BAYESIAN INFERENCE ON POISSON DISTRIBUTION
ONE SAMPLE
TWO SAMPLES (RATIO)
STRATIFIED SAMPLES

Bruno Lecoutre

CNRS and Université de Rouen
ERIS, Laboratoire de Mathématiques Raphaël Salem, UMR 6085
Avenue de l'Université, BP 12
F 76801 Saint-Etienne-du-Rouvray (France)
bruno.lecoutre@univ-rouen.fr

ERIS 2011 ◊◊◊ eris62.eu

Bayesian inference for one sample

We have n observations $\boldsymbol{x} = (x_1, x_2 \dots x_n)$ from a Poisson distribution with parameter λ

$$Pr(x_i|\lambda) = \frac{\lambda^{x_i}}{x_i!} \exp(-\lambda)$$

hence the likelihood

$$\ell(\lambda|\boldsymbol{x}) \propto \lambda^S \exp(-n\lambda)$$

where S is the sufficient statistic

$$S = \sum_{i=1}^{n} x_i.$$

Conjugate prior

An appropriate conjugate prior is a scaled Gamma distribution

$$\lambda \sim \frac{1}{n_0} \text{Gamma}(S_0)$$

with density

$$p(\lambda) \propto \lambda^{S_0 - 1} \exp(-n_0\lambda).$$

Then the posterior density is

$$p(\lambda|\boldsymbol{x}) \propto \ell(\lambda|\boldsymbol{x}) p(\lambda) = \lambda^{S_1 - 1} \exp(-(n_1)\lambda)$$

hence

$$\lambda|\boldsymbol{x} \sim \frac{1}{n_0 + n} \text{Gamma}(S_0 + S).$$

Jeffreys prior

The Jeffreys prior is $p(\lambda) \sim \lambda^{-1/2}$, which corresponds to $S_0 = 1/2$ and $n_0 = 0$, that is a prior uniform in $\sqrt{\lambda}$.

Remark: Gamma and Chi-square distributions

Some authors (see, e.g., Lee, 2004) use the Chi-square distribution prior. We have the equivalence

$$\frac{1}{n_1} \text{Gamma}(S_1) \sim \frac{1}{2n_1} e\chi^2_{S_1}$$

but the parametrization of the Gamma distribution is more natural.

Numerical application

Lee (2004) considered the numbers of misprints spotted on the first few pages of an eraly draft of his book:

$$3 \qquad 4 \qquad 2 \qquad 1 \qquad 2 \qquad 3$$

Assuming that these numbers are a sample from a Poisson distribution of mean λ, we have

$$n = 6 \qquad S = 15$$

For the Jeffreys prior, we have the posterior distribution

$$\lambda|\boldsymbol{x} \sim \frac{1}{6}\text{Gamma}(15.5) \sim \frac{1}{12}\chi^2_{31}.$$

Bayesian interpretation of the frequentist significance test

For Poisson sampling (as well as for Binomial sampling: see Lecoutre, 2008), different non-informative priors have been proposed. In fact, it exists two extreme noninformative priors that are respectively the more unfavorable and the more favorable priors with respect to the null hypothesis $\lambda = \lambda_0$. They are the Gamma distribution with $S_0 = n_0 = 0$ and the Gamma distribution with $S_0 = 1$ and $n_0 = 0$. These priors lead to the Bayesian interpretation of the one-sided Poisson test: the observed significance levels of the inclusive and exclusive conventions are exactly the posterior Bayesian probabilities that λ is greater than λ_0 respectively associated with these two extreme priors.

Indeed, the sampling distribution of S under $H_0 : \lambda = \lambda_0$ (vs $H_a : \lambda < \lambda_0$) is

$$S|H_0 \sim \text{Poisson}(n\lambda)$$

so that the observed p-value is

$p_{inc} = Pr(\text{Poisson}(n\lambda_0) \leq S)$ for the inclusive (conservative) convention

and

$p_{exc} = Pr(\text{Poisson}(n\lambda_0) < S)$ for the exclusive (liberal) convention

the "mid p-value being

$p_{mid} = 1/2(p_{inc} + p_{exc})$.

Using the equality

$$Pr(\text{Poisson}(n\lambda_0) < S) = Pr(\text{Gamma}(S) > n\lambda_0)$$

we get

$p_{inc} = Pr(\text{Poisson}(n\lambda_0) < S+1) = Pr(\frac{1}{n}\text{Gamma}(S+1) > \lambda_0)$

$p_{exc} = Pr(\text{Poisson}(n\lambda_0) < S) = Pr(\frac{1}{n}\text{Gamma}(S) > \lambda_0)$.

Note that the Jeffreys prior Gamma($1/2, 0$ is very naturally intermediate between the two extreme priors Gamma($0, 0$ and $1, 0$, so that it gives a posterior probability, fully justified, close to the observed mid-p-value (which has only *ad hoc* justifications.).

Bayesian inference for two independent groups

We have n_1 observations from a Poisson distribution with parameter λ_1 and n_2 observations from a Poisson distribution with parameter λ_2. The sum of the observations are respectively S_1 an n_1. Assuming two independent marginal conjugate priors

$$\lambda_g \sim \frac{1}{n_g^0}\,\text{Gamma}(S_g^0) \quad (g = 1,2)$$

the marginal posterior distributions are the two independent Gamma distributions

$$\lambda_g|\boldsymbol{x} \sim \frac{1}{n_g^0 + n_g}\,\text{Gamma}(S_g^0 + S_g) \quad (g = 1,2).$$

The posterior distributions of any derived parameter can be approximated by generating a sample from these two marginals. The Jeffreys prior corresponds to $S_1^0 = S_2^0 = 1/2$ and $n_1^0 = n_2^0 = 0$.

Bayesian inference about the ratio $\tau = \lambda_1/\lambda_2$

For the ratio $\tau = \lambda_1/\lambda_2$, we have a particularly simple explicit form, since it involves the distribution of the ratio of two independent Gamma distributions. This is a scaled Beta distribution of the second kind (sometimes called a Beta-prime distribution)

$$\tau|\boldsymbol{x} \sim \frac{n_2^0 + n_2}{n_1^0 + n_1}\,\text{Beta}_{\text{II}}(S_1^0 + S_1, S_2^0 + S_2).$$

It is also a scaled F distribution (with a less natural parametrization)

$$\tau|\boldsymbol{x} \sim \frac{n_2^0 + n_2}{n_1^0 + n_1}\,\frac{S_1^0 + S_1}{S_2^0 + S_2}\,F_{2(S_1^0+S_1),2(S_2^0+S_2)}.$$

We have a Bayesian interpretation of the Fisher ("exact") conditional significance test and of its associated confidence interval. In particular, for the inclusive convention, the lower and upper limits of this confidence interval respectively coincide with the Bayesian lower credible limit of τ associated with the prior ($S_0^1 = 0$ $n_0^1 = 0$ $S_0^2 = 1$ $n_0^2 = 0$) and with the upper limit associated with the prior ($S_1^1 = 0$ $n_0^1 = 0$ $S_0^2 = 0$ $n_0^2 = 0$). For the exclusive convention, we have the reverse result.

Numerical application

The data correspond to "Example 2. Seizure for Rimonabant (pooled data)" from Huang et al. (2008):

$$S^1 = 8 \quad n^1 = 3418.2 \qquad S^2 = 5 \quad n^2 = 2781$$

We get the posterior distributions and the 90% equal tails credible intervals

$\lambda \sim 0.8136\text{Beta}_{\text{II}}(8.0, 6.0) \to [0.447, 2.819]$ for the prior $(0,0,1,0)$
$\lambda \sim 0.8136\text{Beta}_{\text{II}}(9.0, 5.0) \to [0.607, 4.098]$ for the prior $(1,0,0,0)$
$\lambda \sim 0.8136\text{Beta}_{\text{II}}(8.5, 5.5) \to [0.521, 3.376]$ for the Jeffreys prior $(1/2, 0, 1/2, 0)$

The corresponding frequentist 90% confidence intervals are

$[0.447, 4.098]$ ("exact" inclusive)
$[0.607, 2.819]$ (exclusive)
$[0.504, 3.538]$ (mid-p)

Bayesian inference for stratified groups

Within each groupe g, for each stratum i, we have $n_{i<g>}$ observations from a Poisson distribution with parameter $\lambda_{i<g>}$.

Assuming independent marginal conjugate priors

$$\lambda_{i<g>} \sim \frac{1}{n^0_{i<g>}} \text{Gamma}(S^0_{i<g>})$$

the marginal posterior distributions are independent Gamma distributions

$$\lambda_{i<g>}|\boldsymbol{x} \sim \frac{1}{n^0_{i<g>} + n_{i<g>}} \text{Gamma}(S^0_{i<g>} + S_{i<g>}).$$

The posterior distributions of any derived parameter can be approximated by generating a sample from these marginals. The Jeffreys prior corresponds to $S^0_{i<g>} = 1/2$ and $n^0_{i<g>} = 0$.

Bayesian inference about the ratio τ of the weighted λ

If, for each marginal prior, we assume $n^0_{i<g>} = 0$, then the distribution of the ratio $\tau = \lambda_1/\lambda_2$ of the weighted parameters

$$\lambda_g = \frac{\sum_{i<g>} n_{i<g>} \lambda_{i<g>}}{\sum_{i<g>} n_{i<g>}}$$

is again a Beta distribution of the second kind. Indeed, we have in this case

$$n_{i<g>}\lambda_{i<g>}|\boldsymbol{x} \sim \text{Gamma}(S^0_{i<g>} + S_{i<g>})$$

The sum of independent unscaled Gamma distributions is again a Gamma distribution

$$\lambda_g|\boldsymbol{x} \sim \frac{1}{\sum_{i<g>} n_g} \text{Gamma}(S^0_g + S_g)$$

where

$$S^0_g = \sum_{i<g>} S_{i<g>^0} \quad S_g = \sum_{i<g>} S_{i<g>} \quad n_g = \sum_{i<g>} n_{i<g>}$$

and consequently, τ has the distribution

$$\tau|\boldsymbol{x} \sim \frac{n_2}{n_1} \text{Beta}_{\text{II}}(S^0_1 + S_1, S^0_2 + S_2).$$

Numerical application

For the “Example 2. Seizure for Rimonabant” considered above, let us consider the stratified data:

$g_1 : S_{1<1>} = 7\ S_{2<1>} = 1\ S_{3<1>} = 0 \quad n_{1<1>} = 2608\ n_{2<1>} = 780\ n_{3<1>} = 30.2$
$g_2 : S_{1<2>} = 1\ S_{2<2>} = 1\ S_{3<2>} = 3 \quad n_{1<2>} = 2257.7\ n_{2<2>} = 494.1\ n_{3<2>} = 29.2.$

Assuming the Jeffreys prior, we have

$$S^0_1 = S^0_2 = 1.5 \quad S_1 = 8 \quad S_2 = 5 \quad \frac{n_2}{n_1} = \frac{2781}{3418.2} = 0.8136$$

hence the posterior distribution

$$\tau|\boldsymbol{x} \sim 0.8136\,\text{Beta}_{\text{II}}(9.5, 6.5)$$

and the 90% equal tail credible interval $[0.522, 2.938]$ (instead of $[0.521, 3.376]$ for the pooled data).

Appendix

Bayesian inference about the difference $\delta = \lambda_1/\lambda_2$

The general problem is to derive the density of the difference $d = x - y$ of two independent Gamma distributions

$$x \sim e\text{Gamma}(a) \qquad y \sim f\text{Gamma}(b).$$

We have $d|y \sim e\text{Gamma}(a) - y$, hence

$$p(d|y) = \frac{1}{\Gamma(a)}\,(\frac{d+y}{e})^{a-1}\,\exp(-\frac{d+y}{e})$$
$$= \frac{1}{\Gamma(a)}\,e^{1-a}\,\exp(-\frac{d}{e})\,(d+y)^{a-1}\,\exp(-\frac{y}{e})$$

and

$$p(y) = \frac{1}{\Gamma(b)}\,y^{b-1}\,\exp(-y)$$

We get

$$p(d) = \frac{1}{\Gamma(a)\Gamma(b)}\,e^{1-a}\,\exp(-\frac{d}{e})\,\int_0^\infty y^{b-1}\,(d+y)^{a-1}\,\exp(-\frac{e+1}{e}y)dy$$

We have to compute an integral of the form

$$I = \int_0^\infty y^u\,(d+y)^v\,\exp(-hy)\,dy$$

From

$$\exp(-hy) = \sum_{j=0}^{\infty} \frac{(-1)^j}{j!}\,h^j y^j$$

we deduce

$$I = \sum_{j=0}^{\infty} \frac{(-1)^j}{j!}\,h^j \int_0^\infty y^{u+j}\,(d+y)^v\,dy$$

$$I = \sum_{j=0}^{\infty} \frac{(-1)^j}{j!}\,h^j\,d^v \int_0^\infty y^{u+j}\left(1+\frac{y}{d}\right)^v dy$$

References

Lecoutre, B. (2008). Bayesian methods for experimental data analysis. In C. R. Rao, J. Miller & D. C. Rao (Eds.), *Handbook of statistics: Epidemiology and Medical Statistics* (Vol 27), Amsterdam, Nederland: Elsevier, 775-812.

Lee, P. (2004). *Bayesian Statistics: An Introduction* (3rd edition). Oxford University Press, New York.

Huang, X., Wang, L., Boddy, A. & Quan, H. (2008). Methods for rare events analyses. Bridgewater, N: Sanofi-Aventis.