Imperial College
London

Course:	M5M S06 and M4 S18 B2
Setter:	Mortlock
Checker:	Calderhead
Editor:	Calderhead
External:	Griffin
Date:	April 21, 2017

MSc and MSci EXAMINATIONS (MATHEMATICS)

May 2017

M5M S06 and M4 S18 B2

Bayesian Data Analysis

Setter's signature	Checker's signature	Editor's signature
. . . .	. . . .	. . . .

Imperial College London

MSc and MSci EXAMINATIONS (MATHEMATICS)

May 2017

M5M S06 and M4 S18 B2

Bayesian Data Analysis

Date: Wednesday, May 3, 2017 Time: 10:00 – 11:30

This paper has 3 questions, worth 80 marks in total.

The marks for each question are given at the end of the question.

Calculators may not be used.

Statistical tables are not provided.

1. Consider two models, M_1 and M_2, which make predictions for a quantity θ according to

$$\Pr(\theta|M_1) = \delta_{\rm D}(\theta) \qquad \text{and} \qquad \Pr(\theta|M_2) = {\rm N}(\theta; 0, \sigma_2^2),$$

where $\delta_{\rm D}(.)$ is the Dirac delta function (i.e., model 1 is fully specified) and $\sigma_2 \geq 0$ is a fixed/known parameter of model 2. The models are to be assessed on the basis of a measurement x which is subject to observational uncertainty σ, so that its sampling distribution of the data is $\Pr(x|\theta) = {\rm N}(x; \theta, \sigma^2)$.

(i) Show that the Bayes factor in favour of model 1 is

$$B_{1,2} = \frac{\Pr(x|M_1)}{\Pr(x|M_2)} = \left(1 + \frac{\sigma_2^2}{\sigma^2}\right)^{1/2} \exp\left[-\frac{1}{2}\frac{\sigma_2^2\, x^2}{\sigma^2\,(\sigma^2 + \sigma_2^2)}\right].$$

(ii) In the below extreme cases find limiting expressions for $B_{1,2}$ and explain these results in conceptual/qualitative terms. (Assume that the other quantities are finite and non-zero in each case.)

a. $\sigma \to 0$.

b. $\sigma \to \infty$.

c. $\sigma_2 \to 0$.

d. $\sigma_2 \to \infty$.

e. $x \to 0$.

f. $x \to \infty$.

(iii) Assume that M_1 and M_2 are equally probable a priori, and that $\sigma = \sigma_2 = 1$. For what values of x would M_2 have a posterior probability of 0.9 or more?

(30 marks)

2. Post-Brexit, the UK changes its currency away from the current French-inspired decimal system, and introduces a binary system in which there are six new banknotes, worth £1, £2, £4, £8, £16 and £32. You appear on a game show where you are presented with two envelopes, labelled A and B but otherwise indistinguishable. You are told that each envelope contains one of a consecutive pair of the above banknotes, of value m_A and m_B respectively, implying that either $m_B = 2\,m_A$ or $m_B = m_A/2$. You are allowed to open one envelope, say A (without loss of generality), and so know m_A. You are then told that you can either keep this banknote or swap to B, in which case you can only keep the note in that envelope (i.e., you cannot change back).

(i) A previous contestant reasoned that it is a 50/50 chance that envelope B has the higher (or lower) value note, and so the expectated amount in envelope B is $0.5 \times (2\,m_A) + 0.5 \times (m_A/2) = 1.25\,m_A$. Th implication is that you should always swap and keep whatever note is in envelope B. Provide an assessment of this argument and, in particular, whether the result is reasonable.

You decide to take a Bayesian approach to the problem. You hence aim to infer the unknown quantity, m_B, conditional on the available information.

(ii) Provide a justification for adopting the prior that it is equally probable that the more valuable of the two notes is £2, £4, £8, £16 or £32. What is the implied joint prior distribution for m_A and m_B? Give your answer as a contingency table (i.e., a 6×6 array giving the probability of each combination of m_A and m_B).

(iii) Given the above prior, what is the implied posterior distribution of m_B given m_A? What would be your strategy therefore be for maximizing your return from the game?

(iv) Calculate the expected yield of applying the Bayesian strategy you obtained in part (iii). How does this compare to the "always swap" strategy advocated by the previous contestant in part (i)? Comment on the result.

(30 marks)

3. An undergraduate physics student asks you to help them analyse data from an experiment to measure the strength of a radioactive source in the presence of an unknown level of background radiation. The basic experiment consists of two measurements:

- a background-only calibration measurement (with the radioactive source locked away in a lead box) of duration $t_{\rm off}$ which yields $N_{\rm off}$ counts;
- an on-source measurement (with the source exposed and at a some known distance from the detector) of duration $t_{\rm on}$ which yields $N_{\rm on}$ counts, which is the sum of the separate contributions from both the source and the still-present background radiation.

You know that all the unstable atoms in the source have the same decay probability per unit time, and that the half-life of the source is much longer than the duration of the experiment. There are hence only two unknown parameters:

- $\Gamma_{\rm b}$, the rate (i.e., number per unit time) at which background particles hit the detector (the same in both measurements);
- $\Gamma_{\rm s}$, the rate at which particles from the source hit the detector (albeit only during the "on-source" measurement).

The physics student wants to place constraints on the two rates, and in particular $\Gamma_{\rm s}$.

(i) Given the above information, provide a justification for why the sampling distribution for the number of counts, N, from a single measurement of duration t can be written as

$$\Pr(N|\Gamma,t) = \Theta(N)\,\frac{(\Gamma\,t)^N\,\exp(-\Gamma\,t)}{N!},$$

where $\Gamma = \Gamma_{\rm b}$ if the source is locked away and $\Gamma = \Gamma_{\rm b} + \Gamma_{\rm s}$ if the source is exposed, and where $\Theta(x) = 0$ if $x < 0$ and $\Theta(x) = 1$ if $x \geq 0$ is the Heaviside step function.

(ii) The physics student had initially analysed their data by calculating an estimate of the background rate as $\widehat{\Gamma}_{\rm b} = N_{\rm off}/t_{\rm off}$ and then assuming that the true background rate is exactly equal to this value (i.e., $\Gamma_{\rm b} = N_{\rm off}/t_{\rm off}$). Assuming an improper prior on the source rate of the form $\Pr(\Gamma_{\rm s}|I) \propto \Theta(\Gamma_{\rm s})$, calculate the posterior distribution $\Pr(\Gamma_{\rm s}|N_{\rm on}, t_{\rm on}, \widehat{\Gamma}_{\rm b}, I)$ implied by the on-source measurement. Answers may be given in terms of the incomplete gamma function, defined here as $\gamma(t,x) = \int_x^\infty {\rm d}x'\, x'^{t-1}\, e^{-x'}$.

(iii) You inform the physics student that their approach above does not correctly incoroprate the uncertainty in the background rate, and so you set about developing a fully Bayesian calculation to obtain the marginalised posterior distribution $\Pr(\Gamma_{\rm s}|N_{\rm off}, N_{\rm on}, t_{\rm off}, t_{\rm on}, I)$. Outline the steps you would take to obtain this posterior distribution (but do *not* attempt the actual calculation). How would you expect this posterior distribution to compare to that obtained in part (ii)?

(20 marks)

Imperial College
London

Course: M5M S06 and M4 S18 B2
Setter: Mortlock
Checker: Calderhead
Editor: Calderhead
External: Griffin
Date: April 21, 2017

MSc and MSci EXAMINATIONS (MATHEMATICS)

May 2017

M5M S06 and M4 S18 B2

Bayesian Data Analysis

Setter's signature	Checker's signature	Editor's signature
....	....	....

Imperial College London

MSc and MSci EXAMINATIONS (MATHEMATICS)

May 2017

M5M S06 and M4 S18 B2

Bayesian Data Analysis

Date: Wednesday, May 3, 2017 Time: 10:00 – 11:30

This paper has 3 questions, worth 80 marks in total.

The marks for each question are given at the end of the question.

Calculators may not be used.

Statistical tables are not provided.

1. Consider two models, M_1 and M_2, which make predictions for a quantity θ according to

$$\Pr(\theta|M_1) = \delta_{\rm D}(\theta) \qquad \text{and} \qquad \Pr(\theta|M_2) = {\rm N}(\theta; 0, \sigma_2^2),$$

where $\delta_{\rm D}(.)$ is the Dirac delta function (i.e., model 1 is fully specified) and $\sigma_2 \geq 0$ is a fixed/known parameter of model 2. The models are to be assessed on the basis of a measurement x which is subject to observational uncertainty σ, so that the sampling distribution of the data is $\Pr(x|\theta) = {\rm N}(x; \theta, \sigma^2)$.

(i) Show that the Bayes factor in favour of model 1 is

$$B_{1,2} = \frac{\Pr(x|M_1)}{\Pr(x|M_2)} = \left(1 + \frac{\sigma_2^2}{\sigma^2}\right)^{1/2} \exp\left[-\frac{1}{2}\frac{\sigma_2^2\, x^2}{\sigma^2\,(\sigma^2 + \sigma_2^2)}\right].$$

Answer:

sim. seen ⇓

The (marginal) likelihood under model 1 is just

$$\begin{aligned}
\Pr(x|M_1) &= \int_{-\infty}^{\infty} {\rm d}\theta' \Pr(\theta'|M_1)\Pr(x|\theta', \sigma) \\
&= \int_{-\infty}^{\infty} {\rm d}\theta'\, \delta_{\rm D}(\theta')\, \frac{1}{(2\pi)^{1/2}\,\sigma} \exp\left[-\frac{1}{2}\frac{(x - \theta')^2}{\sigma^2}\right] \\
&= \frac{1}{(2\pi)^{1/2}\,\sigma} \exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right) \\
&= {\rm N}(x; 0, \sigma^2).
\end{aligned}$$

(This result could also have been written down directly.)

The marginal likelihood under model 2 is

$$\begin{aligned}
\Pr(x|M_2) &= \int_{-\infty}^{\infty} {\rm d}\theta' \Pr(\theta'|M_2)\Pr(x|\theta', \sigma) \\
&= \int_{-\infty}^{\infty} {\rm d}\theta'\, \frac{1}{(2\pi)^{1/2}\,\sigma_2} \exp\left(-\frac{1}{2}\frac{\theta'^2}{\sigma_2^2}\right) \frac{1}{(2\pi)^{1/2}\,\sigma} \exp\left[-\frac{1}{2}\frac{(x - \theta')^2}{\sigma^2}\right] \\
&= \frac{1}{2\pi\,\sigma\,\sigma_2} \exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2 + \sigma_2^2}\right) \int_{-\infty}^{\infty} {\rm d}\theta' \exp\left[-\frac{1}{2}\frac{\theta'^2 - x/(1 + \sigma^2/\sigma_2^2)}{\sigma^2\,\sigma_2^2/(\sigma^2 + \sigma_2^2)}\right] \\
&= \frac{1}{(2\pi)^{1/2}\,(\sigma^2 + \sigma_2^2)^{1/2}} \exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2 + \sigma_2^2}\right) \\
&= {\rm N}\left(x; 0, \sigma^2 + \sigma_2^2\right).
\end{aligned}$$

(This could also be obtained on more general grounds as the sampling distribution under M_2 is just the convolution of two normal distributions of variance σ_2^2 and σ^2, and so is itself a normal distribution of variance $\sigma^2 + \sigma_2^2$.)

Taking the ratio of the above marginal likelihoods gives the Bayes factor as

$$\begin{aligned}
B_{1,2} &= \frac{\Pr(x|M_1)}{\Pr(x|M_2)} \\
&= \frac{(2\pi)^{1/2}\,(\sigma^2 + \sigma_2^2)^{1/2}}{(2\pi)^{1/2}\,\sigma} \exp\left[-\frac{1}{2}\frac{x^2}{\sigma^2} + \frac{1}{2}\frac{x^2}{\sigma^2 + \sigma_2^2}\right] \\
&= \left(1 + \frac{\sigma_2^2}{\sigma^2}\right)^{1/2} \exp\left[-\frac{1}{2}\frac{\sigma_2^2\, x^2}{\sigma^2\,(\sigma^2 + \sigma_2^2)}\right],
\end{aligned}$$

which is the required result.

(ii) In the below extreme cases find limiting expressions for $B_{1,2}$ and explain these results in conceptual/qualitative terms. (Assume that the other quantities are finite and non-zero in each case.)

a. $\sigma \to 0$.

Answer:

sim. seen ⇓

The required limit is $\lim_{\sigma \to 0} B_{1,2} = 0$.

In this case the measurement is perfect, so (given that $x \neq 0$) model 1 is completely contradicted by the data, leading to a Bayes factor of 0. [If $x = 0$ then model 1's (infinitely) greater predictivity would mean it is (infinitely) favoured.]

b. $\sigma \to \infty$.

Answer:

sim. seen ⇓

The required limit is $\lim_{\sigma \to \infty} B_{1,2} = 1$.

In this limit the measurement error is enormous, implying that the data contains no useful information. There can be no reason (other than model priors, which do not enter into the Bayes factor) to favour one model over the other, so the Bayes factor is 1.

c. $\sigma_2 \to 0$.

Answer:

sim. seen ⇓

The required limit is $\lim_{\sigma_2 \to 0} B_{1,2} = 1$.

In this limit model 2 is the same as model 1: both predict $\theta = 0$. As the models are hence identical there can be no reason (other than model priors, which do not enter into the Bayes factor) to favour one model over the other so the Bayes factor is 1.

d. $\sigma_2 \to \infty$.

Answer:

sim. seen ⇓

The required limit is $\lim_{\sigma_2 \to \infty} B_{1,2} = \infty$.

In this limit model 2 is so unpredictive that, for any finite measurement x, model 1 will better explain the data (even if $|x| \gg 0$), so the Bayes factor in favour of model 1 is infinite.

e. $x \to 0$.

Answer:

sim. seen ⇓

The required limit is $\lim_{x \to 0} B_{1,2} = (1 + \sigma_2^2/\sigma^2)^{1/2}$.

In this limit the data matches the prediction of both models well: $x = 0$ is the mode of both (marginal) sampling distributions. But, as model 1 is more predictive, the Bayes factor is > 1 (by a factor which depends on the ratio of σ and σ_2).

f. $x \to \infty$.

Answer:

sim. seen ⇓

The required limit is $\lim_{x \to \infty} B_{1,2} = 0$.

Both models predict the data very poorly in this case, but model 2, with its broader range of θ values does better, and inifinitely so in this limit, so the Bayes factor in favour of model 1 is 0.

(iii) Assume that M_1 and M_2 are equally probable a priori, and that $\sigma = \sigma_2 = 1$. For what values of x would M_2 have a posterior probability of 0.9 or more?

Answer:

sim. seen ⇓

Bayes's theorem (and the law of total probability) gives

$$\begin{aligned}\Pr(M_2|x, I) &= \frac{\Pr(M_2|I)\Pr(x|M_2)}{\Pr(M_1|I)\Pr(x|M_1) + \Pr(M_2|I)\Pr(x|M_2)} \\ &= \frac{1}{\Pr(x|M_1)/\Pr(x|M_2) + 1} \\ &= \frac{1}{B_{1,2} + 1},\end{aligned}$$

where the prior information that $\Pr(M_1|I) = \Pr(M_2|I) = 1/2$ has been applied, and $B_{1,2}$ is the Bayes factor as defined in the question. Inverting this and inserting the critical posterior probability of 0.9 then gives

$$B_{1,2} = \frac{1}{0.9} - 1 = \frac{1}{9}.$$

The expression for the Bayes factor given in part (i) Taking $\sigma = \sigma_2 = 1$, the Bayes factor defined in part (i) simplifies to be

$$B_{1,2} = 2^{1/2} \exp\left(-\frac{x^2}{4}\right).$$

Inverting this yields

$$x = \pm 2 \left[\ln\left(9 \times 2^{1/2}\right)\right]^{1/2},$$

which can, of course, be written in several different ways.

The posterior probability of M_2 will be 0.9 or higher if $|x| \geq 2[ln(9 \times 2^{1/2})]^{1/2}$..

(30 marks)

2. Post-Brexit, the UK changes its currency away from the current French-inspired decimal system, and introduces a binary system in which there are six new banknotes, worth £1, £2, £4, £8, £16 and £32. You appear on a game show where you are presented with two envelopes, labelled A and B but otherwise indistinguishable. You are told that each envelope contains one of a consecutive pair of the above banknotes, of value m_A and m_B respectively, implying that either $m_B = 2\,m_A$ or $m_B = m_A/2$. You are allowed to open one envelope, say A (without loss of generality), and so know m_A. You are then told that you can either keep this banknote or swap to B, in which case you can only keep the note in that envelope (i.e., you cannot change back).

(i) A previous contestant reasoned that it is a 50/50 chance that envelope B has the higher (or lower) value note, and so the expectated amount in envelope B is $0.5 \times (2\,m_A) + 0.5 \times (m_A/2) = 1.25\,m_A$. Th implication is that you should always swap and keep whatever note is in envelope B. Provide an assessment of this argument and, in particular, whether the result is reasonable.

Answer: unseen ⇓

The assertion that envelope B is equally likely to have double or half the amount in envelope A is a reasonable assignment of the a priori probability (i.e., before any envelope is opened). But once envelope A is opened, the value of m_A could immediately be in contradiction with this 50/50 assignment, most obviously if $m_A = £1$ (in which case m_B must be $£2$) or if $m_A = £32$ (in which case m_B must be $£16$).

Still, ignoring information is not in itself an illegitimate option; the real problem is that this new information (i.e., the value m_A) has been included in an incomplete way, as the probabilities have not been updated. (One way to avoid making such a mistake is to rigorously apply the rules of Bayesian inference.)

The result – that picking an envelope at random and then swapping to the other envelope, which is hence also effectively picked at random, will increase the pay-off by 25% – is hence not reasonable.

You decide to take a Bayesian approach to the problem. You hence aim to infer the unknown quantity, m_B, conditional on the available information.

(ii) Provide a justification for adopting the prior that it is equally probable that the more valuable of the two notes is £2, £4, £8, £16 or £32. What is the implied joint prior distribution for m_A and m_B? Give your answer as a contingency table (i.e., a 6×6 array giving the probability of each combination of m_A and m_B).

Answer: seen ⇓

A prior distribution for the value of the more valuable note can be defined by the probabilities p_2, p_4, p_8, p_{16} and p_{32}. Any set of values would be admissable, provided only that they are all non-negative and sum to 1. With no further information available there is no reason to distinguish any one of the values, so interchangeability or indifference implies assigning

$p_2 = p_4 = p_8 = p_{16} = p_{32} = 1/5$. An alternative, equally valid argument would be that this is the maximum entropy distribution subject to the available information.

The prior $\Pr(m_A, m_B|I)$, where I is encodes the assumption of a uniform distribution of the more valuable note, can be represented by a contingency table as

$\Pr(m_A, m_B\|I)$		m_B					
		£1	£2	£4	£8	£16	£32
	£1	0	0.1	0	0	0	0
	£2	0.1	0	0.1	0	0	0
m_A	£4	0	0.1	0	0.1	0	0
	£8	0	0	0.1	0	0.1	0
	£16	0	0	0	0.1	0	0.1
	£32	0	0	0	0	0.1	0

Note that summing up each of the five pairs of cross diagonal elements gives a probability of 0.2 associated with each of the five possible values of the more valuable note.

(iii) Given the above prior, what is the implied posterior distribution of m_B given m_A? What would be your strategy therefore be for maximizing your return from the game?

Answer:

unseen ⇓

There are several different ways to formulate this problem. One option is to utilise Kronecker delta functions, as was done extensively in the course, but in this case the small size of the parameter space makes contingency tables the clearer option, (although any method of solution is acceptable).

The data, i.e., the observed value of m_A, modifies the prior distribution from part (ii) by selecting whichever row corresponds to the observed value of m_A. Given that the posterior must be normalised for each value of m_A (i.e., across each row), the posterior distribution can be represented as

$\Pr(m_B\|m_A, I)$	m_B					
	£1	£2	£4	£8	£16	£32
$m_A = £1$	0	1.0	0	0	0	0
$m_A = £2$	0.5	0	0.5	0	0	0
$m_A = £4$	0	0.5	0	0.5	0	0
$m_A = £8$	0	0	0.5	0	0.5	0
$m_A = £16$	0	0	0	0.5	0	0.5
$m_A = £32$	0	0	0	0	1.0	0

If $m_A = £1$ then envelope B must have the higher amount; if $m_A = £32$ then envelope B must have the lower amount; for all other values of m_A envelope B is equally likely to have the lower or higher amount.

Hence the optimal strategey is to swap to envelope B, except in the case that $m_A = £32$, in which case this should be retained.

(iv) Calculate the expected yield of applying the Bayesian strategy you obtained in part (iii). How does this compare to the "always swap" strategy advocated by the previous contestant in part (i)? Comment on the result.

Answer:

unseen ⇓

If the strategy from part (iii) were followed then the yield for each of the possible situations would be

m_{kept}		m_B					
		£1	£2	£4	£8	£16	£32
	£1		£2				
	£2	£1		£4			
m_A	£4		£2		£8		
	£8			£4		£16	
	£16				£8		£32
	£32					£32	

Averaging these ten values (since each scenario was equally likely a priori) then gives $\mathrm{E}(m_{\text{kept}}|m_A, I) = £10.90$.

If the strategy from part (ii) were followed then the equivalent table would be

m_{kept}		m_B					
		£1	£2	£4	£8	£16	£32
	£1		£2				
	£2	£1		£4			
m_A	£4		£2		£8		
	£8			£4		£16	
	£16				£8		£32
	£32					£16	

Averaging these values gives $\mathrm{E}(m_{\text{kept}}|m_A, \text{always swap}) = £9.30$.

This is less than expected using the Bayesian strategy as the information made available by finding out the value of m_A was not utilised.

(30 marks)

3. An undergraduate physics student asks you to help them analyse data from an experiment to measure the strength of a radioactive source in the presence of an unknown level of background radiation. The basic experiment consists of two measurements:

- a background-only calibration measurement (with the radioactive source locked away in a lead box) of duration $t_{\rm off}$ which yields $N_{\rm off}$ counts;
- an on-source measurement (with the source exposed and at a some known distance from the detector) of duration $t_{\rm on}$ which yields $N_{\rm on}$ counts, which is the sum of the separate contributions from both the source and the still-present background radiation.

You know that all the unstable atoms in the source have the same decay probability per unit time, and that the half-life of the source is much longer than the duration of the experiment. There are hence only two unknown parameters:

- $\Gamma_{\rm b}$, the rate (i.e., number per unit time) at which background particles hit the detector (the same in both measurements);
- $\Gamma_{\rm s}$, the rate at which particles from the source hit the detector (albeit only during the "on-source" measurement).

The physics student wants to place constraints on the two rates, and in particular $\Gamma_{\rm s}$.

(i) Given the above information, provide a justification for why the sampling distribution for the number of counts, N, from a single measurement of duration t can be written as

$$\Pr(N|\Gamma, t) = \Theta(N) \, \frac{(\Gamma \, t)^N \, \exp(-\Gamma \, t)}{N!},$$

where $\Gamma = \Gamma_{\rm b}$ if the source is locked away and $\Gamma = \Gamma_{\rm b} + \Gamma_{\rm s}$ if the source is exposed, and where $\Theta(x) = 0$ if $x < 0$ and $\Theta(x) = 1$ if $x \geq 0$ is the Heaviside step function.

Answer: sim. seen ♭

Given the independence of the decay events and that their rate is constant (at least for the duration of the measurements), the implication is that the number of recorded decays, N, can be treated as the output of a Poisson process with some mean λ. If λ was known then the sampling distribution would be

$$\Pr(N|\lambda) = \Theta(N) \, \frac{\lambda^N \, \exp(-\lambda)}{N!}.$$

Given that the rates here are defined at the detector, the mean number of expected decays during an observing time t would be $\Gamma \, t$. Identifying this as λ then implies that

$$\Pr(N|\Gamma, t) = \Theta(N) \, \frac{(\Gamma \, t)^N \, \exp(-\Gamma \, t)}{N!},$$

as required.

(ii) The physics student had initially analysed their data by calculating an estimate of the background rate as $\widehat{\Gamma}_\mathrm{b} = N_\mathrm{off}/t_\mathrm{off}$ and then assuming that the true background rate is exactly equal to this value (i.e., $\Gamma_\mathrm{b} = N_\mathrm{off}/t_\mathrm{off}$). Assuming an improper prior on the source rate of the form $\mathrm{Pr}(\Gamma_\mathrm{s}|I) \propto \Theta(\Gamma_\mathrm{s})$, calculate the posterior distribution $\mathrm{Pr}(\Gamma_\mathrm{s}|N_\mathrm{on}, t_\mathrm{on}, \widehat{\Gamma}_\mathrm{b}, I)$ implied by the on-source measurement. Answers may be given in terms of the incomplete gamma function, defined here as $\gamma(t,x) = \int_x^\infty \mathrm{d}x'\, x'^{t-1}\, e^{-x'}$.

Answer:

sim. seen ⇓

With the background assumed to be $\widehat{\Gamma}_\mathrm{b}$, there is only one unknown parameter, Γ_s, and only the on-source measurement contains any information about it. For the on-source measurement the likelihood is, from part (i),

$$\mathrm{Pr}(N_\mathrm{on}|\widehat{\Gamma}_\mathrm{b}, \Gamma_\mathrm{s}, t_\mathrm{on}) = \Theta(N_\mathrm{on})\, \frac{[(\widehat{\Gamma}_\mathrm{b} + \Gamma_\mathrm{s})\, t_\mathrm{on}]^{N_\mathrm{on}} \exp[-(\widehat{\Gamma}_\mathrm{b} + \Gamma_\mathrm{s})\, t_\mathrm{on}]}{N_\mathrm{on}!},$$

where the estimated value has been substituted for the true background rate. The implied posterior is then obtained from Bayes's theorem (and the law of total probability) as

$$\begin{aligned}
\mathrm{Pr}(\Gamma_\mathrm{s}|N_\mathrm{on}, t_\mathrm{on}, \widehat{\Gamma}_\mathrm{b}, I) &= \frac{\mathrm{Pr}(\Gamma_\mathrm{s}|I)\,\mathrm{Pr}(N_\mathrm{on}|\Gamma_\mathrm{s}, t_\mathrm{on}, \widehat{\Gamma}_\mathrm{b})}{\int_0^\infty \mathrm{d}\Gamma_\mathrm{s}\,\mathrm{Pr}(\Gamma_\mathrm{s}|I)\,\mathrm{Pr}(N_\mathrm{on}|\Gamma_\mathrm{s}, t_\mathrm{on}, \widehat{\Gamma}_\mathrm{b})} \\
&= \Theta(\Gamma_\mathrm{s})\, \frac{[(\Gamma_\mathrm{s} + \widehat{\Gamma}_\mathrm{b})\, t_\mathrm{on}]^{N_\mathrm{on}} \exp[-(\Gamma_\mathrm{s} + \widehat{\Gamma}_\mathrm{b})\, t_\mathrm{on}]}{\int_0^\infty \mathrm{d}\Gamma_\mathrm{s}\, [(\Gamma_\mathrm{s} + \widehat{\Gamma}_\mathrm{b})\, t_\mathrm{on}]^{N_\mathrm{on}} \exp[-(\Gamma_\mathrm{s} + \widehat{\Gamma}_\mathrm{b})\, t_\mathrm{on}]} \\
&= \Theta(\Gamma_\mathrm{s})\, \frac{t_\mathrm{on}\, [(\Gamma_\mathrm{s} + \widehat{\Gamma}_\mathrm{b})\, t_\mathrm{on}]^{N_\mathrm{on}} \exp[-(\Gamma_\mathrm{s} + \widehat{\Gamma}_\mathrm{b})\, t_\mathrm{on}]}{\int_{\widehat{\Gamma}_\mathrm{b} t_\mathrm{on}}^\infty \mathrm{d}x\, x^{N_\mathrm{on}}\, e^{-x}} \\
&= \Theta(\Gamma_\mathrm{s})\, \frac{t_\mathrm{on}\, [(\Gamma_\mathrm{s} + \widehat{\Gamma}_\mathrm{b})\, t_\mathrm{on}]^{N_\mathrm{on}} \exp[-(\Gamma_\mathrm{s} + \widehat{\Gamma}_\mathrm{b})\, t_\mathrm{on}]}{\gamma(N_\mathrm{on} + 1, \widehat{\Gamma}_\mathrm{b}\, t)},
\end{aligned}$$

where $\gamma(x,t)$ is the incomplete gamma function, as defined in the question. (Another option for simplifying the integral is to use the binomial formala, in which case the normalisation constant becomes a finite sum.)

(iii) You inform the physics student that their approach above does not correctly incoroprate the uncertainty in the background rate, and so you set about developing a fully Bayesian calculation to obtain the marginalised posterior distribution $\mathrm{Pr}(\Gamma_\mathrm{s}|N_\mathrm{off}, N_\mathrm{on}, t_\mathrm{off}, t_\mathrm{on}, I)$. Outline the steps you would take to obtain this posterior distribution (but do *not* attempt the actual calculation). How would you expect this posterior distribution to compare to that obtained in part (ii)?

Answer:

sim. seen ⇓

The desired marginal posterior distribution could be obtained by

1. Set a prior distribution on the background and source rates, for which the obvious form (implied by part ii) is $\mathrm{Pr}(\Gamma_\mathrm{b}, \Gamma_\mathrm{s}|I) \propto \Theta(\Gamma_\mathrm{b})\,\Theta(\Gamma_\mathrm{s})$.
2. Construct the joint likelihood by exploiting the independence of the two observations, which implies that $\mathrm{Pr}(N_\mathrm{off}, N_\mathrm{on}|\Gamma_\mathrm{b}, \Gamma_\mathrm{s}, t_\mathrm{off}, t_\mathrm{on}) = \mathrm{Pr}(N_\mathrm{off}|\Gamma_\mathrm{b}, t_\mathrm{off})\,\mathrm{Pr}(N_\mathrm{on}|\Gamma_\mathrm{b}, \Gamma_\mathrm{s}, t_\mathrm{on})$, which is then a product of two Poisson distributions of the form given in part (i).

3. Calculate the (not necessarily normalised) joint posterior distribution $\Pr(\Gamma_b, \Gamma_s | N_{\rm off}, N_{\rm on}, t_{\rm off}, t_{\rm on}, I) \propto \Pr(\Gamma_b, \Gamma_s | I) \Pr(N_{\rm off} | \Gamma_b, t_{\rm off}) \Pr(N_{\rm on} | \Gamma_b, \Gamma_s, t_{\rm on})$.
4. Marginalise out the background rate to obtain the desired posterior according to $\Pr(\Gamma_s | N_{\rm off}, N_{\rm on}, t_{\rm off}, t_{\rm on}, I) = \int_0^\infty {\rm d}\Gamma_b \Pr(\Gamma_b, \Gamma_s | N_{\rm off}, N_{\rm on}, t_{\rm off}, t_{\rm on}, I)$.

Some mention must be made of normalisation, although in this case it is sufficient to leave the joint posterior improper and normalise only the marginal posterior after the last step given explicitly above.

The extra source odf uncertainty that comes from imperfect knowledge of the background rate means the full posterior $\Pr(\Gamma_s | N_{\rm off}, N_{\rm on}, t_{\rm off}, t_{\rm on}, I)$ will be broader than the psuedo-posterior $\Pr(\Gamma_s | N_{\rm on}, t_{\rm on}, \widehat{\Gamma}_b, I)$ obtained in part (ii). While a less precise answer might seem desireable, it is in fact a more accurate result that correctly encodes the available information and nothing more.

(20 marks)