ICIC Data Analysis Workshop

Day 1 solutions

September 12, 2013

1 Probabilistic reasoning

1. What probability do you assign that the coin will, when tossed, land heads? (Try and justify your answer.)

   Answer: Most people will instantly go for $\Pr(H) = 0.5$ almost without thinking, but the important point here is not the answer itself, but the reasoning behind it. When pushed, probably the most common justification is slightly tautological, along the lines that "a coin flipped many times will come up heads about half the time". However a more powerful argument relates to the information that you have at this stage, which is that you have no reason to prefer heads to tails – and hence if the labels were swapped (*i.e.*, the sides of the coin were given new names), the outcome should be the same. With only two possibilities this symmetry principle also leads to $\Pr(H) = 0.5$, without needing to invoke hypothetical multiple flips.

   The coin is tossed and the umpire catches it, immediately covering it with his other hand. What is your probability the coin has landed heads? What is the umpire's probability the coin has landed heads?

   Answer: Neither you nor the umpire have gained any additional information, so both of you should still assign $\Pr(H) = 0.5$. The fact that the coin is now definitely either heads or tails is irrelevant, because you don't know which.

   The umpire lifts his hand a little to check the coin, which has landed heads, but you can't see it. What is your probability the coin has landed heads? What is the umpire's probability the coin has landed heads?

   Answer: The umpire can now assign $\Pr(H) = 1$, but you are stuck with $\Pr(H) = 0.5$. This emphasizes that two people can have different – and even potentially contradictory – states of belief because they have different information.

   You are now shown the coin, and also see that it has landed heads. What is your probability the coin has landed heads? What is the umpire's probability the coin has landed heads?

   Answer: Finally both you and the umpire have the same – and complete – information, and can both say that $\Pr(H) = 1$.

Now think through the above sequence in full. At what point did the answers change? Was it related to physical events? Was it related to the information other people had? Was it related to the information you had?

Answer: (Answered sequentially above.)

Now imagine that someone had, before the toss, passed you a note saying that the coin is biased, and will land one way up 99% of the time (but, unhelpfully, that the note doesn't say which way the coin is biased). Answer the above four questions again. How has this new piece of inside information changed your answers? What if you'd noticed the umpire tossing the coin beforehand while you were signing autographs and you'd noticed it had landed heads that time?

Answer: Running through the normal sequence the seemingly strong piece of information that the coin is biased is of no help, as it doesn't break the symmetry of the situation. There's still no reason to prefer heads to tails, so $\Pr(H) = 0.5$ is still your only reasonable answer until you get to see the result of the final flip.

If you catch a glimpse of the umpire's idle pre-match toss then your insider knowledge about the weighted coin is immediately very powerful. Assuming you believe the note that you mysteriously received, there are two possible models to consider: that the coin is 99% biased towards heads (hypothesis h) or that it is 99% biased towards tails (hypothesis t). At the point of having received the note symmetry demands $\Pr(h) = \Pr(t) = 1/2$.

Then you see the umpire's idle flip lands heads. You can now include this information to modify your probabilities to be

$$\Pr(h|H) = \frac{\Pr(h)\Pr(H|h)}{\Pr(h)\Pr(H|h) + \Pr(t)\Pr(H|t)}, \tag{1}$$

where H is the event that the umpire's idle toss landed heads. From the information in the note $\Pr(H|h) = 0.99$ and $\Pr(H|t) = 0.01$, so that $\Pr(h|H) = 0.99$. Clearly you should call "heads" at the toss, and most likely you will be correct. You can also work out the chance you could be wrong – this would happen either if the coin was in fact biased towards tails (a 1% chance) or if the heads-biased coin just happened to land tails when tossed (also a 1% chance). So your probability of calling correctly is 0.98.

(All of the above calculation only assumes that you believe the note claiming that the coin is highly biased. It would also be possible to assign this information some partial credence – or no credence at all, in which case the original results are recovered.)

2. The Monty Hall problem: you have three doors in front of you, two of which have bad prizes and one with a good prize. You select one door, and the host opens another and shows that it has one of the bad prizes. Should you switch?

   Answer:

   Let the doors be labelled a, b, c, where a is the door you choose initially, and b is the door which is opened. Many, if not all, of the probabilities below should be interpreted as 'given that you have chosen a', but for clarity we won't write this explicitly.

Let $p(a)$ = probability that a leads to the desired prize, etc.

Let B be the event that door b gets opened and leads to worthless junk.

What you want is the probability that a leads to the prize, given that b is opened and leads to junk. i.e. the aim is to calculate

$$p(a|B).$$

We can use Bayes' theorem for this:

$$p(a|B) = \frac{p(a,B)}{p(B)} = \frac{p(B|a)p(a)}{p(B)}$$

Now, clearly $p(a) = p(b) = p(c) = 1/3$ (all doors are equally likely, before any experiment is done).

$p(B|a)$ = probability that door b is opened, given that a leads to the prize. Evidently

$$p(B|a) = \frac{1}{2}:$$

Alan could have opened either door b or c, since they both lead to junk.

What about $p(B)$? It is the sum of all the joint probabilities:

$$p(B) = p(B,a) + p(B,b) + p(B,c) = p(B|a)p(a) + p(B|b)p(b) + p(B|c)p(c),$$

each of which we can calculate. $p(a) = p(b) = p(c) = 1/3$, as before, and $p(B|a) = 1/2$ as before. Now

$$p(B|b) = 0:$$

Alan will not open b since it leads to the prize in this case.

$p(B|c)$ is the most interesting. Given that you have chosen a (remember this is implicit throughout), then if c leads to the prize, then Prof Heavens *must* open door b, i.e.

$$p(B|c) = 1$$

So the probability that your original choice a leads to the prize is

$$\begin{aligned} p(a|B) &= \frac{p(B|a)p(a)}{p(B|a)p(a) + p(B|b)p(b) + p(B|c)p(c)} \qquad (2) \\ &= \frac{\frac{1}{2}\frac{1}{3}}{\frac{1}{2}\frac{1}{3} + \left(0 \times \frac{1}{3}\right) + \left(1 \times \frac{1}{3}\right)} \\ &= \frac{1}{3} \end{aligned}$$

So you would double your chances (from 1/3 to 2/3) if you switch to the other door.

3. A pan contains 10 ravioli, of which 9 are filled with pesto and one with ricotta. You put in the pan a further raviolo filled with pesto and cover with an opaque lid. Then you randomly draw a raviolo, eat it and discover that it is filled with pesto. After this procedure, the pan is again in the same state as before. What is now the probability that the next raviolo drawn will be filled with pesto?

   On a different night, you cook a pan of mixed pesto and ricotta ravioli (in equal proportions). One last raviolo remains in your plate, which could be either pesto or ricotta. Your friend tosses into your plate her last raviolo, which she tells you is a pesto-filled one. Then you mix the two ravioli randomly, pick one and realize it's pesto. What is now the probability that the last raviolo in your plate, *before your friend threw hers in*, is pesto? What is the probability that the *very last* raviolo is pesto?

   Answer:

   Note that the following doesn't really answer the first question, which really just requests $P(2nd = p|1st = p) = 9/10$.

   Let us denote by p a pesto raviolo, by r a raviolo filled with ricotta. Then the joint probability of drawing two pesto ravioli is

$$P(1st = p, 2nd = p) = P(2nd = p|1st = p)P(1st = p). \tag{3}$$

   The probability that the first raviolo is filled with pesto is $P(1st = p) = 10/11$, while for the second (conditional) draw $P(2nd = p|1st = p) = 9/10$. So the combined probability is $P(1st = p, 2nd = p) = 9/11$.

   In the second case, for the raviolo in your plate (yr) you have a prior $P(yr = p) = P(yr = r) = 1/2$. After you have picked the raviolo and found it to be pesto ($p = 1$) your posterior probability for the remaining raviolo in your plate is:

$$\begin{aligned} P(yr = p|p = 1) &= \frac{P(p = 1|yr = p)P(yr = p)}{P(p = 1|yr = p)P(yr = p) + P(p = 1|yr = r)P(yr = r)} \\ &= \frac{1}{1 + \frac{P(p=1|yr=r)P(yr=r)}{P(p=1|yr=p)P(yr=p)}} = \frac{1}{1 + \frac{\frac{1}{2}\cdot\frac{1}{2}}{1\cdot\frac{1}{2}}} = 2/3. \end{aligned} \tag{4}$$

   Finally, we can ask whether the final raviolo is pesto, which can we denote "$f = p$". Since we already know the probability for $yr = p$ from the above, the easiest way to do this is by marginalization:

$$\begin{aligned} P(f = p|p = 1) &= \sum_{yr}^{p,r} P(f = p, yr|p = 1) \\ &= \sum_{yr}^{p,r} P(f = p|yr, p = 1)P(yr|p = 1) \\ &= \quad P(f = p|yr = p, p = 1)P(yr = p|p = 1) \\ &\quad + P(f = p|yr = r, p = 1)P(yr = r|p = 1) \\ &= 1 \times 2/3 + 0 \times 1/3 = 2/3 \end{aligned} \tag{5}$$

where $P(f = p|yr = p, p = 1)P(yr = p|p = 1) = 1$ since given $yr = p$ there were two pesto on your plate so that the final raviolo is pesto, and, alternately, $P(f = p|yr = r, p = 1)P(yr = r|p = 1)$, since if $yr = r$ and you've already drawn a pesto ($p = 1$), the remaining one must be ricotta.

4. A body has been found on the Baltimore West Side, with no apparent wounds, although it transpires that the deceased, a Mr Fuzzy Dunlop, was a heavy drug user. The detective in charge suggests to close the case and to attribute the death to drugs overdose, rather than murder.

   Knowing that, of all murders in Baltimore, about 30% of the victims were drug addicts, and that the probability of a dead person having died of overdose is 50% (without further evidence apart from the body) estimate the probability that the detective's hunch is correct. You may assume that in crime-ridden Baltimore all deaths (at least those investigated by the detective) are either by overdose or murder. Do you have to make any other assumptions?

   Answer: Let us denote by $od = 1$ the statement "Mr Dunlop died because of drugs overdose"; by $Dd = 1$ the statement "Mr Dunlop is dead" and by $u = 1$ the statement "Mr Dunlop used drugs".

   We are looking for the posterior probability that Fuzzy Dunlop died of overdose, given that he was a drug addict ($u = 1$) and that he is dead ($Dd = 1$):

$$P(od = 1|Dd = 1, u = 1) = \frac{P(u = 1|od = 1, Dd = 1)P(od = 1|Dd = 1)}{P(u = 1|od = 1, Dd = 1)P(od = 1|Dd = 1) + P(u = 1|od = 0, Dd = 1)P(od = 0|Dd = 1)} \tag{6}$$

   From the problem, we have that the probability of being a drug user and having been murdered (assuming that people only die of either overdose or murder in Baltimore) is $P(u = 1|od = 0, Dd = 1) = 0.3$. Also, the probability of the person having died of overdose (given that we have the body) is 50%, hence $P(od = 1|Dd = 1) = 50\%$ so $P(od = 0|Dd = 1) = 50\%$.

   Finally, we need to estimate the probability that Mr Dunlop was a drug user, given that he died of overdose, $P(u = 1|od = 1, Dd = 1)$. It seems highly unlikely that somebody would die of overdose the first time they try drugs, so perhaps we can assign $P(u = 1|od = 1, Dd = 1) = 0.9$.

   So we have that

$$P(od = 1|Dd = 1, u = 1) = \frac{1}{1 + \frac{P(u=1|od=0,Dd=1)P(od=0|Dd=1)}{P(u=1|od=1,Dd=1)P(od=1|Dd=1)}} = \frac{1}{1 + 3/9} = 75\%. \tag{7}$$

   How sensitive is this conclusion to our estimate for $P(u = 1|od = 1, Dd = 1)$? Changing this to $P(u = 1|od = 1, Dd = 1) = 0.5$ (we are agnostic as to whether a drug overdose is more likely for usual drugs consumers or for novices) gives a posterior $P(od = 1|Dd = 1, u = 1) = 62\%$, while increasing it to $P(u = 1|od = 1, Dd = 1) = 0.99$ (most people overdosing are drugs users) gives $P(od = 1|Dd = 1, u = 1) = 77\%$.

2 Single parameter inference

1. The distribution of flux densities of extragalactic radio sources are distributed as a power-law with slope $-\alpha$, say. In a non-evolving Euclidean universe $\alpha = 3/2$ and departure of α from the value 3/2 is evidence for cosmological evolution of radio sources. This was the most telling argument against the steady-state cosmology in the early 1960's (even though they got the value of a wrong by quite a long way).

   Given observations of radio sources with flux densities S above a known, fixed measurement limit S_0, what is the best estimate for α?

   The model probability distribution for S is

   $$p(S)dS = (\alpha - 1)S_0^{\alpha-1}S^{-\alpha}dS$$

   where the factor $\alpha - 1$ in front of the terms arises from the normalization requirement

   $$\int_{S_0}^{\infty} dS\, p(S) = 1.$$

   So the likelihood function L for n observed sources is

   $$L = \prod_{i=1}^{n}(\alpha - 1)S_0^{\alpha-1}S_i^{-\alpha}$$

   with logarithm

   $$\ln L = \sum_{i=1}^{n}[\ln(\alpha - 1) + (\alpha - 1)\ln S_0 - \alpha \ln S_i].$$

   Maximising $\ln L$ with respect to α:

   $$\frac{\partial}{\partial \alpha}\ln L = \sum_{i=1}^{n}\left(\frac{1}{\alpha - 1} + \ln S_0 - \ln S_i\right) = 0$$

   we find the minimum when

   $$\alpha = 1 + \frac{n}{\sum_{i=1}^{n}\ln\frac{S_i}{S_0}}.$$

   Suppose we only observe one source with flux twice the cut-off, $S_1 = 2S_0$, then

   $$\alpha = 1 + \frac{1}{\ln 2} = 2.44$$

   but with a large uncertainty. Clearly, as $S_i = S_0$ we find $\alpha \to \infty$ as expected. In fact $\alpha = 1.8$ for bright radio sources at low frequencies, significantly steeper than 1.5.

   To get a rough estimate of the width of the credibility interval for α, we can compute

   $$\sigma_\alpha^{-2} \simeq -\frac{\partial^2 \ln L}{\partial \alpha^2} = \sum_{i=1}^{n}\frac{1}{(\alpha - 1)^2},$$

evaluated at the maximum likelihood solution. This is not very accurate as the likelihood shape is not gaussian. For the single source, $\alpha = 2.44$ and $n = 1$, so the (rough) estimate of the error is

$$\sigma_\alpha = 1.44.$$

3 Optional Problems

1. **Optional.** An astronomer wishes to know the (mono-chromatic) flux of a particular source and makes a photometric measurement which registers $N_{\rm src}$ photons. Assume that all the photons have come from the source itself (*i.e.*, there is no background or or source confusion) and that the known calibration constant, C, is such that a source of true flux $F_{\rm src}$ would, on average, yield $F_{\rm src}/C$ photons in such a measurement (*i.e.*, a generic estimate of the source's flux would be $\hat{F}_{\rm src} \simeq C N_{\rm src}$).

   (a) What is the model parameter that the astronomer is trying to infer?

   Answer: The true flux of the source, $F_{\rm src}$. (Even though this is a definite physical number, it is reasonable to consider it's value in probabilistic terms, as it is not uniquely/logically determined by the data.)

   (b) What is/are the datum/data?

   Answer: The datum is $N_{\rm src}$, the number of photons registered in the measurement of the source.

   (c) What is the likelihood [*i.e.*, the probability $\Pr(N_{\rm src}|F_{\rm src})$]?

   Answer: The starting point for answering this question is to see that photons from the source hit the detector at a given rate ($F_{\rm src}/C$ per unit observation time) but that the photons propagate independently. This implies that the number of photons that hit the detector in a given period is Poisson distributed, and so

   $$\Pr(N_{\rm src}|F_{\rm src}) = \frac{(F_{\rm src}/C)^{N_{\rm src}} e^{-F_{\rm src}/C}}{N_{\rm src}!}. \qquad (8)$$

   In the case of bright sources, for which $F_{\rm src}/C \gg 1$, the distribution of $N_{\rm src}$ is still Poisson, although mathematically extremely well approximated as a Gaussian of the form

   $$\Pr(N_{\rm src}|F_{\rm src}) \propto \frac{1}{(F_{\rm src}/C)^{1/2}} e^{-1/2(N_{\rm src}-F_{\rm src}/C)^2/(F_{\rm src}/C)}, \qquad (9)$$

   where, in the large $N_{\rm src}$ limit, it is being treated as a continuous variable. This equation is no longer correctly normalised as an awkward sum over $N_{\rm src}$ must be done; however the relative probabilities of the different possible $N_{\rm src}$ values for a given $F_{\rm src}$ are correct. More importantly, the likelihood is a smooth function of $F_{\rm src}$, and it is this interpretation that will be required for later inference. However, whilst $\Pr(N_{\rm src}|F_{\rm src})$ is a Gaussian in $N_{\rm src}$, it is not Gaussian in terms of $F_{\rm src}$, as $F_{\rm src}$ appears in the normalising constant and in the denominator of the exponential.

It is important not only to obtain the mathematical form of the likelihood but also to understand what it means. It is *not* the probability of $F_{\rm src}$, even though in some cases it might have a similar form (*e.g.*, peaked in the same place, or with a similar spread). It *is* only the probability that $N_{\rm src}$ photons would be received from the source *if* its flux was $F_{\rm src}$.

(d) What prior information might the astronomer have *before* making (or at least making use of) the measurement?

Answer: You, as an astronomer, are very far from total ignorance about astronomical sources and their fluxes. If you know the type of the source (*e.g.*, a quasar or a Galactic star, *etc.*) then previous astronomical knowledge about all sorts of astronomical sources. Even without any particular knowledge about the type of source, there is the generic fact that, due to geometry, there are significantly more faint sources than bright sources. The immediate implication is that, in any situation where the data do not strongly constrain the source's flux, it will be important to include the preponderence of faint sources in the prior.

(e) If the astronomer had access to a catalogue of sources of similar fluxes from a different part of the sky, how might this catalogue be used to generate an appropriate, if approximate, prior distribution for the source's true flux, $F_{\rm src}$?

Answer: The complicated nature of astronomical surveys – and particular their attendant selection effects – makes this a potentially difficult question to answer. However the underlying principle is that the observed flux distribution of the sources in question would serve as a good, if approximate, prior for the flux of the source of interest.

(f) If the distribution of source fluxes was known to increase as $\Pr(F_{\rm src}) \propto F_{\rm src}^{-5/2}$, what would the resultant posterior information on the source's flux be upon combining this knowledge about the source population and the data on the particular source of interest? Is this prior normaliseable (*i.e.*, proper)?

Answer: The prior implied is (up to a normalisation constant)

$$\Pr(F_{\rm src})\Theta(F_{\rm src}) \propto F_{\rm src}^{-5/2}, \tag{10}$$

where $\Theta(x)$ is the Heavyside step function, to ensure that the prior is zero for negative fluxes. This might seem a little fussy, but in exploring an unfamiliar problem it is generally worth being more careful/explicit about the assumptions you're making.

The posterior distribution of the source's true flux would then be (up to a normalisation constant)

$$\Pr(F_{\rm src}|N_{\rm src}) \propto \Theta(F_{\rm src})(F_{\rm src}/C)^{N_{\rm src}-5/2}e^{-F_{\rm src}/C}. \tag{11}$$

In the limit of a large number of photons, the Gaussian approximatin invoked above leads to the posterior

$$\Pr(F_{\rm src}|N_{\rm src}) \propto \Theta(F_{\rm src})F_{\rm src}^{-3}e^{-1/2(N_{\rm src}-F_{\rm src}/C)^2/(F_{\rm src}/C)}. \tag{12}$$

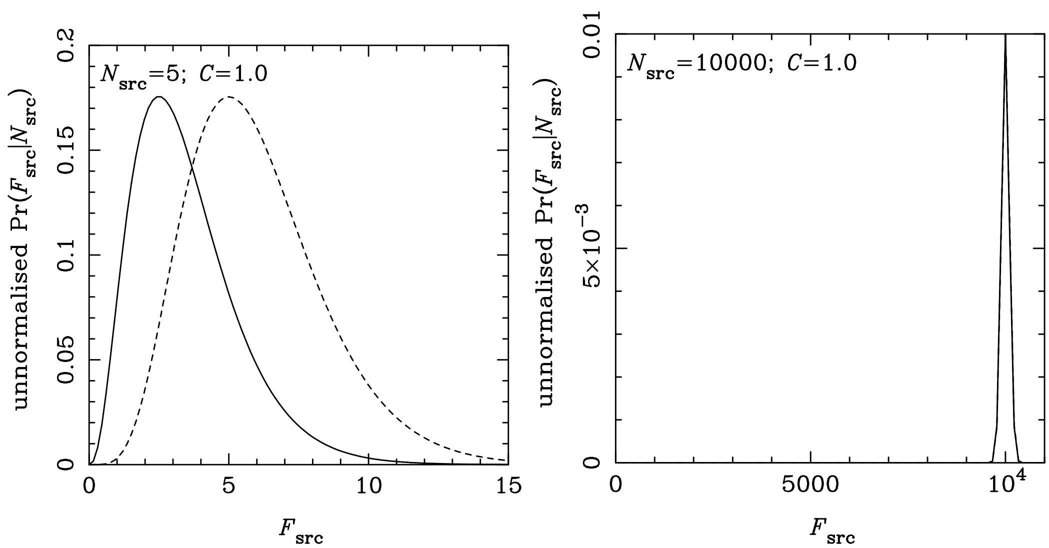

Figure 1: Unnormalised posterior in the source flux, $F_{\rm src}$ in the cases where $N_{\rm src} = 5$ (left) and $N_{\rm src} = 10^4$ (right). In both cases the dashed lines show the likelihood as a function of $F_{\rm src}$.

The prior is not normaliseable unless a minimum flux, $F_{\rm min}$ is assumed (or justified somehow), and so care must be taken with these posteriors to check that they are normaliseable. The obvious potential problem is as $F_{\rm src} \to 0$, as it is here that the improper prior becomes infinite. The prior diverges as a power-law, as does the likelihood, when expressed as a function of $F_{\rm src}$, although the latter is dominant provided $N_{\rm src} > 5/2$, so the posterior is bounded and integrable. The Gaussian approximation does not have this property, however, and the likelihood is finite, if very small, at $F_{\rm src} = 0$, leading to a sharp "spike" in the posterior at $F_{\rm src} = 0$ that contains infinite probability. This is an artefact of the Gaussian approximation to the Poisson likelihood and is not a serious problem in practice.

(g) Assuming, for simplicity, that $C = 1$, plot both the likelihood, $\Pr(N_{\rm src}|F_{\rm src})$, and the posterior distribution, $\Pr(F_{\rm src}|N_{\rm src})$, as a function of $F_{\rm src}$ in i) the case that $N_{\rm src} = 5$ (plausible for an X-ray observation) and ii) the case that $N_{\rm src} = 10^4$ (plausible for an optical observation). Are any of these functions approximately Gaussian? What is the probability that the source has $F_{\rm src} = 0$? What is the probability that the source has $F_{\rm src} < 0$? How did utilising the photometric measurement of the source affect these probabilities?

Answer:

The likelihoods and unnormalised posterior distributions are shown in Fig. 1. In the $N_{\rm src} = 5$ case the full Poisson formula is used; in the $N_{\rm src} = 10^4$ case the Gaussian approximation is adopted. In the latter case the posterior and likelihood are almost indistinguishable and also both very close to Gaussian. The prior does not play a strong role as the high-precision measurement is much more informative. In the $N_{\rm src} = 5$ case, however, the measurement contains far less information and the source is probably fainter than the data might naively be taken to indicate.

(h) What do you think is a reasonable "best estimate" of the source's flux? (There are several plausible answers.) How do these best estimates relate to the naive estimate $\hat{F}_{\rm src} = C N_{\rm src}$? Does this make sense?

Answer: The full answer to any Bayesian parameter estimation problem is the posterior distribution in the parameter(s) of interest. However in many practical situations (*e.g.*, reporting flux estimates of millions of sources) there is no way of assimilating or visualising the full distribution. Hence it is useful to try and condense it into, *e.g.*, an estimated value and an error. That said, there can be no definitive algorithm for doing this. In some cases a few parameters can completely encapsulate the posterior (*e.g.*, the mean/mode/median and standard deviation if it's Gaussian), but in most cases this is not strictly possible.

For singly-peaked distributions it is reasonable to use the peak of the posterior, or the median or the mean. Whichever of these characterising numbers is chosen will be less than the "natural" estimator, $\hat{F}_{\rm src} = C N_{\rm src}$. This result is potentially counter-intuitive, especially if you've gotten used to using sampling statistis. One of the first tests many people would run to test an algorithm being used to estimate some quantity of interest would be to generate lots of fake data with the flux equal to some known $F_{\rm src}$ and then see if the resultant estimates (from the peak or mean or whatever) are centred around the true value. Bayesian estimates do *not* satisfy this test (unless the prior happens to be symmetric about $F_{\rm src}$). The reason is that the prior distribution reflects the distribution of source fluxes in the Universe, which is explicitly contradicted if one simulates data with a single flux value.

Put another way, in any real astronomical measurement most of the sources with photon counts $N_{\rm src}$ will have true fluxes which are less than $F_{\rm src} = C N_{\rm src}$ as there are more faint sources which are randomly scattered bright than there are brighter sources scattered faint. This phenomenon has long been known as Eddington bias, where the term "bias" is used because of the fact that conventional flux estimates are biased high. In terms of Bayesian statistics it would simply be the result of having made a poor choice of prior (that didn't reflect the prevalence of faint sources).

2. The astronomer, having become disillusioned with the lazy data-reporting practices in optical astronomy, has moved into X-ray astronomy. Having found a source of interest, the astronomer falls back on old habits and can't resist trying to do some photometry, just for old time's sake. This is something of a shock, however, both because the expected number of photons from the source is very small (*i.e.*, single figures) and also because there is now an appreciable background (*i.e.*, maybe a third of the photons registered might not have been emitted from the target source). The basic task is the same as in Section 2 – to infer the flux of the source – but now there is the additional complication of a background which must be included in the model. Just as a source of (true) flux $F_{\rm src}$ would provide an average of $\bar{N}_{\rm src} = F_{\rm src}/C$ photons in this measurement, the background flux (in the measurement aperture), $F_{\rm bkg}$, would be expected to contribute $\bar{N}_{\rm bkg} = F_{\rm bkg}/C$ photons in such a measurement.

(a) It is quite possible that the background rate is known precisely (or with so much more accuracy than the measurement that it is effectively exact), so that $F_{\rm bkg}$ can be treated as a known constant. Given a single on-source measurement of $N_{\rm on}$ photons, what is the likelihood and the posterior for the source flux [again assuming the prior $\Pr(F_{\rm src}) \propto F_{\rm src}^{-5/2}$]?

Answer: The distribution of source photons contributing to the measurement is as before; the distribution of background photons is also Poisson, but with a mean of $F_{\rm bkg}$ instead of $F_{\rm src}$. The resultant distribution of the total number of photons in the measurement could be obtained by looking at the separate distributions of $N_{\rm src}$ and $N_{\rm bkg}$ and then marginalizing using the fact that $N_{\rm on} = N_{\rm src} + N_{\rm bkg}$. This process has the advantage that it doesn't require any insight or understanding of the problem.

In this case, however, the sampling distribution can be obtained more quickly by noting that the arrival of the source and background photons is independent. Hence the distribution of $N_{\rm on}$ must be Poisson, and the mean must be $(F_{\rm src} + F_{\rm bkg})/C$. Hence the likelihood is

$$\Pr(N_{\rm on}|F_{\rm src}, F_{\rm bkg}) = \frac{[(F_{\rm src} + F_{\rm bkg})/C]^{N_{\rm on}} e^{-(F_{\rm src}+F_{\rm bkg})/C}}{N_{\rm on}!}. \qquad (13)$$

The prior is not at all modified for the fact that there is now a background source of photons – the prior in the model parameters ought to be completely independent of the new constraining measurement that is being considered. Hence the posterior is (up to a normalisation constant),

$$\Pr(F_{\rm src}|N_{\rm src}) \propto \Theta(F_{\rm src}) F_{\rm src}^{-5/2} [(F_{\rm src} + F_{\rm bkg})/C]^{N_{\rm on}} e^{-(F_{\rm src}+F_{\rm bkg})/C}. \qquad (14)$$

(b) Unfortunately, the uncertainty in the background is often significant; in such cases it must also be measured and its level inferred. The astronomer, faced with an uncertain background, now makes two measurements: one with the telescope aperture centred on the source, which yields $N_{\rm on}$ photons, and one with the telescope aperture pointed at a "blank" patch of sky, which yields $N_{\rm off}$ photons. Although the astronomer is only really interested in $F_{\rm src}$, it is also necessary to include the unknown $F_{\rm bkg}$ in the modelling (to be marginalized over later).

Write down the likelihood [*i.e.*, $\Pr(N_{\rm on}, N_{\rm off}|F_{\rm src}, F_{\rm bkg})$ and the prior [*i.e.*, $\Pr(F_{\rm src}, F_{\rm bkg})$]. (Think carefully about the prior for the background flux. If you have a strongly motivated choice of prior make sure it is justified; if you are less certain try working through the problem with different plausible priors that you think might span the possibilities.)

Answer: The two measurements are statistically independent – different photons are registered by each pointing of the telescope – and so the joint likelihood of the two measurements is the product of the two individual likelihoods. Hence

$$\begin{aligned}
\Pr(N_{\rm on}, N_{\rm off}|F_{\rm src}, F_{\rm bkg}) &= \Pr(N_{\rm on}|F_{\rm src}, F_{\rm bkg}) \Pr(N_{\rm off}|F_{\rm src}, F_{\rm bkg}) && (15) \\
&= \frac{[(F_{\rm src} + F_{\rm bkg})/C]^{N_{\rm on}} e^{-(F_{\rm src}+F_{\rm bkg})/C}}{N_{\rm on}!} \frac{(F_{\rm bkg}/C)^{N_{\rm off}} e^{-(F_{\rm bkg})/C}}{N_{\rm off}!}. \\
&= \frac{[(F_{\rm src} + F_{\rm bkg})/C]^{N_{\rm on}} (F_{\rm bkg}/C)^{N_{\rm off}} e^{-(F_{\rm src}+2F_{\rm bkg})/C}}{N_{\rm on}! N_{\rm off}!}. && (16)
\end{aligned}$$

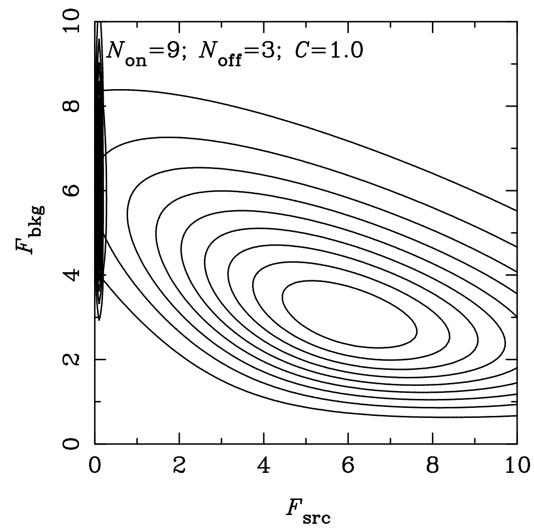

Figure 2: Posterior in the source flux, $F_{\rm src}$, and background flux, $F_{\rm bkg}$, in the cases where $N_{\rm on} = 9$ and $N_{\rm off} = 3$. The contours are at 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1.0 times the peak of the posterior.

The posterior requires the inclusion of the prior. For $F_{\rm src}$ this has already been motivated, but not for $F_{\rm bkg}$. In the absence of other strong arguments a uniform prior is reasonable – although it would be best to look at the results obtained using a variety of plausible priors to see how sensitive they are to the prior. (If they are not it suggests the data are sufficiently informative to give an unambiguous answer, in which case the reason for using a Bayesian analysis is primarily to propagate the information in the data to the final inference. If the results are prior-dependent then the most likely overall truth is that the astronomer does not possess enough information – either from prior knowledge or this measurement – to constrain the source's flux.)

(c) The rest of the problem requires techniques to be discussed on day 2 of the workshop. The full posterior $\Pr(F_{\rm src}, F_{\rm bkg}|N_{\rm on}, N_{\rm off})$ is fairly complicated (whatever prior for the background level was chosen). To explore this distribution without the need for any significant additional programming (or algebra), generate 10^5 samples from the full posterior using MCMC in the case that $N_{\rm on} = 9$ and $N_{\rm off} = 3$ (and, again for convenience, that $C = 1$). Make a scatter plot showing the range of plausible $F_{\rm src}$ and $F_{\rm bkg}$ values. Are they independent or correlated? Can you explain this intuitively? Are these two parameters linked physically at all (*i.e.*, does the flux of a particular source have anything to do with the background)?

Answer:

Adopting the $\Pr(F_{\rm src}) \propto F_{\rm src}^{-5/2}$ prior for the source flux and a uniform prior for the background flux yields

$$\Pr(F_{\rm src}, F_{\rm bkg}|N_{\rm on}, N_{\rm off}) \qquad (17)$$
$$\propto \Theta(F_{\rm src})\Theta(F_{\rm bkg})(F_{\rm src}/C)^{-5/2}[(F_{\rm src} + F_{\rm bkg})/C]^{N_{\rm on}}(F_{\rm bkg}/C)^{N_{\rm off}} e^{-(F_{\rm src}+2F_{\rm bkg})/C}.$$

All but the most mathematically gifted will, presumably, find the overall behaviour of this function difficult to interpret on the basis of the above equation. For this particular question it is reasonably straightforward to code it up directly, which results in the contour plot shown in Fig. 2. The likelihood is peaked at the expected values (*i.e.*, $F_{\rm src} = C(N_{\rm on} - N_{\rm off}) = 6C$ and $F_{\rm bkg} = CN_{\rm off} = 3C$) but is extremely broad due to the very lower numbers of photons in the two measurements. As a result the steeply rising source prior dominates the posterior – one way of explaining this result is that there are so many ultra-faint sources with $F_{\rm src} \ll 1$ that it is more common for one of these to have been observed with a big background spike than it is for a "bright" source with $F_{\rm src} \simeq 6C$ to have been identified.

A a more general way of approaching such problems is to use a Monte Carlo technique instead – this rapidly becoms necessary for problems with more complicated likelihoods or more parameters.

(d) It is only the marginal posterior of the source flux, $\Pr(F_{\rm src}|N_{\rm on}, N_{\rm off})$, that is really of interest. To obtain this marginalized distribution, post-process the MCMC output by making a histogram of the $F_{\rm src}$ values, ignoring the $F_{\rm bkg}$ values.

Answer: To be done on computer