Econ. 517, Econometric Theory I   Fall 2001   Chris Sims

Estimation and testing exercise answers*

(1) If objects arrive at random time intervals, with the time between arrivals independent, then if the expected number of arrivals per unit of time (an hour, we'll say) is λ, the pdf of the number of arrivals that actually occur in an hour is

$$p(n \mid \lambda) = \frac{\lambda^n}{n!} e^{-\lambda} .$$

This is the **Poisson** distribution. A standard recommendation for a "flat" prior on λ in this problem is to make it flat in $\log \lambda$, i.e. to give it the form $d\lambda/\lambda$.[1] Another possibility is a conjugate prior, which here takes the form of a $\Gamma(q, \alpha)$ distribution, i.e. a pdf

$$\frac{\alpha^q \lambda^{q-1} e^{-\alpha\lambda}}{\Gamma(q)} .$$

(a) Show that, if we have an i.i.d. sample n_i, $i = 1, \ldots, k$ from this Poisson pdf, then the posterior mean of λ under the $d\lambda/\lambda$ prior is an unbiased estimator in the non-Bayesian sense.

The likelihood will be the prior "pdf" $1/\lambda$ times k copies of the pdf for an individual observation:

$$\lambda^{\sum_1^k n_i - 1} e^{-k\lambda} .$$

This, as a function of λ, is proportional to a $\Gamma(\sum n_i, k)$ pdf and thus has mean $\bar{n} = \sum n_i / k$, where the expectation is taken with respect to the posterior pdf (conditional distribution for the parameter λ) with the data held fixed. As the problem statement said, the expected value of a Poisson(λ) pdf is λ, each n_i therefore has expectation λ, and their average $\bar{n}$ therefore also has expectation λ, when the expectations are taken with respect to the conditional distribution of n_i with λ held fixed.

(b) Is there any choice of α and q in the conjugate prior that at the same time makes the prior proper (i.e. makes it integrate to one) and makes the posterior mean classically unbiased?

With the conjugate prior, the posterior is proportional to

$$\lambda^{\sum n_i + q - 1} e^{-(k+\alpha)\lambda} .$$

This is proportional to a $\Gamma(\sum n_i + q, k + \alpha)$ pdf, which has a mean of $(\sum n_i + q)/(k + \alpha)$. Taking the expectation of this over $\{n_i\}$, conditional on λ, gives us $(k\lambda + q)/(k + \alpha)$. This can't be made identically equal to λ for all $\lambda > 0$ unless

[1] This is the form of the Jeffreys prior in this problem.

we set $q = \alpha = 0$. But these choices of q and α just give us the improper $d\lambda/\lambda$ prior.

(c) The unbiased estimator in part (1a) will with positive probability be zero. Explain why no Bayesian estimator derived from a proper prior (other than one that puts probability 1 on $\lambda = 0$) and a symmetric loss function could ever, in any sample, produce an estimate $\hat{\lambda} = 0$.

The answer is straightforward with the additional assumption (suggested in the posted explanatory note) that the loss function $L(|\lambda - \hat{\lambda}|)$ is differentiable, assuming also that the posterior is continuous with respect to Lebesgue measure (i.e. that it has a pdf). In that case expected loss is

$$\int_0^\infty L(|\hat{\lambda} - \lambda|)g(\lambda \mid \{n_i\})d\lambda\,,$$

where g is the posterior pdf. The derivative of this expression with respect to $\hat{\lambda}$, evaluated at $\hat{\lambda} = 0$ is, with L differentiable,

$$-\int_0^\infty L'(\lambda)g(\lambda \mid \{n_i\})d\lambda\,.$$

Since by assumption (again, this was explained to be a reasonable assumption in the explanatory posting) the loss function is increasing in λ for $\lambda > 0$, this expression is negative, meaning that increasing $\hat{\lambda}$ will produce a reduction in expected loss, so that $\hat{\lambda} = 0$ cannot be the expected-loss-minimizing estimator. If the posterior does not have a pdf, then we replace the ordinary integral with a Lebesgue integral:

$$-\int_0^\infty L'(\lambda)g(\lambda \mid \{n_i\})G(d\lambda)\,.$$

This expression is also clearly negative. As stated in the explanatory note, the result can also be shown under the assumption that L is convex, in which case differentiability is not needed.

(d) Find the posterior mean, the posterior median, and a 95% posterior probability interval for λ under the prior pdf $\lambda^{-\frac{1}{2}}e^{-\lambda}/\Gamma(\frac{1}{2})$, assuming a sample of five draws with the n_i given by 5,10,10,5,9. To see how sensitive the results are to the prior, repeat the analysis using the $d\lambda/\lambda$ prior. By using the fact that the posterior has the shape of a Γ distribution, you can do this problem without a computer, if you have access to tables of a Γ or χ^2 distribution.

The posterior is proportional to

$$\lambda^{\sum n_i - \frac{1}{2}} e^{-(k+1)\lambda}\,.$$

This is the form of a $\Gamma(\sum n_i + \frac{1}{2}, k+1)$ pdf, and therefore has mean $(\sum n_i + \frac{1}{2})/(k+1) = 6.58$ and median 6.53. The median was obtained by locating the value of x that makes `gammainc(x,39.5)=.5` in Matlab, then dividing it by $k + 1 = 6$. An equal-tailed posterior 95% probability interval can be found, again using `gammainc`, as (4.69,8.79). Note that the interval is asymmetric around the mean, going below it by 1.89, but above it by 2.21. A minimum length interval requires some more work. A pretty good approximation is (4.684,8.778). Note that this is scarcely different from the equal-tailed interval. If you were working with tables, you probably could not detect the difference.

With the $d\lambda/\lambda$ prior, the posterior is in the form of a $\Gamma(39,5)$, so it has posterior mean 7.8 and posterior median 7.73. An equal-tailed 95% interval is (5.546,9.962). The minimum length interval is no doubt also close to this. It might be interesting to note that the sample was actually generated as computer random numbers with $\lambda = 6$.

(e) Are either of the 95% probability intervals you found in (1d) non-Bayesian 95% confidence intervals? Explain your answer. [This is too hard to do as an exercise. It is included here as a point of information. The interval derived from a $d\lambda/\lambda$ prior is in fact a non-Bayesian 95% confidence interval.]

Omitted due to time constraints.

(2) You arrive at home at 7PM to find the phone is not working. It was working 10 hours ago when you left. For some decision you need to make, it is important when the phone service shut down. Your answering machine contains a single message: a call from the phone company telling you that later today the phone service will be shut down. The answering machine says this call arrived at 10AM, one hour after you left.

Here is a model for assessing the uncertainty: The time of shutdown is $Y \in (0, 10)$. Conditional on $Y = y$, the time of arrival of the call from the phone company (which we'll call X) is distributed uniformly on $(0, y)$. The marginal distribution of Y (that is, the distribution you would give it if you didn't know the time of the phone company call) is uniform on $(0, 10)$, meaning you have no idea when the shutdown occurred between when you left and when you returned.

(a) What is the posterior pdf on Y given X, the time of the phone company call?

The prior pdf on Y is 1/10 over the interval $[0, 10]$ and zero elsewhere. The conditional pdf of $X \mid Y$ is $1/y$ on the interval $[0, y]$. So the joint pdf is proportional to $1/y$ on the region where $x \leq y$. The posterior for $Y \mid X$ is then proportional to $1/y$ over the interval $[X, 10]$. This has to be normalized to integrate to 1 in y. So we need to divide by

$$\int_x^{10} \frac{1}{y}\, dy = \log(10/x)\,.$$

(b) What is a minimum-length 95% posterior probability interval for Y for the observed $X = 1$?

The posterior probability between x and y^* can be found, by calculating the integral, to be $(\log y^* - \log 1)/(\log 10 - \log 1)$. Setting this equal to .95 and solving, we get $y^* = 10^{.95} = 8.91$. So the interval is $[1, 8.91]$. Note that the minimum length interval runs all the way back to x, because the pdf is monotone decreasing in y.

(c) Here are two random intervals:

$$\left\{ y \,\middle|\, y > \frac{X}{.95} \text{ and } y < 10 \right\} \qquad \left\{ y \,\middle|\, y < \frac{X}{.05} \text{ and } y \in (X, 10) \right\}.$$

Show that these are both 95% non-Bayesian confidence sets for Y.

The presample distribution of X given the unknown Y, treating Y as the parameter, is uniform on $[0, Y]$. Therefore

$$P[0 < X < .95Y \mid Y] = P[.05Y < X < 10 \mid Y] = .95\,.$$

It is easily seen that these two probabilities translate into assertions that the two intervals above have 95% coverage probability for each possible value of Y.

(d) Compare the behavior of these non-Bayesian intervals to that of the Bayesian interval, particularly for X near 0 or near 10.

The Bayesian interval will always be non-empty and include only a fraction of the interval $[X, 10]$ of possible Y values. If it is minimum-length, it will always exclude an interval of values close to 10. The first classical interval always excludes some values in an interval just above x, and indeed is entirely empty if $X > 9.5$. The second classical interval, if it excludes anything, excludes y values near 10. It will include the entire $[x, 10]$ interval whenever $X \geq .5$. As $x \to 0$, the upper limit of the second classical interval converges linearly to zero. The Bayesian interval has upper limit $10^{.95}x^{.05}$, so it is much less sensitive to variation in x, particularly near $x = 0$. For small x, the Bayesian interval will extend much farther to the right.

(e) Food for thought: Can you think of a way to produce better-behaved non-Bayesian confidence intervals here? (Bad behavior, for example: producing empty confidence intervals or 95% intervals that contain the true value of Y with post-sample probability 1.)

I haven't been able to think of one myself. One usual way to get improvement is to base the non-Bayesian interval on the distribution of the ratio of the likelihood at its maximum to the likelihood at the parameter value, excluding a Y value when the observed likelihood ratio exceeds a 95% critical value. But the second of the two intervals above already has this form.

(3) Suppose our model is

$$\underset{10\times 1}{Y} = \underset{10\times 1}{X}\beta + \varepsilon$$
$$\varepsilon \,|\, \{X, \beta, \sigma\} \quad \sim \quad N(0, \sigma^2 I)\,.$$

We have a conjugate prior pdf given by

$$p(\beta, \sigma) = \begin{cases} .5\varphi(\beta; 100\sigma^2) g(\sigma^2)\, d\beta\, d\sigma^2 & \beta \neq 0 \\ .5 g(\sigma^2) & \beta = 0\,, \end{cases}$$

where $g(\sigma^2) = \sigma^{-2} e^{-1/\sigma^2}$ is the inverse-gamma density with parameters 1,1 and $\varphi(\cdot, a)$ is the standard normal pdf with variance a. The prior density is a density with respect to the measure that is Lebesgue measure over $\sigma > 0$ on the subspace of $\mathbb{R}^2$ where $\beta = 0$ and Lebesgue measure over β, σ elsewhere in the parameter space (which is of course the part of $\mathbb{R}^2$ on which $\sigma > 0$). Supposing our sample produces

$$\hat{\beta} = 1,\ \hat{\sigma}^2 = 1,\ X'X = 2\,.$$

Here $\hat{\beta}$ is the OLS estimator and $\hat{\sigma}^2$ is the sum of squared OLS residuals divided by degrees of freedom, which is here 9. What is the posterior probability of $\beta = 0$? What is the smallest significance level at which $H_0 : \beta = 0$ would be rejected by the usual t test that rejects when

$$\frac{\hat{\beta}}{\sqrt{\hat{\sigma}^2 / X'X}}$$

exceeds a critical value?