}^{-u}u^{m-1}\mathrm{e}^{-v}v^{n-1}\,\mathrm{d}u\,\mathrm{d}v$ and doing a substitution $u = r^2\cos^2\theta$, $v = r^2\sin^2\theta$. After this change of variables the radial part of the integral can be recognised as $\Gamma(m+n)$ immediately. The θ integral is $2\int_0^{\pi/2}\cos^{2m-1}\theta\,\sin^{2n-1}\theta\,\mathrm{d}\theta$, which can be recognized as Beta(m,n) by writing $x=\cos^2\theta$.

2. The joint distribution of (X, Y) is

$$p(x,y) = \frac{1}{\sqrt{2^{n+1}\pi}\Gamma(n/2)} y^{\frac{n}{2}-1} \mathrm{e}^{-\frac{(x^2+y)}{2}}$$

since they are independent. We define two new random variables

$$T = \frac{X}{\sqrt{\frac{Y}{n}}}, \qquad U = Y.$$

The Jacobian matrix for the transformation from (x, y) to (t, u) is

$$J = \begin{pmatrix} \frac{\partial t}{\partial x} & \frac{\partial t}{\partial y} \\ \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{n}{y}} & -\frac{x}{2y}\sqrt{\frac{n}{y}} \\ 0 & 1 \end{pmatrix}$$

from which we deduce the joint pdf of (T, U)

$$p(t,u) = \frac{1}{|J|}p(x,y) = \frac{1}{\sqrt{2^{n+1}n\pi}\Gamma(n/2)} u^{\frac{n-1}{2}} \mathrm{e}^{-\frac{u}{2}\left(1+\frac{t^2}{n}\right)}.$$

We now integrate u out of the distribution to find $p(t)$. We note

$$\int_0^\infty u^{\frac{n-1}{2}} \mathrm{e}^{-\frac{u}{2}\left(1+\frac{t^2}{n}\right)} \mathrm{d}u = \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}} \int_0^\infty \tilde{u}^{\frac{n-1}{2}} \mathrm{e}^{-\frac{\tilde{u}}{2}} \mathrm{d}\tilde{u} = \left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}} \Gamma\left(\frac{n+1}{2}\right) 2^{\frac{n+1}{2}}.$$

Hence we deduce the pdf of t

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

as required.

3. (a) This is the standard Birthday Party Problem. The birthday of each GW source is independent and there are 365 possible birthdays. Therefore there are a total of 365^n possible ways in which the birthdays can be distributed through the year. Out of these possibilities, the number of ways in which all the birthdays are different is the number of ways to choose permutations of size n from a set of 365 possibilities, which is ${}_{365}P_n$. The probability that all the birthdays are different is therefore

$$\frac{365!}{(365-n)!\,365^n}.$$

Evaluating this for $n = 22$ gives 0.524, while for $n = 23$ it gives 0.493, so once 23 events have been observed we are more likely than not to have two on the same day.

(b) If the n events are distinct, then the probability that the new category of event falls on the same date as one of the previous observed events is just $n/365$. If we do not specify that the events are distinct then it is easiest to consider the problem the other way around. The new event singles out 1 date out of 365 that is special. The probability that a particular event in the first category is on a different date is 364/365. The probability that all of the first class of events are on different dates is $(364/365)^n$ and the probability that at least one of the first category of events is on the same day as the new event is $1 - (364/365)^n$. As a LIGO example, the first binary neutron star event was observed after 10 binary black hole events had been observed. The probability that it would be on the same date as a BBH merger is therefore 2.7%, so it would have been surprising if it had coincided with a BBH.

(c) Working in time units of days, the stated rate is $\lambda = 1/7$. The separation of events drawn from a Poisson process with rate λ follows independent $\mathcal{E}(\lambda)$ distributions. After observing n events, we have observed $n-1$ event separations and we are therefore interested in the minimum of $n-1$ independent $\mathcal{E}(\lambda)$ random variables. The probability that this minimum, m, exceeds 1 is

$$\mathbb{P}(m > 1) = (\mathbb{P}(X_1 > 1))^{n-1} = \mathrm{e}^{-\frac{n-1}{7}}.$$

When n is large enough that this is less than 0.5, we are more likely than not to have seen two events separated by less than 24 hours

$$\mathrm{e}^{-\frac{n-1}{7}} < 0.5 \quad \Rightarrow \quad n > 7\ln(2) + 1 = 5.85.$$

So once we have observed 6 events there is a better than 50% chance that there will be two within 24 hours [2].

(d) In this formulation of the problem we ask about time rather than the number of events, so we must marginalise over the latter. After observing for time t, the number of observed events, n, follows a Poisson distribution with rate λt. If $n = 0$ or $n = 1$ the separation of events is definitely more than 1 day. If $n \geq 2$ we must compute the probability that n events distributed randomly in the interval $[0, t)$ have a minimum separation greater than 1 day. Denoting the latter by p_n the probability that the minimum separation is greater than 1 day is

$$\mathbb{P}(m > 1) = \mathrm{e}^{-\lambda t} \left[1 + \lambda t + \sum_{k=2}^{\infty} \frac{(\lambda t)^k}{k!} p_k \right] \tag{1}$$

where $\lambda = 1/7$ as before. It remains to compute p_n, which is the probability that the minimum separation between n points distributed in $[0, t)$ exceeds 1 day. This is equal to the probability that the minimum separation of n points distributed in the interval $[0, 1]$ exceeds $1/t$. There is an extensive literature on the "stick breaking problem", i.e., the distribution of lengths of the pieces of a unit length stick broken at random (see for example L. Holst, *J. Appl. Prob.* **17**, 623-634 (1980), which has been uploaded to the course website along with these solutions for those who are interested). The result we need here is the probability that the first r intervals on a stick broken into $n + 1$ pieces all exceed $x = 1/t$, which is $(1 - rx)^n_+$, where $a_+ = a$ if $a > 0$ and 0 otherwise. In fact, we need the probability for the middle $n - 1$ intervals out of $n + 1$, but as the stick is broken at random the distribution must be symmetric under

[2] One of the submitted solutions answered the different, but also interesting question, of how many days would you have to observe before seeing two events on the same date. In that case, we want to use the probability of seeing less than 2 events on a given day, which is $p = (8/7)\mathrm{e}^{-1/7} = 0.99072$. After n days, the probability that we have seen 2 or more events on a day is $1 - p^n$, which is equal to 0.5 when $n = -\ln(2)/\ln(p) = 74.33$, so we would have to wait 75 days. This is about twice as long as we have to wait to have two events separated by less than 24 hours.

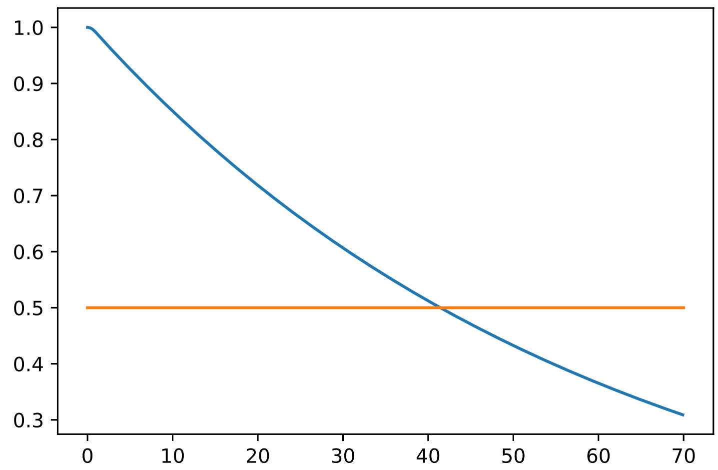

Figure 1: Probability that all event separation will exceed 24 hours as a function of observation time (blue curve). The horizontal orange line indicates a probability of 0.5. The blue curve reaches $p = 0.5$ at $t = 41.4332$.

permutations of the intervals and so this is the same as the probability for the first $n-1$ intervals. We conclude that

$$p_n = \left(1 - (n-1)\frac{1}{t}\right)_+^n .$$

A direct proof of this result is given in the Appendix. Using this result in Eq. (1) we can evaluate the probability as a function of observation time. This is shown in Figure 1. The probability reaches 50% at $t = 41.4332$. During that time, the expected number of observed events is $41.4332/7 = 5.92$, which is close to the $n = 5.85$ found (much more easily) in part (c).

4. Solving

$$\frac{\Pr[X_j = 1]}{\Pr[X_j = 0]} = e^{\rho_j}, \quad \text{and} \quad \Pr[X_j = 0] + \Pr[X_j = 1] = 1$$

for $\Pr[X_j = 0]$ and $\Pr[X_j = 1]$ gives $\Pr[X_j = 0] = \frac{1}{1+e^{\rho_j}}$ and $\Pr[X_j = 1] = \frac{e^{\rho_j}}{1+e^{\rho_j}}$.

Putting these together, we can write $\Pr[X_j = x_j] = \frac{e^{\rho_j x_j}}{(1+e^{\rho_j})}$. The likelihood function for (α, β) is

$$L(\alpha, \beta; \mathbf{x}) = \prod_{j=1}^{n} \Pr[X_j = x_j] = \prod_{j=1}^{n} \frac{e^{\rho_j x_j}}{(1 + e^{\rho_j})} = \frac{\exp\{\sum_{j=1}^{n}(\alpha + \beta z_j)x_j\}}{\prod_{j=1}^{n}(1 + e^{(\alpha+\beta z_j)})}.$$

[Note on sufficiency: let $s_1 = \sum_{j=1}^n x_j$, $s_2 = \sum_{j=1}^n z_j x_j$, $\mathbf{s} = (s_1, s_2)$, $g(\mathbf{s}, \alpha, \beta) = \frac{\exp\{\alpha s_1 + \beta s_2\}}{\prod_{j=1}^n (1+e^{\rho_j})}$ and $h(\mathbf{x}) = 1$. Then, from the Factorization Theorem, $\mathbf{S} = (S_1, S_2) = (\sum_{j=1}^n X_j, \sum_{j=1}^n z_j X_j)$ is sufficient for (α, β).]

To show *minimal* sufficiency, suppose that we have a second set of observations $w_1, w_2, \ldots, w_n$ on X.

The likelihood ratio is

$$\frac{L(\alpha,\beta;\mathbf{x})}{L(\alpha,\beta;\mathbf{w})} = \frac{\exp\left\{\alpha\sum_{j=1}^n x_j + \beta\sum_{j=1}^n z_j x_j\right\}\prod_{j=1}^n(1+e^{\alpha+\beta v_j})}{\prod_{j=1}^n(1+e^{\alpha+\beta z_j})\exp\left\{\alpha\sum_{j=1}^n w_j + \beta\sum_{j=1}^n v_j w_j\right\}}.$$

This will depend on (α,β) unless $\sum_{j=1}^n x_j = \sum_{j=1}^n w_j, \quad \sum_{j=1}^n z_j x_j = \sum_{j=1}^n z_j w_j$. Therefore $(S_1,S_2) = \left(\sum_{j=1}^n X_j, \sum_{j=1}^n z_j X_j\right)$ is minimal sufficient for (α,β).

Note that the explanatory variables are assumed constant and known in each set of observations. If these are not constant or are unknown then the set of sufficient statistics is necessarily larger.

5. The cdf of $X_{(n)}$ is given by

$$\begin{aligned} F(x) &= Pr[X_{(n)} < x] = Pr[X_1 < x, X_2 < x, \ldots, X_n < x] \\ &= Pr[X_1 < x]Pr[X_2 < x] \ \ldots \ Pr[X_n < x] = \left(\frac{x}{\theta}\right)^n, \end{aligned}$$

for $0 \le x \le \theta$, since $X_1,\ldots,X_n$ are independent. Therefore, $f(x) = \frac{dF}{dx} = \frac{nx^{n-1}}{\theta^n}$, for $0 \le x \le \theta$.
For a single observation, $X \sim U[0,\theta]$: $E(X) = \frac{\theta}{2}$ and $\text{var}(X) = E(X^2) - [E(X)]^2 = \frac{\theta^2}{3} - \frac{\theta^2}{4} = \frac{\theta^2}{12}$.
Thus, if $\bar{X}$ is the mean of n observations, we have $E(\bar{X}) = \frac{\theta}{2}$ and $\text{var}(\bar{X}) = \frac{\theta^2}{12n}$, so $E(2\bar{X}) = \theta$ and $\text{var}(2\bar{X}) = \frac{\theta^2}{3n}$.
Therefore, $2\bar{X}$ is unbiased with variance $\to 0$ as $n \to \infty$, hence it is a consistent estimator.

$$\begin{aligned} E(X_{(n)}) &= \int_0^\theta x\frac{nx^{n-1}}{\theta^n}dx = \frac{n}{(n+1)}\theta, \text{ so } E\left(\frac{n+1}{n}X_{(n)}\right) = \theta \\ \text{var}(X_{(n)}) &= \int_0^\theta x^2\frac{nx^{n-1}}{\theta^n}dx - \left[\frac{n}{n+1}\theta\right]^2 = \frac{n\theta^2}{(n+2)} - \frac{n^2\theta^2}{(n+1)^2} \\ &= \frac{n\theta^2}{(n+1)^2(n+2)}[n^2+2n+1-n^2-2n] = \frac{n\theta^2}{(n+1)^2(n+2)} \end{aligned}$$

Thus $\text{var}\left[\frac{n+1}{n}X_{(n)}\right] = \frac{\theta^2}{n(n+2)}$. Hence $\frac{(n+1)}{n}X_{(n)}$ is also an unbaised estimator for θ with variance$\to 0$ as $n \to \infty$, and so is a consistent estimator.

Comment: $\frac{(n+1)}{n}X_{(n)}$ is preferable to $2\bar{X}$ as an estimator for θ as both are unbiased and consistent, but the former can be vastly more efficient.

6. Using θ to denote σ^2, the likelihood is

$$L(\theta) = \frac{(\prod x_i)}{\theta^n}\exp\left\{-\frac{\sum x_i^2}{2\theta}\right\}.$$

The Fisher matrix can be found from

$$
\begin{aligned}
l(\theta) &= -n\log\theta - \frac{\sum x_i^2}{2\theta}, \qquad \frac{\partial l}{\partial\theta} = -\frac{n}{\theta} + \frac{\sum x_i^2}{2\theta^2}, \qquad \frac{\partial^2 l}{\partial\theta^2} = \frac{n}{\theta^2} - \frac{\sum x_i^2}{\theta^3} \\
\Rightarrow \mathbb{E}\left(\frac{\partial^2 l}{\partial\theta^2}\right) &= \frac{n}{\theta^2} - \frac{n\mathbb{E}(X^2)}{\theta^3} \\
\mathbb{E}(X^2) &= \int_0^\infty \frac{x^3}{\theta}\exp\left(-\frac{x^2}{2\theta}\right)dx = \left[-x^2\exp\left(-\frac{x^2}{2\theta}\right)\right]_0^\infty + \int_0^\infty 2x\exp\left(-\frac{x^2}{2\theta}\right)dx \\
&= 2\theta\int_0^\infty \frac{x}{\theta}\exp\left(-\frac{x^2}{2\theta}\right)dx = 2\theta.
\end{aligned}
$$

Giving

$$I_\theta = -\mathbb{E}\left(\frac{\partial^2 l}{\partial\theta^2}\right) = -\left(\frac{n}{\theta^2} - \frac{n2\theta}{\theta^3}\right) = \frac{n}{\theta^2}.$$

The Cramér-Rao lower bound is $\text{var}\,(\widehat{\theta}) \geq \frac{\left(1+\frac{\partial b}{\partial\theta}\right)^2}{I_\theta}$,

$$\text{i.e. } \text{var}\,(\widehat{\theta}) \geq \frac{\theta^2}{n}\left(1+\frac{\partial b}{\partial\theta}\right)^2 \quad \text{where } b = \text{bias}\,(\widehat{\theta}).$$

Since $\frac{\partial l}{\partial\theta} = \frac{n}{\theta^2}(\frac{1}{2n}\sum x_i^2 - \theta) \quad [= I_\theta(\widehat{\theta}-\theta)]$, the bound is attained by the unbiased estimator $\widehat{\theta} = \sum X_i^2/2n$.

7. (a) Likelihood: $L(\theta;\mathbf{x}) = \frac{\prod x_i}{\theta^n}\exp\left(-\frac{1}{2\theta}\sum x_i^2\right)$.
For samples $\mathbf{x}$ and $\mathbf{y}$ consider the ratio $\frac{L(\theta;\mathbf{x})}{L(\theta;\mathbf{y})} = \frac{\prod x_i}{\prod y_i}\exp\left(-\frac{1}{2\theta}(\sum x_i^2 - \sum y_i^2)\right)$.
This does not depend on θ if $\sum x_i^2 = \sum y_i^2$, and thus the statistic $T = \sum X_i^2$ is a minimal sufficient statistic for θ.

(b) Using the Neyman Pearson Lemma, the critical region of the most powerful test of $H_0\colon \theta=\theta_0$ against $H_1\colon \theta = \theta_1$ $(\theta_1 > \theta_0)$ is given by $\frac{L(\theta_1)}{L(\theta_0)} \geq A$, where A is a constant.

$$\text{i.e. } \log L(\theta_1) - \log L(\theta_0) = -n\log\theta_1 - \frac{1}{2\theta_1}\sum x_i^2 + n\log\theta_0 + \frac{1}{2\theta_0}\sum x_i^2 \geq \log A$$

$$\text{i.e. } \frac{1}{2}\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right)\sum x_i^2 \geq \log A + n\log(\frac{\theta_1}{\theta_0})$$

But $\left(\frac{1}{\theta_0} - \frac{1}{\theta_1}\right) > 0$ since $\theta_1 > \theta_0$. Thus, the critical region is of the form $\sum x_i^2 \geq B$, where B is some suitably chosen critical value. Therefore, the test depends on the minimal sufficient statistic T.

For **any** $\theta_1 > \theta_0$, the test of $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ has the **same** form and thus the test is an UMP test of $H_0 : \theta = \theta_0$ against the composite alternative hypothesis $H_1' : \theta > \theta_0$.

(c) $f(x) = \frac{x}{\theta}\exp\left(-\frac{x^2}{2\theta}\right)$. If $y = \frac{x^2}{\theta}$, $\frac{dy}{dx} = \frac{2x}{\theta}$ and thus $f(y) = f(x)|\frac{dx}{dy}| = \frac{1}{2}\exp\left(-\frac{y}{2}\right)$. This is the p.d.f. of an exponential distribution with mean 2 which is a chi-squared distribution with 2 degrees of freedom, i.e. $Y_i \sim \chi_2^2$.
Therefore, under the null hypothesis that $\theta = \theta_0$,

$$\frac{1}{\theta_0}\sum X_i^2 = \sum Y_i \sim \chi^2_{2n},$$

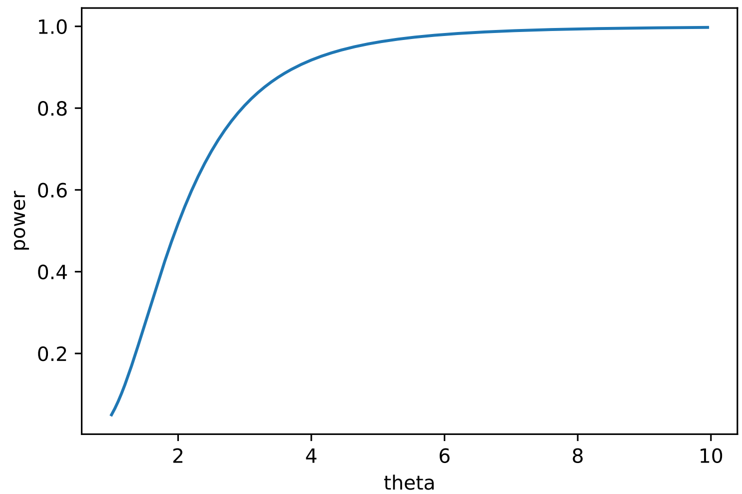

Figure 2: Power of the size 0.05 test as a function of θ. For $\theta = 1$ the power coincides with the size, 0.05, as expected.

using properties of i.i.d. (chi-squared) random variables, which can be used to determine B, i.e. the critical value for a size α test is $B = \theta_0 \chi^2_{2n}(1-\alpha)$.
$H : \theta = 1$ $H' : \theta > 1$: With $n = 5$, the size of the test $\alpha = 0.05$ and $\theta_0 = 1$, the critical value is

$$B = 1 \cdot \chi^2_{10}(0.95) = 18.31.$$

Under the alternative hypothesis, the test statistic, $\sum x_i^2$, is distributed as θ times a χ^2_{10} distribution. Therefore the power of the size α test is

$$\mathbb{P}\left(\chi^2_{10} > \frac{\chi^2_{10}(1-\alpha)}{\theta}\right).$$

This is plotted as a function of θ in Figure 2.

8. The Fisher Matrix is given by

$$\Gamma_{ij} = \left(\frac{\partial h}{\partial \lambda_i}\middle|\frac{\partial h}{\partial \lambda_j}\right), \qquad \text{where } (a|b) = 2\int_0^\infty \frac{\tilde{a}^*(f)\tilde{b}(f) + \tilde{a}(f)\tilde{b}^*(f)}{S_h(f)}\,\mathrm{d}f$$

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

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

The derivatives of the waveform can be computed as

$$\begin{aligned}
\frac{\partial \tilde{h}}{\partial \mathcal{M}} &= \left(\frac{5}{6} - i\left\{\frac{5}{128}(\pi\mathcal{M}f)^{-5/3} + \frac{3}{128}\left(\frac{3715}{756}+\frac{55}{9}\eta\right)\eta^{-\frac{2}{5}}(\pi\mathcal{M}f)^{-2}\right\}\right)\frac{\tilde{h}(f)}{\mathcal{M}} \\
\frac{\partial \tilde{h}}{\partial \eta} &= i\frac{3}{128}(\pi\mathcal{M}f)^{-1}\left(\frac{11}{3}\eta^{-\frac{2}{5}} - \frac{743}{378}\eta^{-\frac{7}{5}}\right)\tilde{h}(f) \\
\frac{\partial \tilde{h}}{\partial \phi_c} &= -i\tilde{h}(f) \\
\frac{\partial \tilde{h}}{\partial t_c} &= 2\pi i f\tilde{h}(f) \\
\frac{\partial \tilde{h}}{\partial D_L} &= -\frac{1}{D_L}\tilde{h}(f).
\end{aligned}$$

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

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

The Fisher Matrix elements are therefore

$$\begin{aligned}
\Gamma_{\mathcal{M}\mathcal{M}} &= \frac{4\mathcal{A}^2}{\mathcal{M}^2\Sigma^2}\left(\frac{25}{36}G\left(\frac{7}{3}\right) + \frac{25}{16384}(\pi\mathcal{M})^{-\frac{10}{3}}G\left(\frac{17}{3}\right)\right. \\
&\qquad + \frac{15}{8192}\left(\frac{3715}{756}+\frac{55}{9}\eta\right)\eta^{-\frac{2}{5}}(\pi\mathcal{M})^{-\frac{11}{3}}G(6) \\
&\qquad\quad \left. + \frac{9}{16384}\left(\frac{3715}{756}+\frac{55}{9}\eta\right)^2\eta^{-\frac{4}{5}}G\left(\frac{19}{3}\right)\right) \\
\Gamma_{\mathcal{M}\eta} &= -\frac{3\mathcal{A}^2}{32\Sigma^2\mathcal{M}}(\pi\mathcal{M})^{-1}\left(\frac{11}{3}\eta^{-\frac{2}{5}} - \frac{743}{378}\eta^{-\frac{7}{5}}\right)\left(\frac{5}{128}(\pi\mathcal{M})^{-\frac{5}{3}}G(5)\right. \\
&\qquad \left. + \frac{3}{128}\left(\frac{3715}{756}+\frac{55}{9}\eta\right)\eta^{-\frac{2}{5}}(\pi\mathcal{M})^{-2}G\left(\frac{16}{3}\right)\right) \\
\Gamma_{\mathcal{M}\phi_c} &= \frac{\mathcal{A}^2}{32\Sigma^2\mathcal{M}}\left(5(\pi\mathcal{M})^{-\frac{5}{3}}G(4) + 3\left(\frac{3715}{756}+\frac{55}{9}\eta\right)\eta^{-\frac{2}{5}}(\pi\mathcal{M})^{-2}G\left(\frac{13}{3}\right)\right) \\
\Gamma_{\mathcal{M}t_c} &= -\frac{2\pi\mathcal{A}^2}{32\Sigma^2\mathcal{M}}\left(5(\pi\mathcal{M})^{-\frac{5}{3}}G(3) + 3\left(\frac{3715}{756}+\frac{55}{9}\eta\right)\eta^{-\frac{2}{5}}(\pi\mathcal{M})^{-2}G\left(\frac{10}{3}\right)\right) \\
\Gamma_{\mathcal{M}D_L} &= \frac{10\mathcal{A}^2}{2\mathcal{M}\Sigma^2 D_L}G\left(\frac{7}{3}\right) \\
\Gamma_{\eta\eta} &= \frac{9\mathcal{A}^2}{4096\Sigma^2}(\pi\mathcal{M})^{-2}\left(\frac{11}{3}\eta^{-\frac{2}{5}} - \frac{743}{378}\eta^{-\frac{7}{5}}\right)^2 G\left(\frac{13}{3}\right)
\end{aligned}$$

$$
\begin{aligned}
\Gamma_{\eta\phi_c} &= -\frac{3\mathcal{A}^2}{32\Sigma^2}(\pi\mathcal{M})^{-1}\left(\frac{11}{3}\eta^{-\frac{2}{5}} - \frac{743}{378}\eta^{-\frac{7}{5}}\right) G\left(\frac{10}{3}\right) \\
\Gamma_{\eta t_c} &= \frac{3\pi\mathcal{A}^2}{16\Sigma^2}(\pi\mathcal{M})^{-1}\left(\frac{11}{3}\eta^{-\frac{2}{5}} - \frac{743}{378}\eta^{-\frac{7}{5}}\right) G\left(\frac{7}{3}\right) \\
\Gamma_{\eta D_L} &= 0, \qquad \Gamma_{\phi_c\phi_c} = \frac{4\mathcal{A}^2}{\Sigma^2}G\left(\frac{7}{3}\right), \qquad \Gamma_{\phi_c t_c} = -\frac{8\pi\mathcal{A}^2}{\Sigma^2}G\left(\frac{4}{3}\right), \qquad \Gamma_{\phi_c D_L} = 0 \\
\Gamma_{t_c t_c} &= \frac{16\pi^2\mathcal{A}^2}{\Sigma^2}G\left(\frac{1}{3}\right), \qquad \Gamma_{t_c D_L} = 0, \qquad \Gamma_{D_L D_L} = \frac{4\mathcal{A}^2}{D_L^2\Sigma^2}G\left(\frac{7}{3}\right)
\end{aligned}
$$

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

Solutions to questions on Stochastic Processes

9. (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

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

C. The rate of change of frequency is given by

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

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

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

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

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

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

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

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

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

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

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

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

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

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 $\Omega_{\rm GW} \sim f^{\frac{10}{3}}$. This is the stellar hardening dominated regime. For $f \gg 1$ the constant term dominates in the denominator and we find $\Omega_{\rm GW} \sim f^{\frac{2}{3}}$. 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 3). 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

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

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.

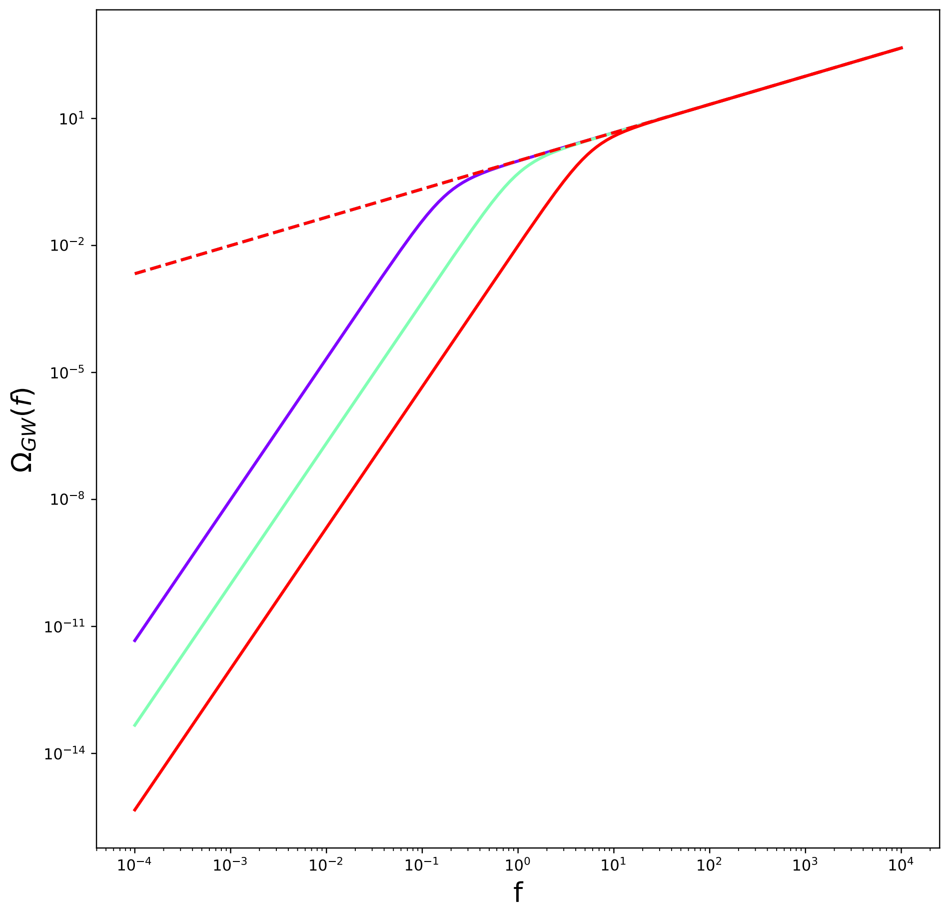

Figure 3: Example backgrounds. We show $\Omega_{\rm GW}(f)$ as a function of frequency for $\lambda = 0.01$ (purple), $\lambda = 1$ (green) and $\lambda = 100$ (red). Also shown, as a dashed red line, is the background in the absence of stellar hardening.

For this choice, we find

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

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

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

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

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

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

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

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

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

The exact background computed from this expression is shown in Figure 4, 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

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

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

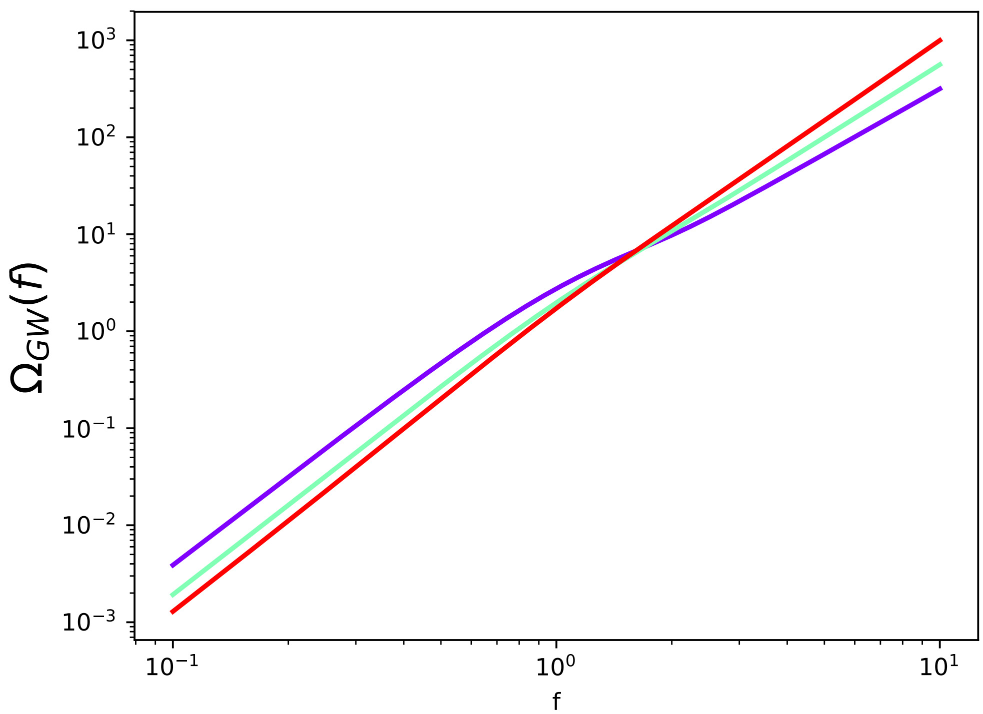

Figure 4: Example backgrounds for the burst population model. We show $\Omega_{\rm GW}(f)$ as a function of frequency for $\lambda = 1$ and three choices of α: $\alpha = -0.75$ (purple), $\alpha = -0.5$ (green) and $\alpha = -025$ (red).

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.

Solutions to additional questions on Frequentist statistics

10. This can be proven by induction. We write

$$I_n = \int_{-\infty}^{\infty} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \mathrm{d}x.$$

Proving the t-distribution is properly normalised is equivalent to proving that

$$I_n = \frac{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)}.$$

Setting $n = n+2$ in the above we find

$$\frac{\sqrt{(n+2)\pi}\Gamma\left(\frac{n}{2}+1\right)}{\Gamma\left(\frac{n+1}{2}+1\right)} = \sqrt{\frac{n+2}{n}} \frac{n}{n+1} \frac{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)}$$

which follows from the identity $\Gamma(n+1) = n\Gamma(n)$. Therefore, if we can show that $I_1 = \sqrt{\pi}\Gamma(1/2) = \pi$, $I_1 = \sqrt{2\pi}/\Gamma(3/2) = \sqrt{2\pi}/(\sqrt{\pi}/2) = 2\sqrt{2}$ and

$$I_{n+2} = \sqrt{\frac{n+2}{n}} \frac{n}{n+1} I_n$$

the result follows by induction. Firstly we note

$$I_1 = \int_{-\infty}^{\infty} \left(1 + x^2\right)^{-1} \mathrm{d}x = \left[\tan^{-1}(x)\right]_{-\infty}^{\infty} = \frac{\pi}{2} + \frac{\pi}{2} = \pi$$

and

$$I_2 = \int_{-\infty}^{\infty} \left(1 + \frac{x^2}{2}\right)^{-\frac{3}{2}} \mathrm{d}x = \int_{-\infty}^{\infty} \sqrt{2}\mathrm{sech}^2 u \mathrm{d}u = \sqrt{2}\left[\tanh(u)\right]_{-\infty}^{\infty} = 2\sqrt{2}.$$

where we used the substitution $x = \sqrt{2}\sinh u$. Finally, we prove the recurrence relation

$$I_n = \int_{-\infty}^{\infty} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}+1}} \mathrm{d}x + \int_{-\infty}^{\infty} \frac{x^2}{n\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}+1}} \mathrm{d}x.$$

We can use a substitution $x^2/n = u^2/(n+2)$ in the first integral to put it in the form of I_{n+2}. For the second term we can integrate by parts, writing $u = x$, $\mathrm{d}v/\mathrm{d}x = x/(n(1+x^2/n)^{(n+3)/2}$. We obtain

$$\begin{aligned} I_n &= \sqrt{n}n + 2I_{n+2} + \frac{1}{n+1}\int_{-\infty}^{\infty} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \mathrm{d}x = \sqrt{n}n + 2I_{n+2} + \frac{1}{n+1} I_n \\ \Rightarrow \quad I_n &= \frac{n+1}{n}\sqrt{\frac{n}{n+2}} I_{n+2} \qquad (2) \end{aligned}$$

as required.

11. The MGF for the exponential distribution can be found via

$$M_X(t) = \mathbb{E}\left[e^{tX}\right] = \int_0^\infty e^{tx}\lambda e^{-\lambda x} = \frac{\lambda}{\lambda - t}.$$

Similarly, for the Gamma(n, λ) distribution we have

$$\begin{aligned} M_X(t) &= \frac{1}{\Gamma(n)}\int_0^\infty e^{tx}\lambda^n x^{n-1}e^{-\lambda x}\,dx \\ &= \left(\frac{\lambda}{\lambda - t}\right)^n \int_0^\infty \frac{1}{\Gamma(n)}(\lambda - t)^n x^{n-1}e^{-(\lambda - t)x}\,dx \\ &= \left(\frac{\lambda}{\lambda - t}\right)^n. \end{aligned}$$

The MGF for a sum of n IID random variables, each of which has MGF $M_X(t)$, is $M_X(t)^n$. Hence we deduce that the sum of n IID $\mathcal{E}(\lambda)$ random variables is a $\Gamma(n, \lambda)$ distribution, as required.

12. The results in this question can also be obtained using results from the theory of stick breaking. We have n birthdays distributed randomly over the year, which we can represent as a circle with unit circumference. The first birthday is arbitrary, but once this is specified it sets a zero point on the circle, which we can think of as representing the two ends of the stick that have been identified with one another. The remaining $(n-1)$ birthdays are distributed randomly around the circle (or along the stick) and therefore the full set of n birthdays represents a random partition of the stick into n pieces. To answer part (a) we need the distribution of the maximum length of a piece, while to answer part (b) we need the distribution of the minimum length of a piece. The corresponding results may also be found in the Appendix.

    (a) To answer this question we need the probability that the pieces of a unit-length stick broken into n parts are all less than $x = 1/26$ (which corresponds to 2 weeks). This is shown in the Appendix to be given by

    $$\sum_{j=0}^{n+1}(-1)^j \binom{n}{j} (1 - jx)_+^{n-1}. \tag{3}$$

    This can be evaluated numerically and is plotted in Figure 5. We conclude that the must be 138 members in the institute before Andrew gets his cake at least every two weeks!

    (b) To answer this question we need the probability that the minimum length of pieces of a stick broken into n parts exceeds x, which is shown in the Appendix to be $(1-nx)_+^{n-1}$. This can be evaluated numerically and is shown in Figure 6. We see that even with as few as $n = 5$ members in the institute there is a greater than 50% chance that the minimum separation between birthdays is less than 2 weeks. So, Alice should employ at most 4 people if she wants to protect her members' health.

13. Let n_{m+1} be the number of items which survive to time mh (so that $n = \sum_{r=1}^{m+1} n_r$),

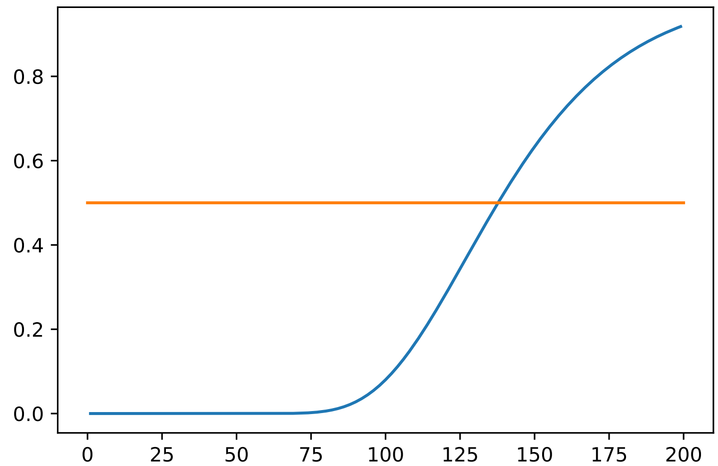

Figure 5: Probability that the longest spacing between birthdays is less than 2 weeks as a function of the number of members of the institute (blue curve). The horizontal orange line indicates a probability of 0.5. The blue curve reaches $p = 0.5$ between $n = 137$ and $n = 138$.

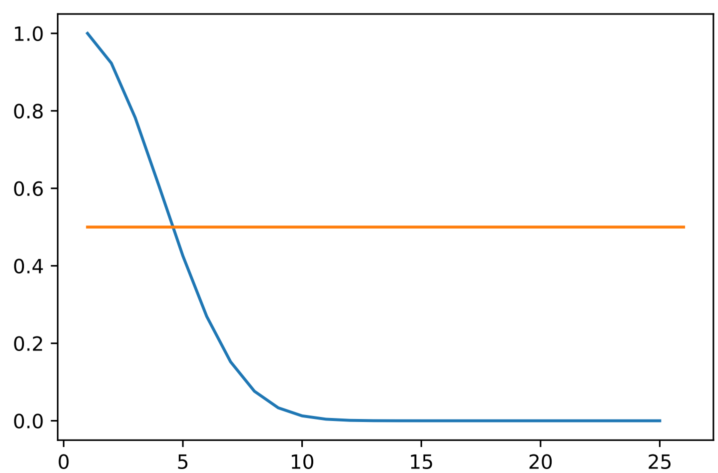

Figure 6: Probability that the shortest spacing between birthdays is greater than 2 weeks as a function of the number of members of the institute (blue curve). The horizontal orange line indicates a probability of 0.5. The blue curve reaches $p = 0.5$ between $n = 4$ and $n = 5$.

and let $\gamma = e^{-h\lambda}$. The probability that an item fails in the interval $((r-1)h, rh)$ is

$$\begin{aligned} p_r &= \Pr((r-1)h < T < rh) \\ &= F_T(rh|\,\lambda) - F_T((r-1)h|\,\lambda) \\ &= (1 - e^{-rh\lambda}) - (1 - e^{-(r-1)h\lambda}) \\ &= \gamma^{r-1} - \gamma^r \\ &= \gamma^{r-1}(1-\gamma) \qquad (r = 1, \ \ldots \ , m); \end{aligned}$$

the probability of surviving to time mh is

$$p_{m+1} = \Pr(T > mh) = e^{-mh\lambda} = \gamma^m.$$

The joint distribution of $(N_1, N_2, \ \ldots \ , N_{m+1})$ is $Mult(n, p_1, \ \ldots \ , p_{m+1})$, so the likelihood function is

$$\begin{aligned} L(\lambda) = n! \prod_{r=1}^{m+1} \frac{p_r^{n_r}}{n_r!} &= \frac{n! \prod_{r=1}^{m} \{\gamma^{r-1}(1-\gamma)\}^{n_r} \times \gamma^{mn_{m+1}}}{\prod_{r=1}^{m+1} n_r!} \\ &= \left\{ \frac{n!}{\prod_{r=1}^{m+1} n_r!} \right\} \cdot \ (\gamma^{s_1}(1-\gamma)^{s_2}) \end{aligned}$$

where $s_1 = \sum_{r=1}^{m+1}(r-1)n_r$, $s_2 = \sum_{r=1}^{m} n_r = n - n_{m+1}$. Therefore, by the Factorization Theorem, (S_1, S_2) is sufficient for λ. [Note: (S_1, N_{m+1}) is also sufficient for λ.]

14. The likelihood for $\boldsymbol{\theta} = (\alpha, \beta)$ is

$$\begin{aligned} L(\alpha, \beta; \mathbf{x}) &= \prod_{i=1}^{n} (\alpha + i\beta) \exp\{-(\alpha + i\beta)x_i\} \\ &= \{\prod_{i=1}^{n} (\alpha + i\beta)\} \exp\{-\alpha \sum_{i=1}^{n} x_i\} \exp\{-\beta \sum_{i=1}^{n} ix_i\}. \end{aligned}$$

Let $\mathbf{s} = (s_1, s_2) = (\sum_{i=1}^n x_i, \sum_{i=1}^n ix_i)$. Using the Factorization Theorem with

$$g(\mathbf{s}, \alpha, \beta) = \{\prod_{i=1}^{n} (\alpha + i\beta)\} \exp\{-\alpha s_1\} \exp(-\beta s_2\} \quad \text{and} \quad h(\mathbf{x}) = 1$$

we see that $\mathbf{S} = (S_1, S_2) = (\sum_{i=1}^n X_i, \sum_{i=1}^n iX_i)$ is sufficient for (α, β).

15. The likelihood is

$$L(\lambda; \mathbf{x}) = \prod_{i=1}^{n} f(x_i \mid \lambda) = \lambda^n e^{-\lambda \sum_{i=1}^n x_i}, \ l(\lambda; \mathbf{x}) = \ln L(\lambda) = n \ln\lambda - \lambda \sum x_i, \qquad (4)$$

$$\text{and} \ \frac{\partial l}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i. \qquad (5)$$

MLE: Equating $\partial l / \partial \lambda$ to zero gives $\widehat{\lambda} = n / \sum x_i$ or $1/\overline{x}$, and it can be verified that this corresponds to a maximum.

Mean: To compute $\mathbb{E}(1/\bar{X})$ we note that $Y = \sum X_i$ has a gamma distribution with p.d.f. $\lambda^n y^{n-1} e^{-\lambda y}/\Gamma(n)$, $y > 0$, and $1/\overline{x}$ is n/y, so

$$\begin{aligned} E\left(\frac{1}{\bar{X}}\right) &= E\left(\frac{n}{Y}\right) = \int_0^\infty \frac{\lambda^n}{\Gamma(n)} y^{n-1} e^{-\lambda y} \frac{n}{y} dy \\ &= \frac{n\lambda}{(n-1)} \int_0^\infty \frac{\lambda^{n-1}}{\Gamma(n-1)} y^{n-2} e^{-\lambda y} dy = \frac{n\lambda}{(n-1)} \end{aligned}$$

Variance: We first compute

$$\mathbb{E}\left[\left(\frac{1}{\bar{X}}\right)^2\right] = \frac{n^2\lambda^2}{(n-1)(n-2)},$$

and then deduce

$$\mathrm{var}\left(\frac{1}{\bar{X}}\right) = \frac{n^2\lambda^2}{(n-1)}\left[\frac{1}{(n-2)} - \frac{1}{(n-1)}\right] = \frac{n^2\lambda^2}{(n-1)^2(n-2)} \to 0 \text{ as } n \to \infty.$$

Cramér-Rao bound: The second derivative of the log-likelihood is

$$\frac{\partial^2 l}{\partial \lambda^2} = -\frac{n}{\lambda^2},$$

which is constant and therefore equal to its expectation value, which is minus the Fisher matrix. Therefore $I_\lambda = n/\lambda^2$.

Bias: The bias is

$$b(\lambda) = \frac{n\lambda}{(n-1)} - \lambda = \frac{\lambda}{(n-1)},$$

so the MLE is biased but asymptotically unbiased.

Consistency: The bias $(\frac{1}{\bar{X}}) \to 0$ and $\mathrm{var}(\frac{1}{\bar{X}}) \to 0$ as $n \to \infty \Rightarrow \frac{1}{\bar{X}}$ is consistent.

Asymptotic efficiency: $\frac{\mathrm{var}(\frac{1}{\bar{X}})}{\frac{\lambda^2}{n}} \to 1$ as $n \to \infty$. Therefore $\frac{1}{\bar{X}}$ is asymptotically efficient.

16. The expectation value of X_1 is

$$\mathbb{E}(X_1) = 0 \times (1-p) + 1 \times p = p$$

so it is an unbiased estimator of p. The variance is

$$\mathrm{var}(X_1) = \mathbf{E}(X_1^2) - p^2 = p - p^2 = p(1-p).$$

The combined likelihood is

$$L(p; \mathbf{x}) = p^{\sum x_i}(1-p)^{n-\sum x_i} = \left(\frac{p}{1-p}\right)^{\sum x_i} (1-p)^n$$

and from the factorisation theorem we recognize $S = \sum X_i$ as a sufficient statistic.

When $X_1 = 1$:

$$\Pr\left[X_1 = 1 \mid \sum_{i=1}^n X_i = t\right] = \frac{\Pr\left[X_1 = 1; \sum_{i=1}^n X_i = t\right]}{\Pr\left[\sum_{i=1}^n X_i = t\right]} = \frac{\Pr\left[X_1 = 1; \sum_{i=2}^n X_i = t - 1\right]}{\Pr\left[\sum_{i=1}^n X_i = t\right]}$$
$$= \frac{\theta \cdot \binom{n-1}{t-1} \theta^{t-1}(1-\theta)^{n-1-t+1}}{\binom{n}{t} \theta^t (1-\theta)^{n-t}} = \frac{t}{n}.$$

When $X_1 = 0$:

$$\Pr(X_1 = 0 \mid \sum X_i = t) = 1 - \Pr(X_1 = 1 \mid \sum X_i = t) = 1 - \frac{t}{n} = \frac{n-t}{n}$$

Note that the conditional distribution of X_1 given $\sum X_i = t$ is independent of θ, as it should be. Therefore

$$\widehat{\theta}_T = \mathbb{E}\left[X_1 \mid \sum_{i=1}^n X_i = t\right] = 0 \cdot \frac{n-t}{n} + 1 \cdot \frac{t}{n} = \frac{t}{n}$$

i.e.

$$\widehat{\theta}_T = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}.$$

We deduce that the sample mean, $\bar{X}$, is a better estimator. It's variance is $p(1-p)/n$, which is smaller than that of X_1, as expected.

17. (a) The likelihood for the observed data is

$$p(\mathbf{y}|\mathbf{X}, \beta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta)\right]$$

and so maximising the likelihood is equivalent to minmising the sum of squares

$$(\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta).$$

Differentiating with respect to (each component of) β and setting the derivatives to zero gives

$$\mathbf{X}^T\mathbf{X}\beta - \mathbf{X}^T\mathbf{y} = 0 \quad \Rightarrow \quad \hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

as required.

(b) The above estimator is a linear combination of normally distributed random variables (the y_i's) and hence is normally distributed. The mean is found via

$$\mathbb{E}(\hat{\beta}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbb{E}(\mathbf{y}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\beta = \beta.$$

The covariance of a linear combination of random variables $\mathbf{A}\mathbf{y}$ is $\mathbf{A}$ cov($\mathbf{y}$) $\mathbf{A}^T$ and so we deduce

$$\text{cov}(\hat{\beta}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}.$$

We deduce

$$\hat{\beta} \sim N\left(\beta, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\right)$$

as required.

(c) We write $\tilde{y}_i = y_i - (\mathbf{X}\beta)_i$ and note

$$\mathbb{E}(\tilde{y}_i\tilde{y}_j) = \mathrm{cov}(\mathrm{y_i}, \mathrm{y_j}) = \sigma^2\delta_{\mathrm{ij}}.$$

The quantity

$$\begin{aligned}
\mathbf{y}^T\mathbf{y} - \hat{\beta}^T\mathbf{X}^T\mathbf{y} &= (\tilde{\mathbf{y}} + \mathbf{X}\beta)^T(\tilde{\mathbf{y}} + \mathbf{X}\beta) - (\tilde{\mathbf{y}} + \mathbf{X}\beta)^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\tilde{\mathbf{y}} + \mathbf{X}\beta) \\
&= \tilde{\mathbf{y}}^T\tilde{\mathbf{y}} - \tilde{\mathbf{y}}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\tilde{\mathbf{y}} \\
&= y_iy_i - y_ix_{ij}(\mathbf{X}^T\mathbf{X})^{-1}_{jk}x_{lk}y_l \qquad (6)
\end{aligned}$$

where we introduced Einstein summation convention in the last line. We now take the expectation value

$$\begin{aligned}
\mathbb{E}\left(\mathbf{y}^T\mathbf{y} - \hat{\beta}^T\mathbf{X}^T\mathbf{y}\right) &= \sigma^2\left(\delta ii - x_{ij}(\mathbf{X}^T\mathbf{X})^{-1}_{jk}x_{ik}\right) \\
&= \sigma^2\mathrm{Tr}\left(\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\right) \\
&= \sigma^2\mathrm{Tr}\left(\mathbf{I}_n - (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\right) = \sigma^2\mathrm{Tr}\left(\mathbf{I}_n - \mathbf{I}_k\right) \\
&= \sigma^2(n-k). \qquad (7)
\end{aligned}$$

Here we use $\mathbf{I}_k$ to denote the $k \times k$ identity matrix. The quoted result follows. As mentioned in the question, the quantity $(n-k)\hat{\sigma}^2$ is independent of $\hat{\beta}$ and follows a χ^2 distribution with $(n-k)$ degrees of freedom. We won't give a detailed proof, but this is most easily seen by decomposing the observations $\mathbf{y}$ into a model-parallel and model-orthogonal piece. In particular

$$\mathbf{y}^T\mathbf{y} - \hat{\beta}^T\mathbf{X}^T\mathbf{y} = \left(\mathbf{y} - \mathbf{X}\hat{\beta}\right)^T\left(\mathbf{y} - \mathbf{X}\hat{\beta}\right).$$

This is the sum of squares of the residual, i.e., the difference between the observed data and the part of it that can be explained by the best-fit model. The elements of the residual, $\mathbf{e} = \left(\mathbf{y} - \mathbf{X}\hat{\beta}\right)$, are linear combinations of Normally distributed random variables and so also follow a Normal distribution. The covariance between the residual and the model parameter estimator is

$$\mathrm{cov}(\mathbf{e}, \hat{\beta}) = \mathrm{cov}\left(\mathbf{y}, \mathbf{y}\right)\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} - \mathbf{X}\mathrm{cov}(\hat{\beta}, \hat{\beta}) = \sigma^2\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} - \sigma^2\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} = 0.$$

While zero covariance does not imply independence in general, this is true for normally distributed random variables. We deduce that $\mathbf{e}$, and hence $\hat{\sigma}^2$, are independent of $\hat{\beta}$. The estimator $\hat{\sigma}^2$ is a sum of squares of zero mean normal random variables and so will follow a chi-squared distribution. However, not all n components of $\mathbf{e}$ can be independent, since we started with n random variables and k of them are used to determine the components of $\hat{\beta}$. A more careful analysis decomposes the observations into a set of k components that lie in the model space, which give $\hat{\beta}$, and a set of $n-k$ components orthogonal to the model space, the sum of squares of which give $\mathbf{e}^T\mathbf{e}$. So the latter is σ^2 times a chi-squared distribution with $n-k$ degrees of freedom.

(d) The estimator $\mathbf{c}^T\hat{\beta}$ is normally distributed with mean

$$\mathbb{E}(\mathbf{c}^T\hat{\beta}) = \mathbf{c}^T\beta$$

and variance

$$\Sigma^2 = \mathbf{c}^T\mathrm{cov}(\hat{\beta},\hat{\beta})\mathbf{c} = \sigma^2\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}.$$

The normalised estimator

$$\frac{\mathbf{c}^T\hat{\beta} - \mathbf{c}^T\beta}{\sigma\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}} \sim N(0,1)$$

is standard normal. We do not know σ, but

$$\hat{\sigma}^2 = \frac{\sigma^2}{n-k}\chi$$

where $\chi \sim \chi^2_{n-k}$. Therefore

$$\frac{\mathbf{c}^T\hat{\beta} - \mathbf{c}^T\beta}{\hat{\sigma}\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}} = \frac{Z}{\sqrt{\chi/(n-k)}}, \quad \text{where } Z \sim N(0,1) \quad \text{and } \chi \sim \chi^2_{n-k}$$

which is the definition of a t-distribution with $(n-k)$ degrees of freedom.
A $100(1-\alpha)\%$ confidence interval for $\mathbf{c}^T\beta$ is then

$$\mathbf{c}^T\beta - \hat{\sigma}\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}\,t_{\frac{\alpha}{2}} < \mathbf{c}^T\beta < \mathbf{c}^T\beta + \hat{\sigma}\sqrt{\mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}}\,t_{\frac{\alpha}{2}}$$

where $t_{\frac{\alpha}{2}}$ is the upper $\alpha/2$ point (i.e., the point corresponding to $(1-\alpha/2)$ in the cdf) of the t_{n-k}-distribution.

18. (a) According to the Neyman-Pearson lemma, the most powerful test of size α for testing the simple null hypothesis $H_0 : \theta = 1$ against the simple alternative hypothesis $H_1 : \theta = \theta_1$ $(\theta_1 > 1)$ has critical regions of the form

$$\begin{aligned}
\{\mathbf{y} : \frac{f(\mathbf{y} \mid \theta_1)}{p(\mathbf{y} \mid \theta = 1)} > K_\alpha\} &= \{\mathbf{y} : \frac{\prod_i p(x_i \mid \theta_1)}{\prod_i p(x_i \mid \theta = 1)} > K_\alpha\} \\
&= \{\mathbf{y} : \frac{\theta_1^{na} e^{-\theta_1 \sum_i z_i x_i}}{e^{-\sum_i z_i x_i}} > K_\alpha\} \\
&= \{\mathbf{y} : \theta_1^{na} e^{(1-\theta_1)\sum_i z_i x_i} > K_\alpha\} \\
&= \{\mathbf{y} : na\log\theta_1 + (1-\theta_1)\sum_i z_i x_i > \log K_\alpha\} \\
&= \{\mathbf{y} : (1-\theta_1)\sum_i z_i x_i > \log K_\alpha - na\log\theta_1\} \\
&= \{\mathbf{y} : \sum_i z_i x_i < C_\alpha\}
\end{aligned}$$

since $(1-\theta_1) < 0$.
Constant C_α can be found from the condition that

$$P(\sum_i z_i Y_i < C_\alpha \mid H_0) = \alpha.$$

To find the distribution of $\sum_i z_i X_i$, we can either use the Central Limit theorem to find the distribution approximately, or we can find it exactly. Since $z_i X_i \sim$

$\Gamma(a,\theta)$ independently, $\sum_i z_i X_i \sim \Gamma(an,\theta)$ or equivalently $\theta \sum_i z_i X_i \sim \Gamma(an,1)$. Therefore, since under H_0 $\theta = 1$,

$$\alpha = P(\sum_i z_i X_i < C_\alpha \mid H_0) = F_{\Gamma(an,1)}(C_\alpha),$$

which implies that $C_\alpha = F^{-1}_{\Gamma(an,1)}(\alpha)$.
Alternatively, using the approximation, we have that $z_i X_i \sim \Gamma(a,\theta)$ implies that $\mathbb{E}(z_i X_i) = z_i \mathbb{E} X_i = a/\theta$ and $Var(z_i X_i) = a/\theta^2$, and hence

$$\sum_i z_i X_i \sim N(na/\theta, na/\theta^2)$$

for large n. Thus,

$$\begin{aligned} \alpha &= P(\sum_i z_i X_i < C_\alpha \mid H_0) = P([\sum_i z_i X_i - na]/\sqrt{na} < [C_\alpha - na]/\sqrt{na} \mid H_0) \\ &\approx \Phi([C_\alpha - na]/\sqrt{na}) = 1 - \Phi([na - C_\alpha]/\sqrt{na}) \end{aligned}$$

which implies that $C_\alpha \approx na - z_\alpha \sqrt{na}$.
Thus, the exact UMP critical regions are

$$\{(x_1,\ldots,x_n) : \sum_i z_i x_i < F^{-1}_{\Gamma(an,1)}(\alpha)\}$$

and the approximate ones are

$$\{(x_1,\ldots,x_n) : \sum_i z_i x_i < na - z_\alpha \sqrt{na}\}.$$

(b) Since the critical regions are independent of θ_1, the preceding test is also UMP for testing $H_0 \colon \theta = 1$ against $H_1 \colon \theta > 1$.

(c) No, since the critical regions of the UMP for testing the simple hypotheses $H_0 \colon \theta = 1$ against the alternative hypothesis $H_1 \colon \theta_1$ for $\theta_1 \neq 1$ depend on θ_1. For $\theta_1 > 1$, the best critical regions are of the form $\{\sum_i z_i x_i < C_\alpha\}$, and for $\theta_1 \in (0,1)$ the best critical regions are of the form $\{\sum_i z_i x_i > C_\alpha\}$, that is, their form is different for different θ_1.

(d) For observed data with $n = 311$, $\sum_i z_i x_i = 571$ and $a = 2$, the 5% exact best critical regions are

$$\{(x_1,\ldots,x_n) : \sum_i z_i x_i < F^{-1}_{\Gamma(622,1)}(0.05) = 581.5538\}$$

and the approximate ones are

$$\{(x_1,\ldots,x_n) : \sum_i z_i x_i < na - z_{0.05}\sqrt{na} = 580.9775\},$$

that is, the null hypothesis is rejected at 5% significance level.
For $\alpha = 0.01$, the best critical regions are

$$\{(x_1,\ldots,x_n) : \sum_i z_i x_i < F^{-1}_{\Gamma(622,1)}(0.01) = 565.4556\}$$

and the approximate ones are

$$\{(x_1,\ldots,x_n) : \sum_i z_i x_i < na - z_{0.01}\sqrt{na} = 563.9811\},$$

that is, the null hypothesis is not rejected at 1% significance level.
Here $\sum_i z_i x_i$ can be viewed as a test statistic, so the corresponding exact p-value is

$$P(\sum_i z_i X_i < \sum_i z_i x_i \mid H_0) = F_{\Gamma(622,1)}(\sum_i z_i x_i) = F_{\Gamma(622,1)}(571) = 0.0183,$$

and the approximate p-value is

$$P(\sum_i z_i X_i < \sum_i z_i x_i \mid H_0) \approx \Phi([\sum_i z_i x_i - an]/\sqrt{an}) = 0.0204.$$

Therefore, according to the exact p-value, the null hypothesis is rejected for $\alpha < 0.0183$ and not rejected otherwise. The data provides some evidence against the null hypothesis, but the evidence is not strong.

(e) The power of the test $H_0\colon \theta = 1$ against the alternative hypothesis $H_1\colon \theta = 3$ as a function of n, with $a = 2$, is

$$\eta(\theta_1) \;=\; P(\sum_i z_i X_i < F^{-1}_{\Gamma(2n,1)}(0.05) \mid H_1 : \theta = 3) = F_{\Gamma(2n,3)}(F^{-1}_{\Gamma(2n,1)}(0.05))$$

since under H_1, $\sum_i z_i X_i \sim \Gamma(an, 3)$.
The smallest n such that the power of the test is greater than 0.9, equals $n = 4$, which can be found numerically, by plotting the power as a function of n. The corresponding power is 0.908 (for $n = 3$, the power is 0.794).

(f) According to the Neyman-Pearson lemma, the most powerful test of size α for testing the simple null hypothesis $H_0 : \theta = \theta_0$ against the simple alternative hypothesis $H_1\colon \theta = \theta_1$ $(\theta_1 > \theta_0)$ has critical regions of the form

$$\begin{aligned} R_\alpha(\theta_0) &= \{\mathbf{y} : \frac{\prod_i f(x_i \mid \theta_1)}{\prod_i f(x_i \mid \theta = \theta_0)} > K_\alpha\} \\ &= \{\mathbf{y} : (\theta_1/\theta_0)^{na} e^{(\theta_0-\theta_1)\sum_i z_i x_i} > K_\alpha\} \\ &= \{\mathbf{y} : (\theta_0 - \theta_1)\sum_i z_i x_i > c_\alpha\} \\ &= \{\mathbf{y} : \sum_i z_i x_i < C_\alpha\} \end{aligned}$$

since $(\theta_0 - \theta_1) < 0$. Using $\theta_0 \sum_i z_i X_i \sim \Gamma(an, 1)$ under the null hypothesis, C_α is given by

$$\alpha = P(\theta_0 \sum_i z_i X_i < \theta_0 C_\alpha \mid H_0) = F_{\Gamma(an,1)}(\theta_0 C_\alpha)$$

that is, $C_\alpha = \theta_0^{-1} F^{-1}_{\Gamma(an,1)}(\alpha)$. For the data given in (d) and $\alpha = 0.1$, $C_\alpha = \theta_0^{-1} F^{-1}_{\Gamma(622,1)}(0.1) = 590.26/\theta_0$.
Therefore, $R_\alpha(\theta_0) = \{\mathbf{y} : \sum_i z_i x_i < 590.26/\theta_0\}$.

By definition, a one-sided 90% confidence interval for θ using the critical regions $R_\alpha(\theta_0)$ is given by

$$\begin{aligned}\{\theta_0 : \mathbf{y} \notin R_\alpha(\theta_0)\} &= \{\theta_0 : \sum_i z_i x_i > 590.26/\theta_0\} = \{\theta_0 : 571 > 590.26/\theta_0\} \\ &= \{\theta_0 : \theta_0 > 590.26/571 = 1.03373\},\end{aligned}$$

that is, the corresponding 90% confidence interval for θ is $(1.0337, \infty)$.

19. (a) Using the Neyman-Pearson Lemma, the most powerful test of the simple null hypothesis $H_0 \colon \lambda = \lambda_0$ against the simple alternative hypothesis $H_1 \colon \lambda = \lambda_1$ $(\lambda_1 > \lambda_0)$ has critical region given by $\frac{L(\lambda_1)}{L(\lambda_0)} \geq A$ where A is a constant.

For a Poisson random sample the likelihood is $L(\lambda) = \text{constant}\cdot\lambda^{\sum x_i}\exp(-n\lambda)$, so the critical region is given by

$$\frac{L(\lambda_1)}{L(\lambda_0)} = \left(\frac{\lambda_1}{\lambda_0}\right)^{\sum x_i} \exp\{-n(\lambda_1 - \lambda_0)\} \geq A, \quad \text{or as } \lambda_1 > \lambda_0, \ \sum y_i \geq B,$$

where B is a constant.

As this is the *same* critical region for *any* $\lambda_1 > \lambda_0$, this is the critical region of a uniformly most powerful (UMP) test of the simple null hypothesis $H_0 \colon \lambda = \lambda_0$ against the composite alternative hypothesis $H_1 \colon \lambda > \lambda_0$.

(b) The MGF of $X_i \sim \text{Pois}(\lambda)$ is $M_X(t) = \mathbb{E}(e^{tX}) = \sum_{x=0}^{\infty} e^{tx}\frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{(e^t\lambda)^x}{x!} = \exp(\lambda(e^t - 1))$. Hence the MGF of $\sum X_i$ is $M_{\sum X_i}(t) = \prod M_{X_i}(t) = \exp(n\lambda(e^t - 1))$ which is the MGF of a Poisson random variable with parameter $n\lambda$.

A test with nominal level of 5% when $n = 10$ and $\lambda_0 = 1$ has critical region $\sum x_i \geq 16$ from tables of Poisson probabilities with $\mu = n\lambda_0 = 10$ ($\alpha = P(\sum X_i \geq 16) = 1 - P(\sum X_i \leq 15) = 1 - 0.9513 = 0.0487$).

An approximate critical value may be obtained using a normal approximation to the distribution of $\sum X_i \sim N(n\lambda, n\lambda)$. The critical region is given by

$$\sum x_i \geq n\lambda_0 + z_{0.05}\sqrt{n\lambda_0} + \frac{1}{2} = 10 + 1.6449\sqrt{10} + \frac{1}{2} = 15.7.$$

The addition of the /2 here is called a *continuity correction.* This is to account for the fact that we are approximating a discrete valued random variable by a continuous distribution.

(c) As $\lambda = 2$, $n\lambda = 20$, so power is $P(\sum_{i=1}^n X_i \geq 16) = 1 - \sum_{k=0}^{15}\frac{(20)^k e^{-20}}{k!} = 1 - 0.1565 = 0.8435$.

(d) We now require a test of $H_0 : \lambda = \lambda_0$ against the alternative $H_1 : \lambda \neq \lambda_0$. No uniformly most powerful test exists as for $\lambda_1 > \lambda_0$ the critical region is $\sum X_i \geq B$ but for $\lambda_1 < \lambda_0$ the critical region is $\sum X_i \leq B^*$, and critical regions are *not* of same form for all λ under alternative hypothesis.

Using a normal approximation to the distribution of $\sum X_i$ when $n = 10$ and $\lambda_0 = 1$, a two-sided test (not UMP though) would have critical values $n\lambda_0 \pm z_{\frac{0.05}{2}}\sqrt{n\lambda_0} \pm \frac{1}{2} = 10 \pm 1.96\sqrt{10} \pm \frac{1}{2} = 3.3$ and 16.7. Note that the additional term of 1/2 is included as a *continuity correction.* This is a standard approach when approximating a discrete random variable using a Normal distribution, which is continuous.

Appendix: Stick breaking

Here we provide proofs of the results that were used in questions 3(d) and 12, relating to the lengths of sticks broken at random.
Firstly we prove that the probability that the minimum length of pieces of a stick, of length L, broken at random into $n+1$ pieces exceeds x is

$$p_n = p(\min\{S_i :\ i = 1, \ldots, n+1\} > x) = \left(1 - (n+1)\frac{x}{L}\right)_+^n$$

Note that we can without loss of generality assume $L = 1$ by rescaling. The result for a stick of length L is found by the replacement $x \to x/L$ in the result for a stick of length 1. We prove this result inductively. For $n = 1$, the stick pieces both exceed length if the point of the break lies in the interval $[x, 1-x]$. There are no points in this interval if $1 - x < x$, i.e., $2x > 1$. Otherwise this interval is a fraction $1 - 2x$ of the total range in which the point could lie. We deduce that $p_1 = (1-2x)_+$, so the result holds for $n = 1$. Now suppose the result holds for some $n = k$ and consider $n = k+1$. The probability that the first break point lies in the interval $[u, u + \mathrm{d}u]$ is

$$(k+1)\mathrm{d}u(1-u)^k$$

which is the number of ways that the first break point can be chosen from the set of $k+1$ break points, times the probability density for that point (which is uniform), times the probability that the other k points all lie in the interval $[u, 1]$. All stick piece lengths exceed x if and only if the first break point on the stick lies beyond x, and all the remaining pieces have length that exceeds x. The latter probability is just the probability that a stick of length $(1-u)$ broken into $k+1$ pieces has no piece smaller than x, which follows from the induction assumption and is equal to $(1-(k+1)x/(1-u))_+^k$. We finally prove the induction step by integrating over u

$$\begin{aligned} p_{k+1} &= (k+1)\int_x^1 (1-u)^k \left(1 - (k+1)\frac{x}{(1-u)}\right)_+^k \mathrm{d}u \\ &= \int_x^{1-(k+1)x} (k+1)(1-u-(k+1)x)^k \mathrm{d}u = (1-(k+2)x)_+^{k+1} \end{aligned} \tag{8}$$

and so the result for $n = k+1$ follows.
Next we prove the result needed in question 3(d), namely that all of the interior intervals exceed x. This is related to the previous result, but is slightly different since we do not care about the first and last intervals, as these do not correspond to event separations, but only to separations with respect to the arbitrary start and end times of the observation interval. We derive the necessary result as follows. The probability that the first point is in the interval $[u, u + \mathrm{d}u]$ and the last point is in the interval $[v, v + \mathrm{d}v]$ is

$$n(n-1)\mathrm{d}u\mathrm{d}v(v-u)^{n-2}$$

which is the number of ways to specify the first and last points, times the probability density for those points, times the probability that all other points lie in the interval $[u, v]$. Given the first and last points lie at u and v, the probability that all internal intervals exceed x is just the probability that all pieces of a stick of length $(v-u)$, broken randomly into $n-1$ pieces, exceed x, which follows form the previous result. The final result follows be integrating over u and v

$$
\begin{aligned}
p_n &= \int_0^1 \int_u^1 n(n-1)\left(1-(n-1)\frac{x}{(v-u)}\right)_+^{n-2}(v-u)^{n-2}\mathrm{d}v\mathrm{d}u \\
&= \int_0^1 \int_{u+(n-1)x}^1 n(n-1)\left(v-u-(n-1)x\right)^{n-2}\mathrm{d}v\mathrm{d}u \\
&= \int_0^{1-(n-1)x} n\left(1-u-(n-1)x\right)^{n-1}\mathrm{d}u \\
&= \left(1-(n-1)x\right)_+^n . \qquad (9)
\end{aligned}
$$

This is the result required for Q3(d), setting $x=1$ and $L=t$, or equivalently $x=1/t$ in the above.
This same result is all that is required to answer Q12(b), but for Q12(a) we need the distribution of the maximum piece length. We first prove the result that the probability that the first r pieces of a stick broken into $n+1$ parts all exceed length x is

$$
(1-rx)_+^n ,
$$

which can also be used to prove the result above, as described in the solution to Q3(d). We again prove this by induction on n. Firstly we show that it is true for $n=1$. In that case the stick has 2 parts so we can have $r=1$ or $r=2$ (the result for $r=0$, which has probability 1, is trivial). For $r=1$, the probability is just the probability that the break point is in the interval $[x,1]$, which is $(1-x)$. For $r=2$, the probability is the probability that the break point is in the interval $[x,1-x]$, which is $(1-2x)_+$, so the result for $n=1$ follows. Now we suppose this holds for $n=k$ and we consider $n=k+1$. The probability that the first break point is in the interval $[u,u+\mathrm{d}u]$ is

$$
(k+1)\mathrm{d}u(1-u)^k
$$

as above. The first r intervals will all be greater than x if this first break point is in the range $[x,1]$, and the pieces defined by the next $r-1$ points are all greater than x. The latter is the probability that a stick of length $(1-u)$ broken into k pieces has the first $r-1$ pieces all longer than x, which is give by the induction assumption as $(1-(r-1)x/(1-u))_+^k$. We obtain the final result by integrating over u

$$
\begin{aligned}
p_{k+1,r} &= \int_x^1 (k+1)(1-u)^k\left(1-(r-1)\frac{x}{(1-u)}\right)_+^k \mathrm{d}u \\
&= \int_x^{1-(r-1)x} (k+1)(1-u-(r-1)x)^k\mathrm{d}u = (1-rx)_+^{k+1} \qquad (10)
\end{aligned}
$$

and so the result follows for $n=k+1$.
This result that we want to compute to answer Q12(a) is the probability that the maximum piece length is less than x. The statement that the r'th stick piece is shorter than x is the complement of the statement that the r'th stick piece is longer than x. Denoting by X_r the event that the r'th stick piece is longer than x, the probability we want to compute is

$$
\mathbb{P}\left(\bar{X}_1 \cap \bar{X}_2 \cap \bar{X}_3 \cap \ldots \cap \bar{X}_n \cap \bar{X}_{n+1}\right)
$$

where an overbar denotes the complement. If we consider two events then it is easy to see (from a Venn diagram or otherwise) that

$$
\mathbb{P}\left(\bar{A} \cap \bar{B}\right) = 1 - \mathbb{P}\left(A\right) - \mathbb{P}\left(B\right) + \mathbb{P}\left(A \cap B\right).
$$

For three events we have

$$\mathbb{P}\left(\bar{A}\cap\bar{B}\cap\bar{C}\right) = 1-\mathbb{P}(A)-\mathbb{P}(B)-\mathbb{P}(C)+\mathbb{P}(A\cap B)+\mathbb{P}(A\cap C)+\mathbb{P}(B\cap C)-\mathbb{P}(A\cap B\cap C)$$

and so on. Therefore the probability we require is

$$\begin{aligned}\mathbb{P}\left(\bar{X_1}\cap\ldots\cap\bar{X_{n+1}}\right) = 1 - \mathbb{P}(X_1) - \cdots - \mathbb{P}(X_{n+1}) + \mathbb{P}(X_1\cap X_2) + \cdots \mathbb{P}(X_n\cap X_{n+1}) - \cdots \\ \cdots + (-1)^{n+1}\mathbb{P}(X_1\cap\ldots\cap X_{n+1}).\end{aligned} \tag{11}$$

Since the breaks are distributed randomly, the probabilities do no depend on the labels of the intervals and so in each group of terms the probabilities are equal and are given by the previous result. We conclude that

$$\mathbb{P}\left(\bar{X_1}\cap\ldots\cap\bar{X_{n+1}}\right) = \sum_{j=0}^{n+1}(-1)^j \begin{pmatrix} n+1 \\ j \end{pmatrix} (1-jx)_+^n . \tag{12}$$

The result required for Q12(a) requires the replacement $n \to n-1$ since the periodic boundary condition means that the stick is broken into n pieces, with $n-1$ break points.