ADA IX Summer School Tutorial: Exercices

Roberto Trotta, Imperial College London

May 2018

Contents

1 Bayesian Reasoning

A cohort chemistry undergraduates are screened for a dangerous medical condition called *Bacillum Bayesianum* (BB). The incidence of the condition in the population (i.e., the probability that a randomly selected person has the disease) is estimated at about 1%. If the person has BB, the test returns positive 95% of the time. There is also a known 5% rate of false positives, i.e. the test returning positive even if the person is free from BB. One of your friends takes the test and it comes back positive. Here we examine whether your friend should be worried about her health.

(i) Translate the information above in suitably defined probabilities. The two relevant propositions here are whether the test returns positive (denote this with a $+$ symbol) and whether the person is actually sick (denote this with the symbol $BB = 1$. Denote the case when the person is healthy as $BB = 0$).

(ii) Compute the conditional probability that your friend is sick, knowing that she has tested positive, i.e., find $P(BB = 1|+)$.

(iii) Imagine screening the general population for a very rare disease, whose incidence in the population is 10^{-6} (i.e., one person in a million has the disease on average, i.e. $P(BB = 1) = 10^{-6}$). What should the reliability of the test (i.e., $P(+|BB = 1)$) be if we want to make sure that the probability of actually having the disease after testing positive is at least 99%? Assume first that the false positive rate $P(+|BB = 0)$ (i.e, the probability of testing positive while healthy), is 5% as in part (a). What can you conclude about the feasibility of such a test?

2 Bayesian Parameter Estimation

2.1 Coin Tossing: Binomial Distribution

A coin is tossed N times and heads come up H times.

(i) What is the likelihood function? Identify clearly the parameter, θ, and the data.

(ii) What is a reasonable, non-informative prior on θ?

(iii) Compute the posterior probability for θ. Recall that θ is the probability that a single flip will give heads. This integral will prove useful:

$$\int_0^1 d\theta \theta^N (1-\theta)^M = \frac{\Gamma(N+1)\Gamma(M+1)}{\Gamma(N+M+2)}. \tag{1}$$

(iv) Determine the posterior mean and standard deviation of θ.

(v) Plot your results as a function of H for $N = 10, 100, 1000$.

2.2 The Gaussian Linear Model

This problem takes you through the steps to derive the posterior distribution for a quantity of interest θ, in the case of a Gaussian prior and Gaussian likelihood, for the 1-dimensional case.

2.2.1 Theory

Let us assume that we have made N independent measurements, $\hat{x} = \{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_N\}$ of a quantity of interest θ (this could be the temperature of an object, the distance of a galaxy, the mass of a planet, etc). We assume that each of the measurements in independently Gaussian distributed with known experimental standard deviation σ. Let us denote the sample mean by $\bar{x}$, i.e.

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} \hat{x}_i. \tag{2}$$

Before we do the experiment, our state of knowledge about the quantity of interest θ is described by a Gaussian distribution on θ, centered around 0 (we can always choose the units in such a way that this is the case). Such a prior might come e.g. from a previous experiment we have performed. The new experiment is however much more precise, i.e. $\Sigma \gg \sigma$. Our prior state of knowledge be written in mathematical form as the following Gaussian pdf:

$$p(\theta) \sim \mathcal{N}(0, \Sigma^2). \tag{3}$$

(i) Write down the likelihood function for the measurements and show that it can be recast in the form:

$$\mathcal{L}(\theta) = L_0 \exp\left(-\frac{1}{2}\frac{(\theta - \bar{x})^2}{\sigma^2/N}\right), \tag{4}$$

where L_0 is a constant that does not depend on θ.

(ii) By using Bayes theorem, compute the posterior probability for θ after the data have been taken into account, i.e. compute $p(\theta|\hat{x})$. Show that it is given by a Gaussian of mean $\bar{x}\frac{\Sigma^2}{\Sigma^2+\sigma^2/N}$ and variance $\left[\frac{1}{\Sigma^2} + \frac{N}{\sigma^2}\right]^{-1}$.
Hint: you may drop the normalization constant from Bayes theorem, as it does not depend on θ

(iii) Show that as $N \to \infty$ the posterior distribution becomes independent of the prior.

(iv) Show that as $N \to \infty$ the mean of the posterior distribution converges to the MLE of the mean for θ. This means that for a large number of measurements, the Bayesian result matches the frequentist MLE result.

2.2.2 Two-dimensional Example

Now we specialize to the case $n = 2$, i.e. we have two parameters of interest, $\theta = \{\theta_1, \theta_2\}$ and the linear function we want to fit is given by

$$y = \theta_1 + \theta_2 x. \tag{5}$$

(In the formalism above, the basis vectors are $X^1 = 1, X^2 = x$).

Table 1: Data sets for the Gaussian linear model exercise. You may assume that all data points are independently and identically distributed with standard deviation of the noise $\sigma = 0.1$.

x	y
0.8308	0.9160
0.5853	0.7958
0.5497	0.8219
0.9172	1.3757
0.2858	0.4191
0.7572	0.9759
0.7537	0.9455
0.3804	0.3871
0.5678	0.7239
0.0759	0.0964

Table 1 gives an array of $d = 10$ measurements $y = \{y_1, y_2, \ldots, y_{10}\}$, together with the values of the independent variable x_i. Assume that the uncertainty in the same for all measurements, i.e. $\tau_i = 0.1$ ($i = 1, \ldots, 10$). You may further assume that measurements are uncorrelated. The data set is shown in the left panel of Fig. 1

(i) Assume a Gaussian prior with Fisher matrix $P = \text{diag}\left(10^{-2}, 10^{-2}\right)$ for θ.

Find the posterior distribution for θ given the data, and plot it in 2 dimensions in the (θ_1, θ_2) plane (see right panel of Fig. 1).

Use the appropriate contour levels to demarcate 1, 2 and 3 sigma joint credible intervals of the posterior.

(ii) In a language of your choice, write an implementation of the Metropolis-Hastings Markov Chain Monte Carlo algorithm, and use it to obtain samples from the posterior distribution.

(iii) If you are already familiar with Metropolis-Hastings, write an implementation of Hamiltonian Monte Carlo instead.

(iv) Plot *equal weight* samples in the (θ_1, θ_2) space, as well as marginalized 1-dimensional posterior distributions for each parameter.

(v) Compare the credible intervals that you obtained from the MCMC with the analytical solution.

2.3 Poisson counts

2.3.1 Maximum Likelihood Approach

An astronomer measures the photon flux from a distant star using a very sensitive instrument that counts single photons. After one minute of observation, the

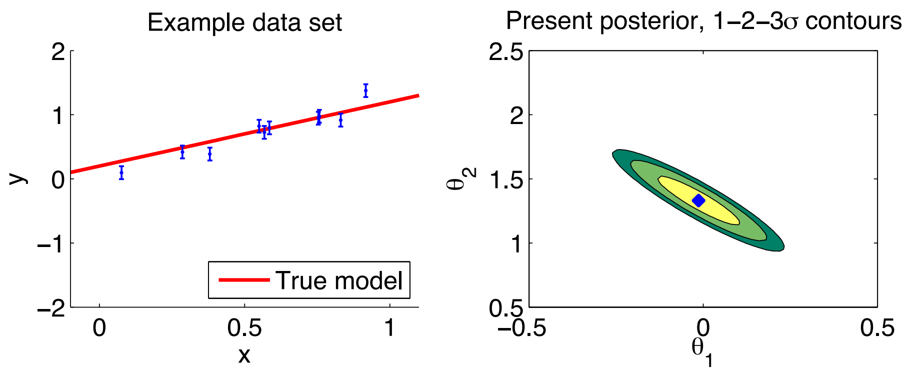

Figure 1: Left panel: data set for the Gaussian linear problem. The solid line shows the true value of the linear model from which the data have been generated, subject to Gaussian noise. Right panel: 2D credible intervals from the posterior distribution for the parameters. The the blue diamond is the Maximum Likelihood Estimator, whose value for this data set is $x = -0.0136, y = 1.3312$.

instrument has collected $\hat{r}$ photons. One can assume that the photon counts, $\hat{r}$, are distributed according to the Poisson distribution. The astronomer wishes to determine λ, the emission rate of the source.

(i) What is the likelihood function for the measurement? Identify explicitly what is the unknown parameter and what are the data in the problem.

(ii) If the true rate is $\lambda = 10$ photons/minute, what is the probability of observing $\hat{r} = 15$ photons in one minute?

(iii) Find the Maximum Likelihood Estimate for the rate λ (i.e., the number of photons per minute). What is the maximum likelihood estimate if the observed number of photons is $\hat{r} = 10$?

2.3.2 The On/Off Problem

Upon reflection, the astronomer realizes that the photon flux is the superposition of photons coming from the star plus "background" photons coming from other faint sources within the field of view of the instrument. The background rate is supposed to be known, and it is given by λ_b photons per minute. This can be estimated e.g. by pointing the telescope away from the source (the "off" measurement) and measuring the photon counts there, where the telescope is only picking up background photons. This estimate of the background comes with an uncertainty, of course, but we'll ignore this for now. She then points to the star again, measuring $\hat{r}_t$ photons in a time t_t (this is the "on" measurement).

(i) What is her maximum likelihood estimate of the rate λ_s from the star in this case? *Hint:* The total number of photons $\hat{r}_t$ is Poisson distributed with rate $\lambda = \lambda_s + \lambda_b$, where λ_s is the rate for the star.

(ii) What is the source rate (i.e., the rate for the star) if $\hat{r}_t = 30$, $t_t = 2$ mins, and $\lambda_b = 12$ photons per minute?

(iii) Is it possible that the measured average rate from the source (i.e., $\hat{r}_t/t_t$) is less than λ_b? Discuss what happens in this case and comment on the physicality of this result.

2.3.3 On/Off Problem: Bayesian version

We revisit the On/Off problem but this time from a Bayesian perspective, which fully and automatically accounts for uncertainty in the background rate estimate.

We consider first the "off" measurement, which collects $n_{\rm off}$ photons in a time $t_{\rm off}$.

(i) Assuming a uniform prior on the background rate b, find the posterior distribution for b from the off measurement.

(ii) Now consider the "on" measurement, which collects a number $n_{\rm on}$ of photons during a time $t_{\rm on}$. This is a measurement for the combined rate $s+b$ (where s denotes the source rate). Write down the likelihood function for this measurement.

(iii) Assume again a uniform prior on s, and a prior on b given by the posterior of the "off" measurement[1], find the (unnormalized) joint posterior distribution for s, b, and show that is is given by the expression:

$$p(s, b|n_{\rm on}, t_{\rm on}) \propto (s+b)^{n_{\rm on}} b^{n_{\rm off}} \exp(-st_{\rm on}) \exp(-b(t_{\rm on}+t_{\rm off})) \text{ for } s, b \geq 0. \tag{6}$$

(iv) Compute analytically the marginal posterior pdf for the signal, s, by integrating the joint posterior over b, i.e.

$$p(s|n_{\rm on}, t_{\rm on}) = \int_0^\infty p(s, b|n_{\rm on}, t_{\rm on}) db. \tag{7}$$

. Plot the resulting marginal distribution for the signal s for the following two cases, and compare the result with the MLE result:

(a) $n_{\rm on} = 5, t_{\rm on} = 1, n_{\rm off} = 2, t_{\rm off} = 1$

(b) $n_{\rm on} = 2, t_{\rm on} = 2, n_{\rm off} = 3, t_{\rm off} = 1$

[1] The posterior for the "off" measurement can be used as prior on b for the "on" measurement. Alternatively, you can write down the joint posterior on s, b conditional on both measurements, with an ur-prior on b that is just the uniform prior (i.e., the prior that you used for the "off" measurement). Both procedures will give the same result, as they should (consistency of Bayesian reasoning is always in-built). Convince yourself that this is indeed the case!

(c) $n_{\text{on}} = 8, t_{\text{on}} = 1, n_{\text{off}} = 2, t_{\text{off}} = 4$

Hint: use the binomial expansion: $(s+b)^{n_{\text{on}}} = \sum_{k=0}^{n_{\text{on}}} \binom{n_{\text{on}}}{k} s^{n_{\text{on}}-k} b^k$.

(v) Write a code to perform MCMC sampling of the joint posterior for s, b (in Python you may want to use the `PyMC` package). Plot equal-weight samples from the posterior in parameter space for $n_{\text{on}} = 10, t_{\text{on}} = 2, n_{\text{off}} = 3, t_{\text{off}} = 1$. Marginalize over b numerically and compare the resulting numerical estimate with the analytical result above.

3 Toy Cosmological Parameter Inference (Harder)

Supernovae type Ia can be used as standardizable candles to measure distances in the Universe. This series of problems explores the extraction of cosmological information from a simplified SNIa toy model.

The cosmological parameters we are interested in constraining are

$$\mathscr{C} = \{\Omega_m, \Omega_\Lambda, h\} \tag{8}$$

where Ω_m is the matter density (in units of the critical energy density) and Ω_Λ is the dark energy density, assumed here to be in the form of a cosmological constant, i.e. $w = -1$ at all redshifts. In the following, we will fix $h = 0.72$ for simplicity, where the Hubble constant today is given by $H_0 = 100h$km/s/Mpc.

In an FRW cosmology defined by the parameters $\mathscr{C}$, the distance modulus μ (i.e., the difference between the apparent and absolute magnitudes, $\mu = m - M$) to a SN at redshift z is given by

$$\mu(z, \mathscr{C}) = 5 \log \left[\frac{D_L(z, \Omega_m, \Omega_\Lambda, h)}{\text{Mpc}} \right] + 25, \tag{9}$$

where D_L denotes the luminosity distance to the SN. Recalling that $D_L = c d_L / H_0$, We can rewrite this as

$$\mu(z, \mathscr{C}) = \eta + 5 \log d_L(z, \Omega_m, \Omega_\Lambda), \tag{10}$$

where

$$\eta = -5 \log \frac{100h}{c} + 25 \tag{11}$$

and c is the speed of light in km/s. We have defined the dimensionless luminosity distance

$$d_L(z, \Omega_m, \Omega_\Lambda) = \frac{(1+z)}{\sqrt{|\Omega_\kappa|}} \text{sinn}\{ \sqrt{|\Omega_\kappa|} \int_0^z \mathrm{d}z' [(1+z')^3 \Omega_m + \Omega_\Lambda + (1+z')^2 \Omega_\kappa]^{-1/2} \}. \tag{12}$$

The curvature parameter is given by the constraint equation

$$\Omega_\kappa = 1 - \Omega_m - \Omega_\Lambda \tag{13}$$

and the function

$$\mathrm{sinn}(x) = \begin{cases} x & \text{for a flat Universe } (\Omega_\kappa = 0); \\ \sin(x) & \text{for a closed Universe } (\Omega_\kappa < 0); \\ \sinh(x) & \text{for an open Universe } (\Omega_\kappa > 0). \end{cases} \tag{14}$$

We now assume that from each SNIa in our sample we get a measurement of the distance modulus with Gaussian noise[2], i.e., that the likelihood function for each SN i ($i = 1, \dots, N$) is of the form

$$\mathcal{L}_i(z_i, \mathscr{C}, M) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{1}{2}\frac{(\hat{\mu}_i - \mu(z_i, \mathscr{C}))^2}{\sigma_i^2}\right). \tag{15}$$

The observed distance modulus is given by $\hat{\mu}_i = \hat{m}_i - M$, where $\hat{m}_i$ is the observed apparent magnitude and M is the intrinsic magnitude of the SNIa. We assume that each SN observation is independent of all the others.

The provided data file[3] (`SNe_simulated.dat`) contains simulated observations from the above simplified model of $N = 300$ SNIa. The two columns give the redshift z_i and the observed apparent magnitude $\hat{m}_i$. The observational error is the same for all SNe, $\sigma_i = \sigma = 0.4$ mag for $i = 1, \dots, N$.

A plot of the data set is shown in the left panel of Fig. 2. The characteristics of the simulated SNe are designed to mimic currently available datasets (see [11, 1, 10, 14, 3]).

(i) We assume that the intrinsic magnitude[4] is known and fix $M = M_0 = -19.3$ and that $h = 0.72$. We also assume that the observational error is known, given by the value above.

   Using a language of your choice, write a code to carry out an MCMC sampling of the posterior probability for $(\Omega_m, \Omega_\Lambda)$ and plot the resulting 68% and 95% posterior regions, both in 2D and marginalized to 1D, using uniform priors on $(\Omega_m, \Omega_\Lambda)$ (be careful to define them explicitly).

   You should obtain a result similar to the 2D plot shown in the right panel of Fig. 2.

(ii) † Add the quantity σ (the observational error) to the set of unknown parameters and estimate it from the data along with $\mathscr{C}$. Notice that since σ is a "scale parameter", the appropriate (improper) prior is $p(\sigma) \propto 1/\sigma$ (see [4] for a justification).

---

[2]We neglect the important issue of applying the empirical corrections known as Phillip's relations to the observed light curve. This is of fundamental important in order to reduce the scatter of SNIa within useful limits for cosmological distance measurements, but it would introduce a technical complication here without adding to the fundamental scope of this exercise.

[3]Thanks to Marisa March for help with the simulation.

[4]In reality the SNe intrinsic magnitude is not fixed, but there is an "intrinsic dispersion" (even after Phillips' corrections) reflecting perhaps intrinsic variability in the explosion mechanism, or environmental parameters which are currently poorly understood.

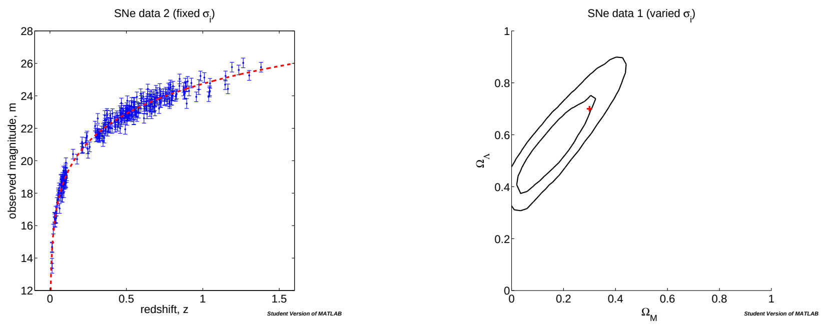

Figure 2: Left: Simulated SNIa dataset, `SNe_simulated.dat`. The solid line is the true underlying cosmology. Right: constraints on Ω_m, Ω_Λ from this dataset, with contours delimiting 2D joint 68% and 95% credible regions (uniform priors on the variables Ω_m, Ω_Λ, assuming $M = M_0$ fixed and $h = 0.72$). The red cross denotes the true value.

(iii) The location of the peaks in the CMB power spectrum gives a precise measurement of the angular diameter distance to the last scattering surface, divided by the sound horizon at decoupling. This approximately translates into an effective constraint (see [15], Fig. 20) on the following degenerate combination of Ω_m and Ω_Λ:

$$1.41\Omega_\Lambda + \Omega_m = 1.30 \pm 0.04. \tag{16}$$

Add this constraint (assuming a Gaussian likelihood, with the above mean and standard deviation) to the SNIa likelihood and plot the ensuing combined 2D and 1D limits on $(\Omega_m, \Omega_\Lambda)$.

(iv) The measurement of the baryonic acoustic oscillation scale in the galaxy power spectrum at small redshift gives an effective constraint on the angular diameter distance D_A out to $z \sim 0.3$. This measurement can be summarized as [2]:

$$D_A(z = 0.57) = (1408 \pm 45)\ \text{Mpc}. \tag{17}$$

Add this constraints (again assuming a Gaussian likelihood) to the above CMB+SNIa limits and plot the resulting combined 2D and 1D limits on $(\Omega_m, \Omega_\Lambda)$.
Hint: recall that $D_L(z) = (1 + z)^2 D_A(z)$.

References

[1] Amanullah, R., Lidman, C., Rubin, D., Aldering, G., Astier, P., et al.: Spectra and Light Curves of Six Type Ia Supernovae at 0.511 ¡ z ¡ 1.12

and the Union2 Compilation. Astrophys.J. **716**, 712–738 (2010). DOI 10.1088/0004-637X/716/1/712

[2] Anderson, L., et al.: The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: measuring D_A and H at z = 0.57 from the baryon acoustic peak in the Data Release 9 spectroscopic Galaxy sample. Mon. Not. Roy. Astron. Soc. **439**(1), 83–101 (2014). DOI 10.1093/mnras/stt2206

[3] Betoule, M., et al.: Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples. Astron. Astrophys. **568**, A22 (2014). DOI 10.1051/0004-6361/201423413

[4] Box, G.E.P., Tiao, G.C.: Bayesian Inference in Statistical Analysis. John Wiley & Sons, Chicester, UK (1992)

[5] Foreman-Mackey, D., Hogg, D.W., Lang, D., Goodman, J.: emcee: The MCMC Hammer. Pub. Astron. Soc. Pac. **125**, 306–312 (2013). DOI 10.1086/670067

[6] Gelman, A., Roberts, G., Gilks, W.: Efficient Metropolis Jumping Rules. In: J. Bernardo, J. Berger, A. Dawid, A. Smith (eds.) Bayesian statistics 5, vol. 30, pp. 599–607. Oxford University Press (1996)

[7] Goodman, J., Weare, J.: Ensemble samplers with affine invariance. Comm. App. Math. Comp. Sci. **5**, 65 (2010)

[8] Hastings, W.K.: Monte Carlo sampling methods using Markov chains and their applications. Biometrika **57**, 97–109 (1970)

[9] Kelly, B.C.: Some Aspects of Measurement Error in Linear Regression of Astronomical Data. ApJ **665(2)**, 1489–1506 (2007)

[10] Kessler, R., Becker, A., Cinabro, D., Vanderplas, J., Frieman, J.A., et al.: First-year Sloan Digital Sky Survey-II (SDSS-II) Supernova Results: Hubble Diagram and Cosmological Parameters. Astrophys.J.Suppl. **185**, 32–84 (2009). DOI 10.1088/0067-0049/185/1/32

[11] Kowalski, M., et al.: Improved Cosmological Constraints from New, Old and Combined Supernova Datasets. Astrophys.J. **686**, 749–778 (2008). DOI 10.1086/589937

[12] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E.: Equation of state calculations by fast computing machines. J. Chem. Phys. **21**, 1087–1092 (1953)

[13] Neal, R.: Mcmc using hamiltonian dynamics. In: S. Brooks, A. Gelman, G. Jones, X.L. Meng (eds.) Handbook of Markov Chain Monte Carlo. Chapman and Hall/CRC Press (2011)

[14] Rest, A., et al.: Cosmological Constraints from Measurements of Type Ia Supernovae discovered during the first 1.5 yr of the Pan-STARRS1 Survey. Astrophys. J. **795**(1), 44 (2014). DOI 10.1088/0004-637X/795/1/44

[15] Spergel, D.N., et al.: Wilkinson Microwave Anisotropy Probe (WMAP) three year results: implications for cosmology. Astrophys. J. Suppl. **170**, 377 (2007). DOI 10.1086/513700

4 Solutions to Selected Problems

4.1 Bayesian Reasoning

Let $BB = 1$ denote the proposition that your friend has the virus, and $BB = 0$ that she does not. We use $+$ $(-)$ to denote the test returning a positive (negative) result.

(i) We know from the reliability of the test that

$$P(+|BB=1) = 0.95 \tag{18}$$
$$P(+|BB=0) = 0.05 \text{ hence} \tag{19}$$
$$P(-|BB=0) = 0.95. \tag{20}$$

Given that 1% of the population has the virus, the probability of being one of them (before taking the test) is $P(BB=1) = 0.01$, while $P(BB=0) = 0.99$.

(ii) The probability of your friend having the virus after she has tested positive is thus

$$P(BB=1|+) = \frac{P(+|BB=1)P(BB=1)}{P(+)}. \tag{21}$$

We can compute the denominator as follows:

$$P(+) = P(+|BB=1)P(BB=1)+P(+|BB=0)P(BB=0) = 0.95{\cdot}0.01+0.05{\cdot}0.99 = 0.059. \tag{22}$$

Therefore the probability that your friend has the virus is much less than 95%, namely

$$P(BB=1|+) = \frac{0.95 \cdot 0.01}{0.059} = 0.16 = 16\%. \tag{23}$$

(iii) From the above, we have that

$$P(BB=1|+) = \frac{P(+|BB=1)P(BB=1)}{P(+|BB=1)P(BB=1)+P(+|BB=0)P(BB=0)}. \tag{24}$$

We want to achieve 99% probability that the person has BB given that they tested positive, i.e., $P(BB=1|+) = 0.99$, and we need to solve the above equation for the reliability, i.e. $P(+|BB=1)$.

We first assume that $P(+|BB=0) = 0.05$, as in part (a). Since $P(BB=1) = 10^{-6}$, it follows that $P(BB=0) = 1 - P(BB=1) \approx 1$. Then Eq. (24) becomes

$$0.99 = \frac{P(+|BB=1)10^{-6}}{P(+|BB=1)10^{-6} + 0.05 \times 1} \approx \frac{P(+|BB=1)10^{-6}}{0.05}. \tag{25}$$

It is clear that this equation has no solution for $P(+|BB=1) \leq 1$. This means that for a 5% false positive rate and for the given incidence

$P(BB=1)$ it is impossible to obtain a test that is 99% reliable. Therefore in order to achieve 99% reliability, the false positive rate, $P(+|BB=0)$, has to be reduced, as well.

4.2 Bayesian Parameter Estimation

4.2.1 Coin Tossing

(i) The likelihood function for p_H is given by the binomial

$$\mathcal{L}(p_H) = P(H|p_H, N) = \binom{N}{H} p_H^H (1-p_H)^{N-H}. \tag{26}$$

The parameters is $\theta = p_H$ and the data is H (for a given N).

(ii) A reasonable, non-informative prior choice is a "flat prior" over p_H, *i.e.*, uniform over $0 \le p_H \le 1$.

(iii) With the above prior, the posterior is numerically identical to the likelihood, apart from the normalizing constant (the evidence), and we obtain:

$$P(p_H|H,N) = \frac{\mathcal{L}(p_H)}{Z} \tag{27}$$

where

$$Z = \binom{N}{H} \int_0^1 \mathrm{d}p_H \; p_H^H (1-p_H)^{N-H} = \binom{N}{H} \frac{H!(N-H)!}{(N+1)!}. \tag{28}$$

(iv) The posterior mean is given by

$$\int dp_H p_H p(p_H|H,N). \tag{29}$$

This can be computed numerically with a simple 1-dimensional integrator. For the case where the posterior is approximately Gaussian (valid if $N \gg 1$ and H/N is sufficiently away from either 0 or 1), then it is simple to show that the posterior mean tends to the maximum likelihood estimate, $p_H = H/N$.

A measure of the uncertainty of our estimate for p_H is the standard deviation of the posterior, which becomes very close to Gaussian for large N and not too small H, as apparent from Fig. 3. We can estimate the standard deviation by expanding the posterior to second order in p_H around the maximum, and the standard deviation is then given by minus the curvature of the log-posterior at the peak:

$$P(p_H|H,N) \approx P_0 \exp\left(-\frac{1}{2}\frac{(p_{\rm ML}-p)^2)}{\Sigma^2}\right), \tag{30}$$

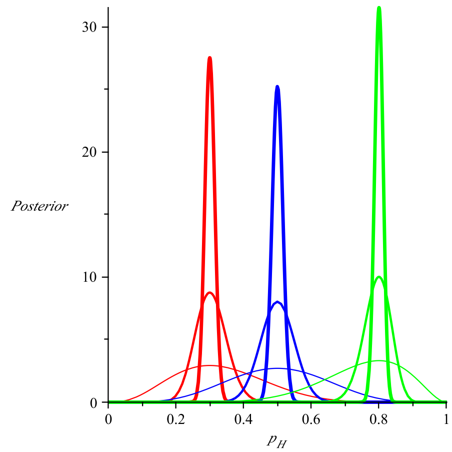

Figure 3: Posterior probability for the coin bias for $N = 10, 100, 1000$ (from thin to thick) and $H/N = 0.3, 0.5, 0.8$ (from left to right).

where

$$\begin{aligned}\Sigma^{-2} &= -\frac{\partial^2 \ln P(p_H|H,N)}{\partial p^2}\Big|_{p=p_{\mathrm{ML}}} = -\frac{\partial}{\partial p}\left(\frac{H}{p} - \frac{N-H}{1-p}\right)\Big|_{p=p_{\mathrm{ML}}} \\ &= \frac{H - 2Hp + p^2 N}{p^2(1-p)^2}\Big|_{p=p_{\mathrm{ML}}} = \frac{N}{\frac{H}{N}(1-H/N)}.\end{aligned} \tag{31}$$

and p_{ML} is the maximum of the posterior, given by $p_{\mathrm{ML}} = H/N$ (as apparent from derivating Eq. (27) wrt p_H and setting it to 0). So the standard deviation of the posterior is approximately given by

$$\Sigma \approx \frac{(\frac{H}{N}(1-H/N))^{1/2}}{\sqrt{N}}. \tag{32}$$

(v) For $N = 100$, the posterior on p_H is plotted in Fig. 3 for a few choices of H and $N = 10, 100, 1000$.

4.2.2 Gaussian Linear Model

To be done numerically.

4.2.3 Poisson Counts

(i) (a) The likelihood function is given by the Poisson distribution

$$\mathcal{L}(\hat{r}) = P(\hat{r}|\lambda) = \frac{(\lambda t)^{\hat{r}}}{\hat{r}!}\exp(-\lambda t), \tag{33}$$

where t is the time of observation in minutes. The unknown parameter is the source strength λ (in units of photons/min), while the data are the observed counts, $\hat{r}$.

(b) We can compute the requested probability by substituting in the Poisson distribution above the values for $\hat{r}$ and λ, obtaining:

$$P(\hat{r} = 15|\lambda = 10, t = 1 \text{ min}) = 0.0347. \tag{34}$$

(c) The maximum likelihood estimate is obtained by finding the maximum of the log likelihood as a function of the parameter (here, the rate λ). Hence we need to find the value of λ such that:

$$\frac{\partial \ln \mathcal{L}(\hat{r})}{\partial \lambda} = 0. \tag{35}$$

The derivative gives

$$\frac{\partial \ln \mathcal{L}(\hat{r})}{\partial \lambda} = \frac{\partial}{\partial \lambda}\left(\hat{r}\ln(\lambda t) - \ln \hat{r}! - \lambda t\right) = \hat{r}\frac{t}{\lambda t} - t = 0 \Leftrightarrow \lambda_{MLE} = \frac{\hat{r}}{t}. \tag{36}$$

So the maximum likelihood estimator for the rate is the observed number of counts divided by the time. In this case, $t = 1$ min so the MLE for λ is 10 photons per minute.

(ii) (a) The likelihood function now needs to be modified to account for the fact that the observed counts are the superposition of the background rate and the source rate (the star). According to the hint, the likelihood for the total number counts, $\hat{r}_t$, is Poisson with rate $\lambda = \lambda_s + \lambda_b$, and thus

$$P(\hat{r}_t|\lambda = \lambda_s + \lambda_b) = \frac{(\lambda t_t)^{\hat{r}_t}}{\hat{r}_t!}\exp(-\lambda t_t). \tag{37}$$

Similarly to what we have done above, the MLE estimate for λ_s is found by setting to 0 the derivative of the log likelihood wrt λ_s:

$$\frac{\partial \ln P(\hat{r}_t|\lambda = \lambda_s + \lambda_b)}{\partial \lambda_s} = \hat{r}_t\frac{t_t}{(\lambda_s + \lambda_b)t_t} - t_t = 0 \Leftrightarrow \lambda_s = \frac{\hat{r}_t}{t_t} - \lambda_b. \tag{38}$$

So the MLE for the source is given by the observed average total rate ($\frac{\hat{r}_t}{t_t}$) minus the background rate.

(b) Inserting the numerical results, we have that $\lambda_s = 3$. The MLE estimate for λ_s gives a negative rate if $\hat{r}_t/t_t < \lambda_b$, which is clearly non-physical.

(c) However, this can definitely happen because of downwards fluctuations in the number counts due to the Poisson nature of the signal (even if the background is assumed to be known perfectly). So this is an artefact of the MLE estimator (nothing to do with physics! We *know* that the actual physical source rate has to be a non-negative quantity!). The solution is to use Bayes theorem instead.

(a) The (improper) prior distribution is given by

$$p(b) = \Theta(b) \equiv \begin{cases} 1, & \text{if } b \geq 0 \\ 0, & \text{otherwise,} \end{cases} \tag{39}$$

where $\Theta(b)$ denotes the above step-function (verify explicitly that the final result does not change if instead you assume an upper cut-off for the prior, b_{max}, and then take the limit $b_{\text{max}} \to \infty$).

The (un-normalized) posterior is the product of the prior and the likelihood, namely:

$$p(b|n_{\text{off}}, t_{\text{off}}) \propto (bt_{\text{off}})^{n_{\text{off}}} \exp(-bt_{\text{off}})\Theta(b) \tag{40}$$

(b) The same considerations apply as before, except that the likelihood is now for the total rate $s+b$ and the data are n_{on} (t_{on} can be considered as an experimental parameter in this context, as it's supposed to be known; correspondingly, we put in on the right-hand-side of the conditioning bar in the likelihood):

$$p(n_{\text{on}}|s, b, t_{\text{on}}) = \frac{((s+b)t_{\text{on}})^{n_{\text{on}}} \exp(-(s+b)t_{\text{on}})}{n_{\text{on}}!}. \tag{41}$$

(c) The joint posterior for both s, b is given by the expression (dropping normalization terms that do not depend on s, b):

$$p(s, b|n_{\text{on}}, t_{\text{on}}) \propto p(s|b, n_{\text{on}}, t_{\text{on}})p(b) = p(n_{\text{on}}|s, b, t_{\text{on}})p(s)p(b), \tag{42}$$

where the prior on the background rate, $p(b)$, can be taken to be the posterior from the "off" measurement, i.e., $p(b) = p(b|n_{\text{off}}, t_{\text{off}})$. Replacing $p(b)$ from Eq (40) and $p(n_{\text{on}}|s, b, t_{\text{on}})$ from Eq. (41), and taking again a uniform (improper) prior on s, $p(s) = \Theta(s)$, we obtain:

$$\begin{aligned} p(s, b|n_{\text{on}}, t_{\text{on}}) &\propto ((s+b)t_{\text{on}})^{n_{\text{on}}} \exp(-(s+b)t_{\text{on}})(bt_{\text{off}})^{n_{\text{off}}} \exp(-bt_{\text{off}})\Theta(b)\Theta(s) \\ &\propto (s+b)^{n_{\text{on}}} b^{n_{\text{off}}} \exp(-st_{\text{on}}) \exp(-b(t_{\text{on}} + t_{\text{off}}))\Theta(b)\Theta(s), \end{aligned} \tag{43}$$

where in the second line we have dropped factors that do not depend on s, b.

(d) We make use of the hint and write (using again a uniform, improper prior on $s, b \geq 0$, i.e. $\Theta(s)$ and $\Theta(b)$):

$$p(s|n_{\rm on}, t_{\rm on}) \propto \int_0^\infty \sum_k^{n_{\rm on}} \binom{n_{\rm on}}{k} s^k b^{n_{\rm on}+n_{\rm off}-k} \exp(-st_{\rm on}) \exp(-b(t_{\rm on}+t_{\rm off}))db \tag{44}$$

$$= \sum_k^{n_{\rm on}} \binom{n_{\rm on}}{k} \exp(-st_{\rm on}) s^k \int_0^\infty dbb^{n_{\rm on}+n_{\rm off}-k} \exp(-b(t_{\rm on}+t_{\rm off})) \tag{45}$$

$$= \sum_k^{n_{\rm on}} \binom{n_{\rm on}}{k} \frac{e^{-st_{\rm on}} s^k}{(t_{\rm on}+t_{\rm off})^{n_{\rm on}+n_{\rm off}-k}} \int_0^\infty dx x^{n_{\rm on}+n_{\rm off}-k} e^{-x} \tag{46}$$

$$\propto \sum_k^{n_{\rm on}} t_{\rm on} \frac{e^{-st_{\rm on}}(t_{\rm on}s)^k}{k!} \frac{(n_{\rm on}+n_{\rm off}-k)!}{(n_{\rm on}-k)!} \left(1+\frac{t_{\rm off}}{t_{\rm on}}\right)^k \tag{47}$$

$$= \sum_k^{n_{\rm on}} C_k t_{\rm on} \frac{e^{-st_{\rm on}}(t_{\rm on}s)^k}{k!} \tag{48}$$

where we have used the Gamma integral $\Gamma(n) = \int_0^\infty x^{n-1}e^{-n} = (n-1)!$. The final result shows that our probability for the signal s is a weighted average of the posterior probability for the signal s that one would obtain if one had observed $k = 0, \ldots, n_{\rm on}$ counts in the on measurement, where the weighting factor C_k is the probability that the the number of counts k is attributed to the signal.

The cases (a)-(c) are plotted in Fig. 4.

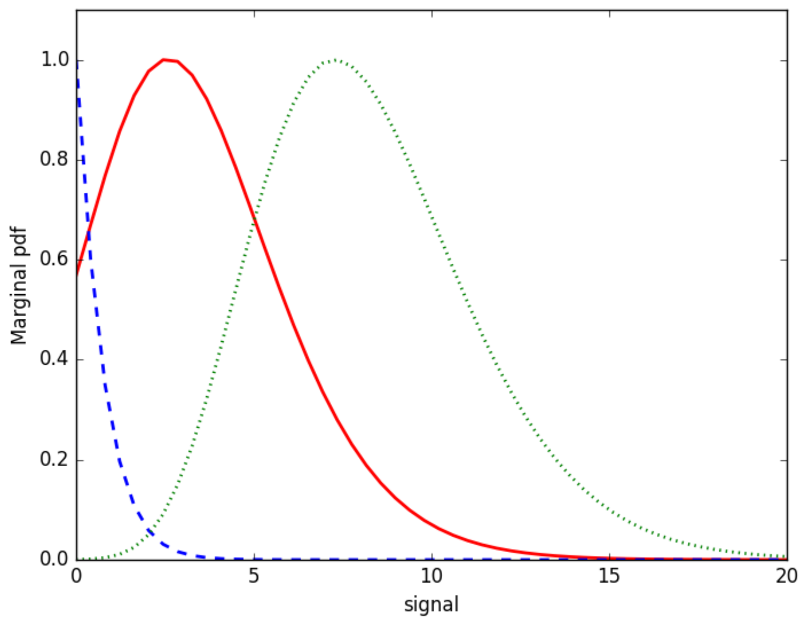

Figure 4: Marginal posterior distribution for the signal s (marginalized over background b and normalized to the peak), for the three cases: (a) red (b) blue and (c) green.