The Dark and Visible Side of the Universe Summer School

Roberto Trotta, Imperial College London

Texel, NL, July 2017

Contents

1 Annotated Bibliography

- For general introduction to inference and Bayesian methods, see [1].
- On the comparison of p-values with Bayesian methods for discovery, see [2].
- On posterior predictive p-values and failures of classical Likelihood Ratio Tests, see [3].
- On the Bayesian analysis of energy spectra, see [4].
- For a Bayesian analysis of the on/off problem, see [5].
- For details on procedures of setting confidence limits, see [6].
- For a comparison of Feldman & Cousins with Bayesian posterior for the on/off problem, see [8].
- For Bayesian Hierarchical (linear) models (with Gaussian noise), see [9].

2 Exercises

2.1 MLE for Poisson counts

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 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 $\hat{r} = 10$?

2.2 Counting experiment – multiple measurements

(i) An experiment counting particles emitted by a radioactive decay measures r particles per unit time interval. The counts are Poisson distributed. If λ is the average number of counts per per unit time interval, write down the appropriate probability distribution function for r.

(ii) Now we seek to determine λ by repeatedly measuring for M times the number of counts per unit time interval. This series of measurements yields a sequence of counts $\hat{\mathbf{r}} = \{\hat{r}_1, \hat{r}_2, \hat{r}_3, ..., \hat{r}_M\}$. Each measurement is assumed to be independent. Derive the combined likelihood function for λ, $\mathcal{L}(\lambda) = P(\hat{\mathbf{r}}|\lambda)$, given the measured sequence of counts $\hat{\mathbf{r}}$.

(iii) Use the Maximum Likelihood Principle applied to the the log likelihood $\ln \mathcal{L}(\lambda)$ to show that the Maximum Likelihood estimator for the average rate λ is just the average of the measured counts, $\hat{\mathbf{r}}$, i.e.

$$\lambda_{\mathrm{ML}} = \frac{1}{M} \sum_{i=1}^{M} \hat{r}_i \,.$$

(iv) By considering the Taylor expansion of $\ln \mathcal{L}(\lambda)$ to second order around λ_{ML}, derive the Gaussian approximation for the likelihood $\mathcal{L}(\lambda)$ around the Maximum Likelihood point and show that it can be written as

$$\mathcal{L}(\lambda) \approx L_0 \exp\left(-\frac{1}{2}\frac{M}{\lambda_{\mathrm{ML}}}(\lambda - \lambda_{\mathrm{ML}})^2\right) \,,$$

where L_0 is a normalization constant.

(v) Compare with the equivalent expression for M Gaussian-distributed measurements to show that the variance σ^2 of the Poisson distribution is given by $\sigma^2 = \lambda$.

2.3 On/Off problem

Upon reflection, the astronomer of Problem 2.1 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.

2.4 Feldman & Cousins confidence belt construction

Consider the construction of the confidence intervals using the ordering principle described by Feldman & Cousins [7].

(i) What is the 90% Feldman & Cousins confidence interval for the signal rate, if the (known) background is $b = 4.0$ and you have measured $n = 5$ counts? (Hint: use the tables at the back of the paper!)

(ii) If the background rate is $b = 5.0$, how many counts do you need to measure before you can get a 2-sided confidence interval (away from 0) using the Feldman & Cousins construction? (Hint: use the tables at the back of the paper!)

(iii) Experiment 1 expects $b = 0.0$ background counts, and measures $n = 0$ counts. Experiment 2 has a larger expected background, $b = 10.0$, and also measures $n = 0$ counts. What is the Feldman & Cousins 90% upper limit on the signal mean in each case? Comment on whether this makes sense.

(iv) Write a code to reproduce Figure 7 in [7], showing the 90% confidence belt for the Poisson case, with a known background rate $b = 3.0$.

2.5 Bayesian reasoning: Warm-up

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?

2.6 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_{off} photons in a time t_{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_{on} of photons during a time t_{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:

$$p(s, b|n_{\text{on}}, t_{\text{on}}) \propto (s + b)^{n_{\text{on}}} b^{n_{\text{off}}} \exp(-st_{\text{on}}) \exp(-b(t_{\text{on}} + t_{\text{off}})). \quad (1)$$

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

$$p(s|n_{\text{on}}, t_{\text{on}}) = \int_0^\infty p(s, b|n_{\text{on}}, t_{\text{on}}) db. \quad (2)$$

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

(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 $n_{\text{on}} = 10, t_{\text{on}} = 2$, $n_{\text{off}} = 3, t_{\text{on}} = 1$. Marginalize over b numerically and compare the resulting numerical estimate with the analytical result above.

3 Solutions to selected exercises

3.1 MLE for Poisson counts

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

$$\mathcal{L}(\hat{r}) = P(\hat{r}|\lambda) = \frac{(\lambda t)^{\hat{r}}}{\hat{r}!} \exp(-\lambda t), \quad (3)$$

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, $\hat{r}$.

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

$$P(\hat{r} = 15|\lambda = 10, t = 1 \text{ min}) = 0.0347. \quad (4)$$

(iii) 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:

$$\frac{\partial \ln \mathcal{L}(\hat{r})}{\partial \lambda} = 0. \quad (5)$$

[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!

The derivative gives

$$\frac{\partial \ln \mathcal{L}(\hat{r})}{\partial \lambda} = \frac{\partial}{\partial \lambda} \left(\hat{r} \ln(\lambda t) - \ln \hat{r}! - \lambda t \right) = \hat{r} \frac{t}{\lambda t} - t = 0 \Leftrightarrow \lambda_{MLE} = \frac{\hat{r}}{t}. \quad (6)$$

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.

3.2 Counting experiment – multiple measurements

(i) The discrete PMF for the number of counts r of a Poisson process with average rate λ is (assuming a unit time, $t = 1$ throughout)

$$P(r) = \frac{\lambda^r}{r!} e^{-\lambda}$$

(ii) In this case

$$P(\hat{r}_i | \lambda) = \frac{\lambda^{\hat{r}_i}}{\hat{r}_i!} e^{-\lambda} ,$$

for each independent measurement $\hat{r}_i$. So the joint likelihood is given by (as measurements are independent)

$$\mathcal{L}(\lambda) = \prod_{i=1}^{M} P(\hat{r}_i | \lambda) = \prod_{i=1}^{M} \frac{\lambda^{\hat{r}_i}}{\hat{r}_i!} e^{-\lambda} . \quad (7)$$

(iii) The Maximum Likelihood Principle states that the estimator for λ can be derived by finding the maximum of the likelihood function. The maximum is found more easily by considering the log of the likelihood

$$\ln \mathcal{L}(\lambda) = \sum_{i=1}^{M} \left[\hat{r}_i \ln(\lambda) - \ln(\hat{r}_i!) - \lambda \right] .$$

with the maximum given by the condition $d\ln \mathcal{L}/d\lambda = 0$.

We have

$$\begin{aligned} \frac{d\ln \mathcal{L}}{d\lambda} &= \sum_{i=1}^{M} \left[\frac{\hat{r}_i}{\lambda} - 1 \right] \\ &= \frac{1}{\lambda} \sum_{i=1}^{M} \hat{r}_i - M . \end{aligned}$$

So the Maximum Likelihood (ML) estimator for λ is

$$\lambda_{\mathrm{ML}} = \frac{1}{M} \sum_{i=1}^{M} \hat{r}_i ,$$

which is just the average of the observed counts.

(iv) The Taylor expansion is

$$\ln \mathcal{L}(\lambda) = \ln \mathcal{L}(\lambda_{\rm ML}) + \left.\frac{d\ln \mathcal{L}}{d\lambda}\right|_{\lambda=\lambda_{\rm ML}} (\lambda - \lambda_{\rm ML}) + \frac{1}{2}\left.\frac{d^2\ln \mathcal{L}}{d\lambda^2}\right|_{\lambda=\lambda_{\rm ML}} (\lambda - \lambda_{\rm ML})^2 + \dots .$$

By definition the linear term vanishes at the maximum so we just need the curvature around the ML point

$$\frac{d^2\ln \mathcal{L}}{d\lambda^2} = -\sum_{i=1}^{M} \frac{\hat{r}_i}{\lambda^2},$$

such that

$$\left.\frac{d^2\ln \mathcal{L}}{d\lambda^2}\right|_{\lambda=\lambda_{\rm ML}} = -\frac{1}{\lambda_{\rm ML}^2}\sum_{i=1}^{M} \hat{r}_i = -\frac{M\lambda_{\rm ML}}{\lambda_{\rm ML}^2} = -\frac{M}{\lambda_{\rm ML}}.$$

Putting this into the Taylor expansion gives

$$\ln \mathcal{L}(\lambda) = \ln \mathcal{L}(\lambda_{\rm ML}) - \frac{1}{2}\frac{M}{\lambda_{\rm ML}}(\lambda - \lambda_{\rm ML})^2,$$

which gives an approximation of the likelihood function around the ML point

$$\mathcal{L}(\lambda) \approx L_0 \exp\left(-\frac{1}{2}\frac{M}{\lambda_{\rm ML}}(\lambda - \lambda_{\rm ML})^2\right),$$

(the normalisation constant L_0 is irrelevant).

So the likelihood is approximated by a Gaussian with variance

$$\Sigma^2 = \frac{\lambda_{\rm ML}}{M}.$$

(v) Comparing this with the standard result for the variance of the mean for the Gaussian case, i.e.

$$\Sigma^2 = \frac{\sigma^2}{M},$$

where M is the number of measurements and σ is the standard deviation of each measurement, we can conclude that the variance of the Poisson distribution itself is indeed

$$\sigma^2 = \lambda.$$

References

[1] R. Trotta, "Bayesian Methods in Cosmology", Lecture notes for the 44th Saas Fee Advanced Course on Astronomy and Astrophysics,"Cosmology with wide-field surveys" (March 2014), to be published by Springer. arXiv:1701.01467 [astro-ph.CO].

[2] L. Demortier & L. Lyons, "Testing Hypotheses in Particle Physics: Plots of p_0 Versus p_1", eprint arXiv:1408.6123.

[3] R. Protassov, D. A. van Dyk, A. Connors, V. L. Kashyap and A. Siemiginowska, Astrophys. J. **571**, 545 (2002) doi:10.1086/339856 [astro-ph/0201547].

[4] D. A. van Dyk, A. Connors, V. L. Kashyap and A. Siemiginowska, Astrophys. J. **548**, 224 (2001) doi:10.1086/318656 [astro-ph/0008170].

[5] Loredo, T. J. 1993, in Statistical Challenges in Modern Astronomy, eds. G.J. Babu & E.D. Feigelson (Springer-Verlag: New York), 275. Available from: `http://www.astro.cornell.edu/staff/loredo/bayes/promise.pdf` (accessed June 30th 2017).

[6] Workshop on Confidence Limits, CERN proceedings F. James, L. Lyons & Y. Perrin (Eds), (2000). Available from: `http://cds.cern.ch/record/411537/files/CERN-2000-005.pdf?version=2` (accessed June 30th 2017).

[7] G. J. Feldman and R. D. Cousins, Phys. Rev. D **57**, 3873 (1998) doi:10.1103/PhysRevD.57.3873 [physics/9711021 [physics.data-an]].

[8] W. A. Rolke, Nucl. Instrum. Meth. A **806**, 318 (2016) doi:10.1016/j.nima.2015.10.028 [arXiv:1505.07027 [physics.data-an]].

[9] B. C. Kelly, Astrophys. J. **665**, 1489 (2007) doi:10.1086/519947 [arXiv:0705.2774 [astro-ph]].