AA12M Statistical Astronomy (STA) problem sheet #2 (with solutions)

This problem sheet covers Bayesian and frequentist methods for parameter estimation and hypothesis testing. Bayesian questions are tagged with a little 'B', frequentist questions with an 'F'. Answers to some problems are shown in curly brackets when appropriate. Solutions are shown immediately after their questions (which are in smaller type), with a horizontal line separating each problem.

1)B Explain how Bayes' Theorem is used for parameter estimation in Bayesian Probability Theory.

A telescope is constructed to look for gamma ray bursters (GRBs). Given that GRBs can appear from any direction (i.e., are distributed isotropically on the sky) what is the prior probability distribution that a GRB will be seen at a particular declination δ (hint: consider the fraction of the sky at this declination)? If $\mu = \sin\delta$, show that the prior for μ is uniform for $-1 < \mu < 1$.

Solution to (1) For the first section, see your lecture notes. In Bayesian probability theory probability represents a degree of belief in a proposition. Bayes' theorem tells us how new data modulates our prior beliefs.

The GRBs are isotropic, so the pdf should be invariant under rotation, i.e., there should be no preferred directions in the sky. If $p(\Omega)\,\mathrm{d}\Omega$ is the pdf for finding a GRB in the differential solid angle $\mathrm{d}\Omega$ in direction Ω then isotropy means that $p(\Omega)\,\mathrm{d}\Omega =$ constant. Correct normalisation gives $p(\Omega) = 1/(4\pi)$.

Now let's change the variables to declination, δ, and right ascension, ϕ. We know from spherical polar coordinates that $\mathrm{d}\Omega = \cos\delta\,\mathrm{d}\delta\,\mathrm{d}\phi$. By standard change of variable

$$p(\delta,\phi)|\,\mathrm{d}\delta\,\mathrm{d}\phi| = p(\Omega)|\,\mathrm{d}\Omega|,$$

so

$$p(\delta,\phi) = \frac{\cos\delta}{4\pi}.$$

We can marginalise over ϕ easily, as the pdf is independent of ϕ, so that

$$p(\delta) = \int_0^{2\pi} p(\delta,\phi)\,\mathrm{d}\phi = \frac{1}{2}\cos\delta.$$

As instructed, now change the variable to $\mu = \sin\delta$. Note that $|\,\mathrm{d}\mu| = |\cos\delta\,\mathrm{d}\delta|$. So now

$$p(\mu) = p(\delta)\left|\frac{\mathrm{d}\delta}{\mathrm{d}\mu}\right| = \frac{\cos\delta}{2\cos\delta} = \frac{1}{2}.$$

So the prior for μ is uniform (with the value 0.5) in the range $-1 < \mu < 1$.

---

2)B Briefly, what is *marginalisation* in the context of Bayesian inference?

In a photon counting experiment we are told that, over a time interval T, exactly two photons have struck our detector. Given no other information, write down and justify the joint probability distribution function (pdf) for the two arrival times t_1 and t_2.

By sketching this pdf, or otherwise, show that the probability that they arrived within a time τ of each other is

$$P(\tau) = 1 - (1 - \tau/T)^2.$$

Given the extra information that they did indeed arrive within a time τ of each other, sketch the pdfs for the time:

(a) of the first arrival

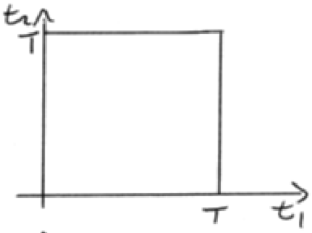
(b) of the second arrival

(c) that a photon arrives.

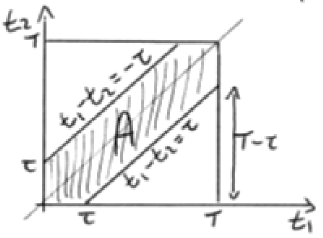
Solution to (2)

Marginalisation : If $p(x,y)$ is a joint Bayesian pdf then $p(x) = \int p(x,y)\,dy$ is the marginal pdf for x [2]

No information on arrival times $\Rightarrow$ UNIFORM joint pdf,



$$p(t_1,t_2) = \frac{1}{T^2} \quad (0 \leq t_1 \leq T;\ 0 \leq t_2 \leq T)$$
$$0 \quad \text{otherwise}$$

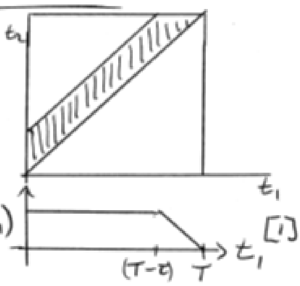


Area A corresponds to $|t_1-t_2| \leq \tau$

The excluded region has area

$$T^2 - A = 2 \cdot \frac{1}{2}(T-\tau).(T-\tau)$$

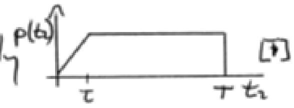
so $A = T^2 - (T-\tau)^2$ [3]



a) Lets say photon 1 arrive 1st $\Rightarrow t_1 < t_2$

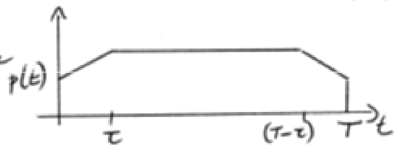
Marginalising shaded region on to t_1 axis, [1]

b) Similarly



[1]

c) Marginalising onto t_1 without the constraint that $t_1 < t_2$:



[1]

---

3)B Use your favorite computing resource (Python, Matlab, Wolfram Alpha, ...) to determine the fraction of the probability contained within a ($\pm$)1, 2 and 3 σ zone around the mean of a Gaussian posterior probability distribution.

Solution to (3) All these are integrals of the form

$$I_a = \int_{-a\sigma}^{a\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/(2\sigma^2)}\, dx$$

for $a = 1$, 2 and 3. Putting $y = x/\sigma$ gives

$$I_a = \int_{-a}^{a} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy.$$

[I_a is related to the ‘error function’, erf(z), by $I_a = \text{erf}(a/\sqrt{2})$]. The results (from your computing resource!) are:

$I_1 = 0.6827$ corresponding the the ‘68%’ region of probability,
$I_2 = 0.9545$ corresponding the the ‘95%’ region of probability,
$I_3 = 0.9973$ corresponding the the ‘99.7%’ region of probability.

4)B From what you know, write down sensible prior probabilities/pdfs for the following quantities:

- The number of sweets in a jar.
- The mass of Neptune.
- The radius of Neptune.

Solution to (4) A deliberately vague question to make you think about the meaning and subjective nature of priors. There is no single correct answer, as a prior encodes one's current state of knowledge, and different people are in different states! Two people with identical states of knowledge should, however, assign identical prior probabilities. Priors may be subjective, but they are also deterministic.

- The number of sweets in a jar: the number of sweets in sensible jars can be as high as several thousand, but we have no information other than our experience of sweets and jars. Let us say that the maximum number that is possible is L, then if we have no reason to prefer n sweets over m sweets all numbers less than or equal to L are equally probable (including zero), so that
$$p(n) = \frac{1}{L+1}, \quad (0 \le n \le L).$$
You may think that a sharp cutoff for L is not justified, so that very big jars (or very small sweets) are just unlikely rather than impossible. In this case you would modify the pdf to roll-off at high n rather than stop abruptly. Note the normalisation would have to be re-evaluated. Again, this is fine so long as you state your prior precisely.

- The mass of Neptune: you may know this off the top of your head, in which case you may be able to quote a rather tight prior around that numerical value with some spread describing your uncertainty in the value. Alternatively you may have no idea other than it's somewhere between the mass of the Earth and the mass of the Sun. If you really think every mass between the two is equally likely (corresponding to adding another atom onto the pile) then we are back to the sweet problem, and the pdf is uniform between its limits.

- The radius of Neptune: again you face a choice of reasonable priors. However there is more than one way of looking at the problem. If you believe that the density of planets is constant (you would be wrong to believe this, but it may be your state of belief at the start!) then the mass and radius of the planet are simply related as $m \propto r^3$. By change of variable,
$$p(r) \propto p(m)\left|\frac{\mathrm{d}m}{\mathrm{d}r}\right| \propto p(m)r^2.$$
In the previous problem we took $p(m)$ as uniform within some range, so to be consistent $p(r)$ must be $\propto r^2$ within the same corresponding range in r. Sometimes probabilities can be more easily assigned by thinking of an associated variable, assigning a probability to that quantity using a physical argument, and then changing variable to the one you want. Sometimes two equally plausible assignments are inconsistent, in which case you need to think harder.

---

5)F The Central Limit Theorem states that the sum of N random variables drawn from almost any distribution will itself be a random variable which, if N is sufficiently large, will be Normally distributed. Extend this idea to the *product* of N random variables(all > 0), and determine the distribution of this product (called the Lognormal distribution).

Why might the masses of bodies in the rings of Saturn have a Lognormal distribution?

Solution to (5) The Central Limit Theorem tells us that if $X = x_1 + x_2 + x_3 + ... + x_N$ and if x_i are all random variables drawn from the same parent population with a well defined mean and variance then, for sufficiently large N, the pdf of X becomes Normal (i.e., it has a Central distribution). So let $Y = x_1 \cdot x_2 \cdot ... \cdot x_N \quad (x_i > 0)$, then $\ln Y = \ln x_1 + \ln x_2 + ... + \ln x_N$ and $Z = \ln Y = z_1 + z_2 + ... + z_N$, where z_i are random variables. We can apply the Central Limit Theorem to Z. Its pdf will be Normal with some mean μ and variance σ^2 so that

$$p(Z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(Z-\mu)^2}{2\sigma^2}\right].$$

If we now change variable to $Y = \mathrm{e}^Z$ then $\mathrm{d}Z/\mathrm{d}Y = 1/Y$ so that

$$p(Y) = p(Z)\left|\frac{\mathrm{d}Z}{\mathrm{d}Y}\right| = \frac{p(Z)}{Y} = \frac{1}{Y}\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(\ln Y-\mu)^2}{2\sigma^2}\right],$$

with $Y > 0$.

This is the Lognormal distribution. If we think of large rocks in Saturn's rings undergoing collisions, then we may expect the rocks to split into two on each collision with some pdf for the relative sizes of the two fragments. If x above represents a particular splitting fraction, then Y represents the size of a fragment after N collisions, relative to its initial size. If all the rocks began the same size and underwent the same number of collisions then the fragments we see today would have a Lognormal distribution. Rather a lot of 'ifs' there, and in fact the distribution of rock sizes in Saturn's rings is not Lognormal as far as we know, but the Lognormal distribution is rather common in nature.

---

6)B Outline the reasoning behind 'least-squares' parameter estimation within a Bayesian framework.
For a set of data, $\{Y_k\}$, with associated error bars $\{\sigma_k\}$, taken at known 'positions' $\{x_k\}$ derive the best slope (m_0) and intercept (c_0) for a straight line fit.

Solution to (6) See your lecture notes for the first part. It is a result of seeking the most likely value of the posterior when the prior is uniform and the likelihood is Gaussian.

The analysis for m_0 and c_0 is done in Sivia, page 69: let the straight line we want to fit have value y_k at position x_k so that

$$y_k = mx_k + c.$$

Our job is to find the best values of slope and intercept. Least-squares minimises the quantity

$$\chi^2 = \sum_{k=1}^{N} \frac{(mx_k + c - Y_k)^2}{\sigma_k^2}.$$

We do the minimisation by considering

$$\frac{\partial \chi^2}{\partial m} = \sum_{k=1}^{N} \frac{2(mx_k + c - Y_k)x_k}{\sigma_k^2} \quad \text{and} \quad \frac{\partial \chi^2}{\partial c} = \sum_{k=1}^{N} \frac{2(mx_k + c - Y_k)}{\sigma_k^2}.$$

By setting both these partial derivatives to zero at $m = m_0$ and $c = c_0$, we get two simultaneous equations in m_0 and c_0 that can be written as

$$\begin{pmatrix} \alpha & \gamma \\ \gamma & \beta \end{pmatrix} \begin{pmatrix} m_0 \\ c_0 \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix},$$

where $\alpha = \sum 2x_k^2/\sigma_k^2$, $\beta = \sum 2/\sigma_k^2$, $\gamma = \sum 2x_k/\sigma_k^2$, $a = \sum 2x_kY_k/\sigma_k^2$ and $b = \sum 2Y_k/\sigma_k^2$. By matrix inversion the solutions are

$$m_0 = \frac{\beta a - \gamma b}{\alpha\beta - \gamma^2} \quad \text{and} \quad c_0 = \frac{\alpha b - \gamma a}{\alpha\beta - \gamma^2}.$$

---

7)B N observations of the flux density of a quasar, $\{x_k\}$, are affected by interstellar scintillation which introduces Gaussian errors of (unknown) variance σ^2. Explain what is meant by the *likelihood* of these data, and show that, if the measurements are independent, the likelihood is

$$p(\{x_k\}|\mu,\sigma,I) = (\sigma\sqrt{2\pi})^{-N}\exp\left[-\frac{1}{2\sigma^2}\sum_{k=1}^{N}(x_k-\mu)^2\right],$$

where μ is the true flux density of the quasar.

Explain the importance of the joint posterior pdf of μ and σ for parameter estimation. What is the meaning of the *marginal* posterior pdf for μ alone? Show that, if the priors for μ and σ are uniform for values > 0 and zero otherwise, the marginal posterior pdf for μ is

$$p(\mu|\{x_k\},I) \propto \int_0^\infty t^{N-2}\exp\left[-\frac{t^2}{2}\sum_{k=1}^{N}(x_k-\mu)^2\right]\mathrm{d}t,$$

where $t = 1/\sigma$. Evaluate the un-normalised value of this,* given the standard result

$$\int_0^\infty x^n\exp(-ax^2)\,\mathrm{d}x \propto a^{-(n+1)/2}.$$

By examining the maximum of $L = \ln\left[p(\mu|\{x_k\},I)\right]$, show that the best estimate for μ is

$$\mu_0 = \frac{1}{N}\sum_{k=1}^{N}x_k,$$

and that the uncertainly in this is $S/\sqrt{N}$ where

$$S^2 = \frac{1}{N-1}\sum_{k=1}^{N}(x_k-\mu_0)^2.$$

Comment on how this result compares to the situation where σ is known, as derived in the notes.

Solution to (7) This analysis is done in Sivia p. 55. The joint likelihood of independent data is just the joint probability of the data, given all other parameters. For each measurement the likelihood is

$$p(x|\mu,\sigma,I) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

The joint pdf is the product of these for each x_k, hence

$$p(\{x_k\}|\mu,\sigma,I) = (\sigma\sqrt{2\pi})^{-N}\exp\left[-\frac{1}{2\sigma^2}\sum_{k=1}^{N}(x_k-\mu)^2\right].$$

By Bayes' Theorem

$$p(\mu,\sigma|\{x_k\},I) \propto p(\mu,\sigma|I)p(\{x_k\}|\mu,\sigma,I).$$

We are told that the joint prior for μ and σ, $p(\mu,\sigma|I)$, is uniform for positive values, zero otherwise. Using this and marginalising over σ we get

$$p(\mu|\{x_k\},I) \propto \int_0^\infty \sigma^{-N}\exp\left[-\frac{1}{2\sigma^2}\sum_{k=1}^{N}(x_k-\mu)^2\right]\mathrm{d}\sigma.$$

Now putting $t = 1/\sigma$ and $\mathrm{d}\sigma = -\mathrm{d}t/t^2$ we get

$$p(\mu|\{x_k\},I) \propto \int_0^\infty t^{N-2}\exp\left[-\frac{t^2}{2}\sum_{k=1}^{N}(x_k-\mu)^2\right]\mathrm{d}t,$$

*The answer you get is basically Student's t distribution, derived from Bayesian principles.

as required. The standard integral given in the question lets us express this as

$$p(\mu|\{x_k\}, I) \propto \left[\frac{1}{2}\sum_{k=1}^{N}(x_k-\mu)^2\right]^{-(N-1)/2} .$$

Now set

$$L = \ln\left[p(\mu|\{x_k\}, I)\right] = -\frac{N-1}{2}\ln\left[\sum_{k=1}^{N}(x_k-\mu)^2\right] + c,$$

$$\frac{\mathrm{d}L}{\mathrm{d}\mu} = \frac{(N-1)\sum_{k=1}^{N}(x_k-\mu)}{\sum_{k=1}^{N}(x_k-\mu)^2} = 0 \quad \text{at } \mu = \mu_0\ ,$$

where μ_0 is the most probable value of μ. The solution to this is

$$\mu_0 = \frac{1}{N}\sum_{k=1}^{N} x_k$$

as stated.

For the variance of this result we consider

$$\frac{\mathrm{d}^2L}{\mathrm{d}\mu^2} = \frac{-N(N-1)}{\sum_{k=1}^{N}(x_k-\mu)^2},$$

and use the result

$$\sigma_\mu^2 = \left(-\left.\frac{\mathrm{d}^2L}{\mathrm{d}\mu^2}\right|_{\mu_0}\right)^{-1} = \frac{1}{N(N-1)}\sum_{k=1}^{N}(x_k-\mu_0)^2.$$

Hence $\sigma_\mu = S/\sqrt{N}$ where

$$S^2 = \frac{1}{N-1}\sum_{k=1}^{N}(x_k-\mu_0)^2.$$

In the lectures we considered the same problem but where the variance of the original observations, σ^2, was known. In these circumstances $\sigma_\mu = \sigma/\sqrt{N}$. So the above analysis has naturally deduced that the right thing to do is estimate the variance from the spread in the data and use S in place of σ. In frequentist statistics S is the *unbiased estimate* of σ.

---

8)B A spacecraft is sent to a moon of Saturn, and, using a penetrating probe, detects a liquid sea under the surface at 1 atmosphere pressure and a temperature of $-3\,^\circ$C. However the thermometer has a random fault, so that the temperature reading may differ from the true temperature by as much as $\pm 5\,^\circ$C with uniform probability within that range.

(a) Assuming the liquid is water (background information I_1), which is a liquid for $0\,^\circ\mathrm{C} < T < 100\,^\circ\mathrm{C}$,

(i) write down, and draw, a sensible prior for the temperature of the liquid, $p(T|I_1)$.

(ii) write down, the likelihood of the data, given the instrument's troublesome performance and sketch its variation with T.

(iii) compute, and draw, the normalised posterior pdf for the temperature.

(b) Now repeat the above analysis assuming the liquid is ethanol (background information I_2) which is liquid at one atmosphere between $-80\,^\circ$C and $80\,^\circ$C, and comment on the difference in the results.

(c) The calibration error is found to be such that subsequent readings have independent errors within the range $\pm 5\,^\circ$C.

(i) By applying the central limit theorem determine, and sketch, the likelihood of the average of 100 such readings.

(ii) Given that this average reading is $-1.2\,^\circ$C again calculate, and sketch, the posterior pdfs under the hypotheses that the liquid is water or ethanol.

(d) (advanced) determine the relative odds that the liquid is water or ethanol following the single measurement and the average measurement.

Solution to (8) This problem is loosely based on an example by John Skilling. John has done much to promote the cause of Bayesian analysis and, in particular, Maximum Entropy methods.

(a) First water, I_1,

(i) Here $p(T|I_1) = 0.01$ for $0\,^\circ\text{C} < T < 100\,^\circ\text{C}$, zero otherwise. If it's water the temperature *must* be in this range. It has the value 0.01 to get the normalisation right.

(ii) Let the measurement be d. Then its likelihood, $p(d|T, I_1)$, is a top hat, centred on T, of width $10\,^\circ$C and height 1/10. So $p(d|T, I_1) = 0.1$ for $|d - T| \leq 5$, zero otherwise. Here, $d = -3$, so the likelihood is 0.1 for $|3 + T| \leq 5$, zero otherwise.

(iii) The posterior is the normalised product of the prior and the likelihood. The prior only allows temperatures $\geq 0\,^\circ$C. The likelihood only allows temperatures between $-5\,^\circ$C and $+2\,^\circ$C. The posterior is therefore $p(T|d, I_1) = 0.5$ for $0 \leq T \leq 2\,^\circ$C, zero otherwise. Note we can also evaluate the *evidence*, $p(d|I_1)$. This is a measure of how good the original question was. The evidence here is $\int p(d|T, I_1)p(T|I_1)\,\mathrm{d}T = 0.002$. We can assess the significance of this later.

(b) Now for ethanol, I_2, the prior is $p(T|I_2) = 1/160$ for $-80\,^\circ\text{C} < T < 80\,^\circ\text{C}$, zero otherwise. The likelihood is the same as for I_1, and the posterior is now $p(T|d, I_2) = 0.1$ for $-8 \leq T \leq 2\,^\circ$C, zero otherwise. Again we can evaluate the evidence as $\int p(d|T, I_2)p(T|I_2)\,\mathrm{d}T = 0.006\,25$. This is a factor ~ 3 greater than for water, so all things being equal ethanol is a better bet.

(c) To apply the CLT we need the mean and variance of the individual errors that we combine when we take the mean of 100 samples. The question states that the pdf of the errors is uniform with a width of $10\,^\circ$C. The mean error is clearly zero and the variance is

$$\sigma_1^2 = \int_{-5}^{5} \frac{x^2}{10}\,\mathrm{d}x = \frac{25}{3} = 8.33.$$

(i) After 100 measurements the variance of the mean of these should be (by the CLT) $\sigma_{100}^2 = \sigma_1^2/100$, so $\sigma_{100} = 0.289$. We can now write the likelihood of this averaged data, d_m, as a Gaussian (again, by the CLT) :

$$p(d_\mathrm{m}|T) = \frac{1}{\sqrt{2\pi}\sigma_{100}} \exp\left[-\frac{(d_\mathrm{m} - T)^2}{2\sigma_{100}^2}\right].$$

(ii) If $d_\mathrm{m} = -1.2\,^\circ$C then for both I_1 and I_2

$$p(d_\mathrm{m} = -1.2\,^\circ\mathrm{C}|T) \propto \exp\left[-\frac{1}{2}\left(\frac{-1.2 - T}{0.289}\right)^2\right].$$

This is strongly peaked around $T = -1.2\,^\circ$C, so the water hypothesis (I_1) will struggle as the likelihood and prior fight each other. We can of course still determine the posterior distribution for T under the water hypothesis. It is

$$p(T|d_\mathrm{m}, I_1) = \frac{\exp\left[-\frac{1}{2}\left(\frac{-1.2-T}{0.289}\right)^2\right]}{0.000\,011\,9}$$

for $0 \le T \le 100\,^\circ$C, zero otherwise. The normalising constant was computed in Maple, but you could use tables. The most probable temperature for the water hypothesis is $0\,^\circ$C – the best trade-off between prior prejudice and the data!

For ethanol the corresponding expression is

$$p(T|d_{\mathrm{m}}, I_2) = \frac{\exp\left[-\frac{1}{2}\left(\frac{-1.2-T}{0.289}\right)^2\right]}{0.724}$$

for $-80 \le T \le 80\,^\circ$C, zero otherwise. The most probable temperature is $-1.2\,^\circ$C for this hypothesis, dominated by the data.

(d) Taking the relative prior odds of I_1 and I_2 to be one (i.e., assuming there is no reason to favour one over the other before the data are available), the posterior relative odds is simply the ratio of the two evidences as defined above. We have seen that for the single measurement this ratio is about 3:1 in favour of ethanol. Once 100 measurements have been averaged the odds favour ethanol very strongly. Using the same equations, and Maple, the odds ratio comes out to be about 38 000:1 in favour of ethanol! Both hypotheses deliver answers, but the ethanol hypothesis is *much* more convincing.

---

9)B An important topic in X-ray astronomy is the determination of the X-ray background rate, b (i.e., the rate of arrival of X-rays from the background sky).

An X-ray telescope observes a 'blank' area of sky and counts n X-ray photons in a time T. The likelihood of this observation follows the Poisson distribution,

$$p(n|b, I) = \frac{(bT)^n \mathrm{e}^{-bT}}{n!}.$$

Taking b to be a scale parameter, assign it a prior, $p(b|I)$, and determine the normalised posterior for b. You will need to use

$$\int_0^\infty x^m \mathrm{e}^{-ax}\,\mathrm{d}x = \frac{m!}{a^{m+1}} \qquad (a > 0;\ m = 0, 1, 2 \ldots).$$

Show that the mean of this posterior is n/T, and that its standard deviation is the mean divided by $\sqrt{n}$. Repeat this analysis using a uniform prior for b. Do the two results differ substantially?

Solution to (9) If b is a true scale parameter then $p(b|I) \propto 1/b$ by definition. So the posterior of b is

$$\begin{aligned} p(b|n, I) &\propto p(b|I)p(n|b, I) \\ &\propto \frac{1}{b}\frac{(bT)^n \mathrm{e}^{-bT}}{n!} \\ &= cb^{n-1}\mathrm{e}^{-bT} \end{aligned}$$

where c is the normalising constant. Normalising we get

$$\begin{aligned} 1 &= c\int_0^\infty b^{n-1}\mathrm{e}^{-bT}\,\mathrm{d}b \\ &= c\frac{(n-1)!}{T^n}, \end{aligned}$$

so finally

$$p(b|n, I) = \frac{T^n}{(n-1)!} b^{n-1}\mathrm{e}^{-bT}.$$

By definition of the mean

$$\langle b\rangle = \int_0^\infty b\,p(b)\,\mathrm{d}b = \frac{T^n}{(n-1)!}\int_0^\infty b^n \mathrm{e}^{-bT}\,\mathrm{d}b = \frac{n}{T}.$$

Similarly, the variance is

$$\sigma^2 = \langle b^2 \rangle - \langle b \rangle^2.$$

For which we need

$$\langle b^2 \rangle = \int_0^\infty b^2 p(b)\,\mathrm{d}b = \frac{T^n}{(n-1)!}\int_0^\infty b^{n+1}\mathrm{e}^{-bT}\,\mathrm{d}b = \frac{n(n+1)}{T^2},$$

so $\sigma^2 = n/T^2$ and $\sigma = \langle b \rangle/\sqrt{n}$.

Using a uniform prior gives $\langle b \rangle = (n+1)/T$ and $\sigma = \langle b \rangle/\sqrt{n+1}$. For $n \gg 1$ the two priors give indistinguishable results: the data overwhelms the prior ignorance.

---

10)F The fraction, X, of the surface of a star covered in starspots is modelled as a random variable with pdf (with k constant)

$$p(x) = \frac{k}{\sqrt{x(1-x)}}, \qquad 0 < x < 1$$

(a) Determine k so that $p(x)$ is properly normalised. $\{1/\pi\}$

(b) Find the expected fraction of the surface covered in starspots. $\{1/2\}$

(c) What is the probability that the fraction covered is less than 25%? $\{1/3\}$

Solution to (10) To normalise $p(x)$ we require that

$$\int_0^1 \frac{k}{\sqrt{x(1-x)}}\,\mathrm{d}x = 1.$$

The trick here is to use the substitution $x = \sin^2\theta$, so that $\mathrm{d}x = 2\sin\theta\cos\theta\,\mathrm{d}\theta$. The integral now becomes

$$1 = 2k\int_0^{\pi/2}\mathrm{d}\theta,$$

so $k = 1/\pi$.

The expectation value of x is just

$$\langle x \rangle = \int_0^1 x\,p(x)\,\mathrm{d}x,$$

and this can be integrated using the same substitution as above to give $\langle x \rangle = 1/2$.

The probability that a fraction t of the surface is covered by spots is

$$P(t) = \int_0^t p(x)\,\mathrm{d}x = \frac{2}{\pi}\sin^{-1} t^{1/2},$$

using the same substitution as before. When $t = 1/4$ this reduces to $P(x < 0.25) = 1/3$.

---

11)B The redshift of a quasar is measured to be z_1 with a standard deviation of σ.

(a) Assuming the uncertainty in this data is Gaussian, explain why a uniform prior probability distribution for z implies a normalised posterior probability distribution for z of

$$p(z|z_1) = \frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left[-\frac{(z-z_1)^2}{2\sigma^2}\right]$$

(b) A second measurement, z_2, also of standard deviation σ, is made with the same setup. Write down an expression for the likelihood of z based solely in this second measurement, $p(z_2|z)$, and hence show that the new posterior probability distribution is

$$p(z|z_1,z_2) \propto \frac{1}{2\pi\sigma^2}\exp\left[-\frac{(z-z_1)^2+(z-z_2)^2}{2\sigma^2}\right]$$

(c) Show that this can be written in the form

$$p(z|z_1, z_2) \propto \frac{1}{2\pi\sigma^2} \exp\left[-\frac{A(z-z_\mathrm{m})^2 + c}{2\sigma^2}\right]$$

and evaluate A, z_m, and c. Hence show that $p(z|z_1, z_2)$ is also Gaussian with a mean of $(z_1+z_2)/2$ and a standard deviation of $\sigma/\sqrt{2}$, commenting on these results.

(d) Usually, one might expect that the two measurements z_1 and z_2 would be within about σ of each other, but on this occasion $z_1 - z_2 = 10\sigma$. Sketch graphs of $p(z|z_1)$, $p(z|z_2)$ and $p(z|z_1, z_2)$ for this situation, and give an interpretation of what may be going on here. Why is the uncertainty in the final posterior so (relatively) small?

(e) Define the evidence (or 'marginal likelihood') for this dataset, and briefly describe the role it plays in Bayesian parameter estimation. Show that the evidence in this problem is proportional to $\exp\left[-c/(2\sigma^2)\right]$, and comment on how its value depends on the magnitude of $z_1 - z_2$.

Solution to (11)

(a) Bayes Theorem states that

$$p(z|z_1) = \frac{p(z)p(z_1|z)}{p(z_1)}.$$

If $p(z)$ is uniform, and the likelihood is Gaussian, then

$$p(z|z_1) \propto \exp\left(-\frac{(z-z_1)^2}{2\sigma^2}\right).$$

Normalising this correctly we get

$$p(z|z_1) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{(z-z_1)^2}{2\sigma^2}\right).$$

(b) The likelihood of z is

$$p(z_2|z) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{(z-z_2)^2}{2\sigma^2}\right).$$

By Bayes Theorem

$$p(z|z_1, z_2) = \frac{p(z|z_1)p(z_2|z, z_1)}{p(z_2|z_1)}) = \frac{1}{p(z_2|z_1)} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(z-z_1)^2 + (z-z_2)^2}{2\sigma^2}\right).$$

(c) Let

$$(z-z_1)^2 + (z-z_2)^2 = A(z-z_\mathrm{m})^2 + c$$
$$2z^2 - 2z(z_1+z_2) + z_1^2 + z_2^2 = Az^2 - 2Azz_\mathrm{m} + Az_\mathrm{m}^2 + c$$

so by matching coefficients we see that $A = 2$ and $z_\mathrm{m} = (z_1+z_2)/2$. Also, solving for c we get

$$c = \frac{1}{2}\left(z_1^2 + z_2^2\right) - z_1 z_2,$$

so

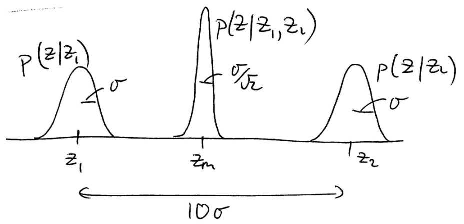
$$p(z|z_1, z_2) = \frac{1}{p(z_2|z_1)} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{A(z-z_\mathrm{m})^2 + c}{2\sigma^2}\right)$$

If we look at the terms that depend in z, we get

$$p(z|z_1, z_2) \propto \frac{1}{2\pi\sigma^2} \exp\left(-\frac{A(z-z_\mathrm{m})^2 + c}{2\sigma^2}\right).$$

This is a Gaussian, mean z_m standard deviation $\sigma/\sqrt{2}$, which is exactly to be expected. We have simply averaged two measurements, decreasing the uncertainty in the value of z by $\sqrt{2}$.

(d) The two measurements seem inconsistent with each other:



and the final posterior does not appear consistent with, or justified by, the two individual posteriors. The point is that we have *assumed* that our data model is correct (i.e., our value for σ, Gaussian errors etc). The procedure is interpreting z_1, z_2 as very improbable measurements consistent with $p(z|z_1, z_2)$ but right in the wings of that distribution.

(e) Bayes says

$$p(z|z_1, z_2, I) = \frac{p(z|I)p(z_1, z_2|z, I)}{p(z_1, z_2|I)}$$

The denominator here is the *evidence*

$$E = p(z_1, z_2|I) = \int p(z|I)p(z_1, z_2|z, I)\ \mathrm{d}z.$$

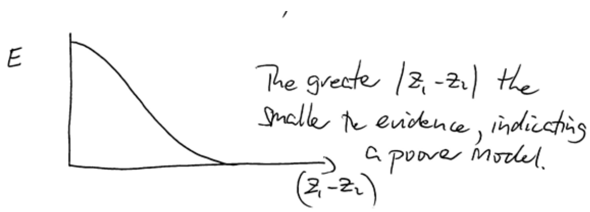
The evidence is the probability of the data for *any* model parameter value. It is a measure of the appropriateness of the model.

$$\begin{aligned}
E &= \int p(z|I)p(z_1, z_2|z, I)\ \mathrm{d}z \\
&= \int \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(z - z_\mathrm{m})^2 - c/2}{\sigma^2}\right)\ \mathrm{d}z \\
&\propto \exp\left(-\frac{c}{2\sigma^2}\right) \times \text{numerical factor depending only on } \sigma
\end{aligned}$$

so

$$\begin{aligned}
\ln E &= -\frac{c}{2\sigma^2} + \text{const} \\
&= \frac{1}{2\sigma^2}\left(z_1 z_2 - \frac{1}{2}(z_1^2 + z_2^2)\right) + \text{const} \\
&= -\frac{1}{4\sigma^2}(z_1 - z_2)^2 + \text{const.} \\
\frac{\partial \ln E}{\partial z_1} &= \frac{z_2 - z_1}{2\sigma^2} \\
\frac{\partial^2 \ln E}{\partial z_1^2} &= -\frac{1}{2\sigma^2}.
\end{aligned}$$

So E is maximised when $z_2 = z_1$, i.e., when both measurements agree.



12)F Let p denote the probability that a particular outcome will happen in any single experiment (called the probability of a *success*). The probability, $p(r)$, of exactly r successes in n experiments is given by the *binomial* distribution:

$$p(r) = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r}, \qquad r = 0, 1, \ldots, n$$

$p(r)$ has mean $\mu = np$ and variance $\sigma^2 = np(1-p)$.

Suppose that 60% of the stars in the Hubble Space Telescope guide star catalogue are binaries, use a binomial distribution model to calculate the probability that a random sample of 5 stars from the guide star catalogue contains a) 0, b) 1, c) 2, d) 3, e) 4, f) 5 binary stars.

How large a sample should be chosen in order that the probability of the sample containing at least two *non*-binary stars is greater than 99%? {14}

Solution to (12) We are told that $p = 0.6$, so by straight substitution

$$\begin{aligned}
\text{Prob(0 binaries)} &= (0.6)^0(0.4)^5 &&= 0.01024 \\
\text{Prob(1 binary)} &= 5(0.6)^1(0.4)^4 &&= 0.0.0768 \\
\text{Prob(2 binaries)} &= 10(0.6)^2(0.4)^3 &&= 0.2304 \\
\text{Prob(3 binaries)} &= 10(0.6)^3(0.4)^2 &&= 0.3456 \\
\text{Prob(4 binaries)} &= 5(0.6)^4(0.4)^1 &&= 0.2592 \\
\text{Prob(5 binaries)} &= (0.6)^5(0.4)^0 &&= 0.0778
\end{aligned}$$

If there are n experiments, then the probability that there are at least 2 non-binaries is 1 minus the probability that there are either n binaries or $n-1$ binaries (the only two situations that do not have at least 2 non-binaries). The number of stars needed to satisfy the condition that the probability is > 0.99 of there being 2 non-binaries is the solution to

$$(0.6)^n + n(0.6)^{n-1}(0.4) < (1 - 0.99).$$

Direct substitution shown that this is first satisfied for $n = 14$.

---

13)F In a meteor search program, four photographic plates were exposed on each observing night and examined for meteor trails. Over a one year period, 150 nights of data were accumulated with the following results:

No of plates with trails	0	1	2	3	4
No of nights	30	62	46	10	2

The number, r, of plates recording meteor trails on any given night is assumed to follow a binomial distribution. The minimum number is 0, the maximum is 4. By equating the sample mean value of r for the above observations with the expected value for a binomial distribution, estimate the parameter, p, the probability of a single plate recording a meteor trail. {0.32}

Hence determine the *predicted* number of nights on which r plates record trails under the binomial model ($r = 0, \ldots, 4$).

Solution to (13) Sample mean is

$$\hat{r} = \frac{0 \times 30 + 1 \times 62 + 2 \times 46 + 3 \times 10 + 4 \times 2}{150} = 1.28$$

The expectation value for a binomial distribution is np (see previous question), so setting $np = \hat{r}$, with $n = 4$ plates, we get $\hat{p} = 0.32$.

The predicted number of plates is just $N(r) = 150 \times p(r|\hat{p})$, again with $n = 4$. To the nearest integer, $N(0) = 32$, $N(1) = 60$, $N(2) = 43$, $N(3) = 13$ and $N(4) = 2$.

---

14) The distribution of (natural) log distance of galaxies in a survey is found to be normal with mean μ and variance σ^2. Derive the pdf, $p(r)$, of the galaxy distance, r, and determine the expected value and variance of r.

Solution to (14) Let $l = \ln r$. Then $p(r)|\,\mathrm{d}r| = p(l)|\,\mathrm{d}l|$ and we are told that $p(l)$ is Normal. By change of variable $p(r)$ is therefore Lognormal (see previous questions) and has the form

$$p(r) = \frac{1}{r}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{\ln r - \mu}{\sigma}\right)^2\right].$$

The expectation value of r is $E(r) = \int_0^\infty r\,p(r)\,\mathrm{d}r$. This can be most easily evaluated by making the substitution $r = \mathrm{e}^l$ again to give

$$E(\mathrm{e}^l) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2\sigma^2}(l-\mu)^2 + l\right]\mathrm{d}l.$$

This integral can be evaluated by completing the square to give

$$E(r) = E(\mathrm{e}^l) = \exp(\mu + \sigma^2/2).$$

The variance we get from $\mathrm{var}(r) = E(r^2) - [E(r)]^2$. Again by completing the square the first term can be integrated, giving a final answer of

$$\mathrm{var}(r) = \exp(2\mu + \sigma^2)\left[\exp(\sigma^2) - 1\right].$$

---

15) Let X and Y be random variables. Find expressions for the following in terms of the variance and covariance of X and Y.

(a) $\mathrm{var}(aX)$ and $\mathrm{var}(aY)$, where a is a constant.

(b) $\mathrm{cov}(aX, aY)$, where a is a constant.

(c) $\mathrm{cov}(X, X+Y)$

(d) $\mathrm{cov}(X+Y, X-Y)$

Show that $\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\,\mathrm{cov}(X,Y)$.

What property of the correlation coefficient, ρ, is indicated by your solution to (a) and (b)?

Solution to (15)

(a) $\mathrm{var}(aX) = E\left[(aX)^2\right] - [E(aX)]^2 = a^2E(X^2) - a^2\left[E(X)\right]^2 = a^2\,\mathrm{var}(X)$. Similarly $\mathrm{var}(aY) = a^2\,\mathrm{var}(Y)$.

(b) $\mathrm{cov}(aX, aY) = E(aXaY) - E(aX)E(aY) = a^2E(XY) - a^2E(X)E(Y) = a^2\,\mathrm{cov}(X,Y)$.

(c) By similar manipulations $\mathrm{cov}(X, X+Y) = \mathrm{var}(X) + \mathrm{cov}(X,Y)$,

(d) and $\mathrm{cov}(X+Y, X-Y) = \mathrm{var}(X) - \mathrm{var}(Y)$.

Finally $\mathrm{var}(X+Y) = E\left[(X+Y)^2\right] - [E(X+Y)]^2 = E(X^2) + 2E(XY) + E(Y^2) - [E(X)]^2 - 2E(X)E(Y) - [E(Y)]^2 = \mathrm{var}(X) + \mathrm{var}(Y) + 2\,\mathrm{cov}(X,Y)$.

The solutions to a) and b) indicate that the correlation coefficients is independent of a scale factor.

---

16)F The angular diameter, θ, of the expanding photosphere of a type II supernova is observed simultaneously with two different telescopes, A and B, at a number of epochs, with the following results (in arcsec):

A	101.8	102.8	111.0	113.5	114.4	114.8	114.5	116.2	120.2	123.5
B	99.2	103.1	114.8	111.6	110.1	110.3	110.7	114.3	117.6	119.2

It is suspected that, due to incorrect flat-fielding, the results from telescope B differ systematically from those of telescope A. The following model is constructed:

$$\theta(A)_i = \alpha + \beta\,\theta(B)_i + e_i \tag{1}$$

where α and β are constants and the errors, e_i, are normally distributed with mean zero and dispersion, $\sigma = 1.7$.

Determine least-squares estimates for α and β with this model. Use the χ^2 statistic to test the goodness of fit of the model to the data at the 1% significance level.

Solution to (16) Using the least-squares formulas $\hat{\alpha} = -0.097 \pm 10.3$ and $\hat{\beta} = 1.02 \pm 0.09$. Now we evaluate

$$\chi^2 = \sum_{i=1}^{10} \left\{ \frac{\theta(A)_i - \left[\hat{\alpha} + \hat{\beta}\theta(B)_i\right]}{1.7} \right\}^2 = 20.52.$$

The critical value for χ^2 is 20.1 for the 1% significance level so the null hypothesis is (just!) rejected at the 1% level.

---

17)B Observations made over several decades have resulted in a value of Hubble's constant H with a (Bayesian) probability

$$p(H|I) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma_\mathrm{p}} \exp\left[-\frac{(H - H_\mathrm{p})^2}{2\sigma_\mathrm{p}^2}\right],$$

where H_p and σ_p are constants and I represents the relevant background information. A new technique measures Hubble's constant using an entirely different method, reporting a value H_d with uncertainty σ_d. Taking the probability of this new data to be Gaussian, show that the updated value of H is described by

$$p(H|H_\mathrm{d}, I) = \frac{1}{\sqrt{2\pi}}\frac{1}{\Sigma} \exp\left[-\frac{(H - m)^2}{2\Sigma^2}\right],$$

where

$$\frac{1}{\Sigma^2} = \frac{1}{\sigma_\mathrm{p}^2} + \frac{1}{\sigma_\mathrm{d}^2}$$

and

$$m = \Sigma^2\left(\frac{H_\mathrm{d}}{\sigma_\mathrm{d}^2} + \frac{H_\mathrm{p}}{\sigma_\mathrm{p}^2}\right).$$

Comment on this result when $\sigma_\mathrm{d} \ll \sigma_\mathrm{p}$.

Solution to (17) The question supplies us with a normalised prior for H:

$$p(H|I) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma_\mathrm{p}} \exp\left[-\frac{(H - H_\mathrm{p})^2}{2\sigma_\mathrm{p}^2}\right],$$

The new data gives the likelihood of H. We are told this is gaussian, so

$$p(H_\mathrm{d}|H, I) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma_\mathrm{d}} \exp\left[-\frac{(H - H_\mathrm{d})^2}{2\sigma_\mathrm{d}^2}\right].$$

By Bayes theorem, the posterior pdf is proportional to the product of these two and is therefore also a Gaussian. Setting

$$\begin{aligned} L &= -\frac{(H - H_\mathrm{p})^2}{2\sigma_\mathrm{p}^2} - \frac{(H - H_\mathrm{d})^2}{2\sigma_\mathrm{d}^2} \\ &\equiv -\frac{(H - m)^2}{2\Sigma^2}, \end{aligned}$$

Our quest is simply to evaluate m and Σ. Differentiating L twice wrt H:

$$\frac{\partial^2 L}{\partial H^2} \equiv -\frac{1}{\Sigma^2} = -\frac{1}{\sigma_\mathrm{p}^2} - \frac{1}{\sigma_\mathrm{d}^2}.$$

The maximum of the pdf occurs at $H = m$. Differentiating once and setting to zero:

$$0 = -\frac{(m - H_\mathrm{p})}{\sigma_\mathrm{p}^2} - \frac{(m - H_\mathrm{d})}{\sigma_\mathrm{d}^2}$$
$$m\left(\frac{1}{\sigma_\mathrm{p}^2} + \frac{1}{\sigma_\mathrm{d}^2}\right) = \frac{H_\mathrm{d}}{\sigma_\mathrm{d}^2} + \frac{H_\mathrm{p}}{\sigma_\mathrm{p}^2}$$
$$m = \Sigma^2\left(\frac{H_\mathrm{d}}{\sigma_\mathrm{d}^2} + \frac{H_\mathrm{p}}{\sigma_\mathrm{p}^2}\right).$$

When $\sigma_\mathrm{d} \ll \sigma_\mathrm{p}$ our data is much better than our prior, so the result should be dominated by the new information. The above expressions reduce to

$$\Sigma = \sigma_\mathrm{d} \quad ; \quad m = H_\mathrm{d},$$

as one would expect in those circumstances.

---

18)B A model exists for a new type of X-ray source. If these sources are distributed homogeneously in the Universe (i.e., with constant mean density) show that the Bayesian probability density function (pdf) for the radial distance, r, to a source is

$$p(r) \propto r^2.$$

The sources are thought to radiate isotropically so that the flux F from a source at a distance r is

$$F = \frac{L}{4\pi r^2},$$

where L is the intrinsic luminosity of the source. Given that the pdf of L for any particular source is $p(L)$ show (by considering the joint pdf of r and L) that the joint pdf of L and F for an *observed* source is

$$p_\mathrm{obs}(L, F) \propto p(L)L^{3/2}F^{-5/2}.$$

You may use the general result that

$$p(y, z) = p(u, v)\begin{vmatrix} \frac{\partial u}{\partial y} & \frac{\partial u}{\partial z} \\ \frac{\partial v}{\partial y} & \frac{\partial v}{\partial z} \end{vmatrix}.$$

To be observed, the flux from a source must be greater than F_min. Assuming the pdf for source luminosity is exponential, i.e.,

$$p(L) \propto \exp(-L/L_0),$$

and given $\int_0^\infty x^{3/2}\mathrm{e}^{-x}\,\mathrm{d}x = 3\sqrt{\pi}/4$, determine the normalised pdf for $p_\mathrm{obs}(L, F)$.
Show that $p(L) \neq p_\mathrm{obs}(L)$ and comment on why this is so, physically. What is the most probable observed luminosity?
Poisson statistics tells us that the probability of receiving n photons in a time τ s from a source of mean flux F is

$$P(n|F,\tau) = \frac{\mathrm{e}^{-F\tau}(F\tau)^n}{n!},$$

where F has units of photons per second. During an observation lasting one second, 6 photons are seen from one of the new sources mentioned in part (b). Determine the un-normalised posterior pdf for the flux of this source, independent of L.
What is the most probable flux of the source? Comment on why it is significantly less than 6 photons per second?

Solution to (18) Homogeneity implies that the probability that a source is in any particular volume of space is proportional to that volume. In spherical polar coordinates

$$p(r, \theta, \phi)\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\phi \propto r^2 \sin\theta\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\phi.$$

Marginalising over θ and ϕ we get

$$p(r) = \iint p(r, \theta, \phi|\mathrm{uniform})\,\mathrm{d}\theta\,\mathrm{d}\phi$$
$$\propto \iint r^2 \sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi$$
$$\propto r^2.$$

For any particular source, two parameters r and L are not related, in the sense that $p(L|r) = p(L)$. We can therefore say that

$$\begin{aligned} p(L,r) &= p(L)p(r) \\ &\propto r^2 p(L). \end{aligned}$$

We now change the parameters to L and F using

$$\begin{aligned} p_{\rm obs}(L,F) &= p(L,r)\left| \begin{matrix} \frac{\partial L}{\partial L} & \frac{\partial L}{\partial F} \\ \frac{\partial r}{\partial L} & \frac{\partial r}{\partial F} \end{matrix} \right| \\ &= p(L)p(r)\left(\frac{\partial r}{\partial F} - \frac{\partial L}{\partial F}\frac{\partial r}{\partial L}\right) \\ &= p(L)p(r).\frac{4\pi r^3}{L} \\ &\propto \frac{p(L)}{L}\left(\frac{L}{F}\right)^{5/2} \\ &\propto p(L)L^{3/2}F^{-5/2}. \end{aligned}$$

Normalising:

$$\int_{F_{\rm min}}^{\infty} F^{-5/2}\ {\rm d}F = \frac{2}{3}F_{\rm min}^{-3/2}$$

and

$$\begin{aligned} \int_0^{\infty} L^{3/2}\exp(-L/L_0)\ {\rm d}L &= L_0^{5/2}\int_0^{\infty} x^{3/2}\exp(-x)\ {\rm d}x \\ &= \frac{3\pi^{1/2}L_0^{5/2}}{4}, \end{aligned}$$

so

$$p_{\rm obs}(L,F) = \frac{2}{\sqrt{\pi}}F_{\rm min}^{3/2}L_0^{-5/2}F^{-5/2}L^{3/2}\exp(-L/L_0).$$

$p_{\rm obs}(L,F)$ factorises nicely, so the marginal pdf for $p_{\rm obs}(L,F)$ is simply

$$p_{\rm obs}(L) = \frac{4}{3\sqrt{\pi}}L_0^{-5/2}L^{3/2}\exp(-L/L_0).$$

Clearly there is an extra $L^{3/2}$ term in there, not present in $p(L)$. Taking logs

$$\ln p_{\rm obs}(L) = \frac{3}{2}\ln L - \frac{L}{L_0} + \text{const},$$

and differentiating wrt L we get a maximum in the pdf at at $L = 3L_0/2$. The reason is that we are now looking at the probability of L in an *observed* source, and there are many more of these far away, so it is more probable that we are seeing a relatively distant but (improbably) luminous source rather than a close-by low luminosity source.

We use Bayes Theorem for this:

$$p(F|n) = \frac{p(F)P(n|F)}{P(n)} \propto p(F)P(n|F).$$

From part (c) we know that

$$p_{\rm obs}(L,F) \propto p(L)L^{3/2}F^{-5/2},$$

so (maginalising implicitly over L) $p(F) \propto F^{-5/2}$. Therefore

$$\begin{aligned} p(F|n) &\propto F^{-5/2}.\frac{e^{-F}F^n}{n!} \\ &\propto F^{n-5/2}e^{-F}. \end{aligned}$$

If $n = 6$ then

$$p(F|n) \propto F^{7/2}e^{-F}.$$

We want the value of F at the maximum of the pdf, F_{mp}.

$$\begin{aligned} p(F|n) &\propto F^{7/2}e^{-F} \\ \ln p &= \frac{7}{2}\ln F - F \\ \frac{d\ln p}{dF} &= \frac{7}{2F_{\text{mp}}} - 1 = 0 \\ F_{\text{mp}} &= \frac{7}{2}. \end{aligned}$$

We saw 6 photons in 1 second, so it would be reasonable at first to assume the flux of the source was about 6/s. However most sources are far away and dim, so it turns out that it's more likely that this is a dim distant source, with a mean flux less than 6/s, which has generated a larger than average number of photons in this particular second, than a close source behaving nominally.
