Posterior odds on restrictions

Christopher A. Sims
Princeton University
sims@princeton.edu

February 25, 2018

Why is this any different from general posterior odds on models?

- The restricted model is a subset of the parameter space of the unrestricted model.

- In such cases it may be (though this is not always true) that it is natural to take the prior on the restricted model to be the unrestricted prior, conditional on the restricted subset of the parameter space.

Aside on conditional density pitfalls

- Warning: When the restricted parameter space has measure zero in the unrestricted parameter space, conditioning on the zero-measure set can imply different conditional distributions over the set.

- Conditional probabilities are really conditional on sigma-fields, not on sets. If X and Y are i.i.d. $N(0,1)$, we think of the conditional distribution of X over the set where $x > 0$, $y = 0$ as the half-normal. But the same set has a different conditional density if X and Y are transformed to polar coordinates ρ, θ. The set is then $\theta = 0$, $\rho > 0$, and the conditional density given $\theta = 0$ is different.

- We focus on the case of linear restrictions in $\mathbb{R}^k$ as parameter space, with the usual convention that the conditional density of Z given $RZ = \gamma$ is that implied by treating the sigma-field of the conditional probability as that generated by sets of the form $\{z \mid Rz \in (\gamma_1, \gamma_2)\}$.

The linear case: The prior

We suppose there is a prior density $\pi()$ over the unrestricted parameter space. To simplify the discussion at first we assume θ has two components (θ_1, θ_2) and that the restriction of interest is $\theta_2 = 0$. The conditional density of $\theta_1 \mid \theta_2 = 0$ is then

$$q(\theta_1 \mid \theta_2 = 0) = \frac{\pi(\theta_1, 0)}{\int \pi(\theta_1, 0)\, d\theta_1}\,.$$

The full prior then has density $(1-\mu)\pi$ over the full parameter space and μq over the restricted space. (Here we are integrating the density over a base measure that is Lebesgue on the full space plus Lebesgue on the restsricted space.)

Linear case: The posterior

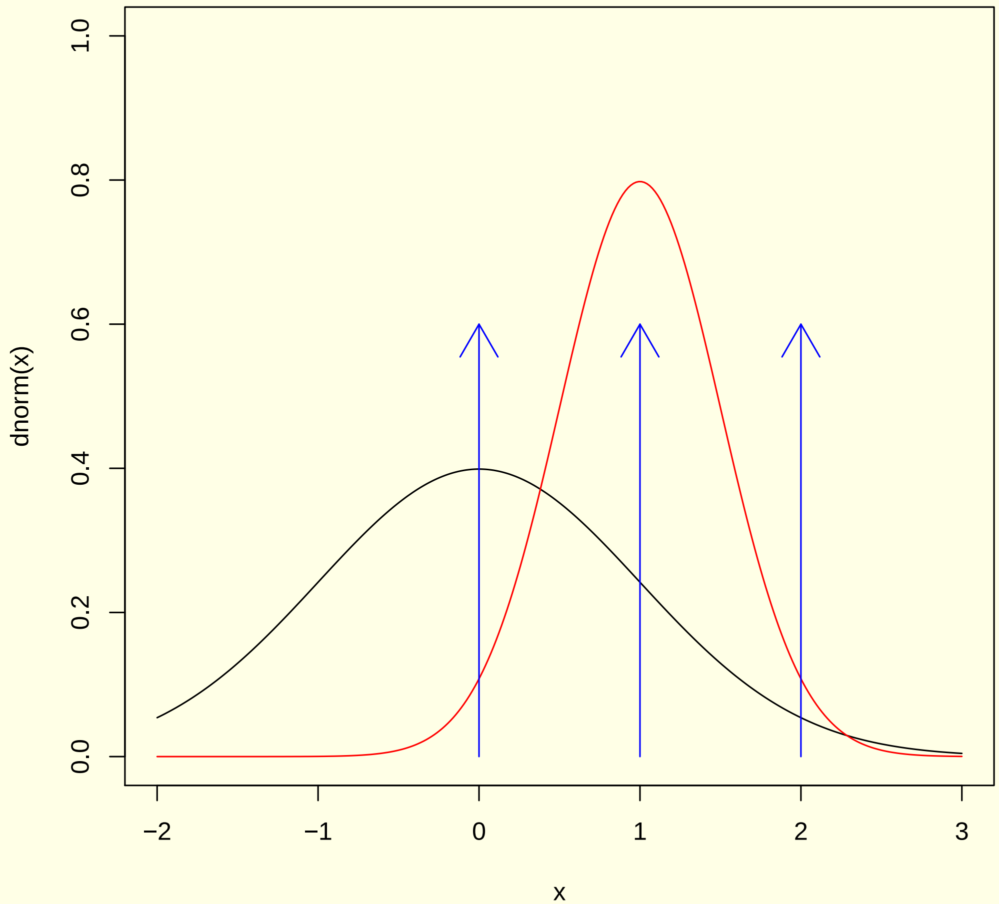
If $\ell(\theta)$ is the likelihood function, the posterior kernel is $\pi(\theta)\ell(\theta)(1-\mu)$ over the full parameter space, and $\pi(\theta_1,0)q(\theta_1)\ell(\theta_1,0)\mu$ over the restricted space. The posterior odds on the restricted space is therefore

$$\frac{\mu\int \ell(\theta_1,0)q(\theta_1)\,d\theta_1}{(1-\mu)\int \ell(\theta)\pi(\theta)\,d\theta}\,.$$

which can be rewritten as

$$\frac{\mu}{1-\mu}\cdot\frac{\int \ell(\theta_1,0)\pi(\theta_1,0)\,d\theta_1}{\int \ell(\theta)u\pi(\theta)\,d\theta)}\cdot\frac{1}{\int \pi(\theta_1,0)\,d\theta_1}$$

In words, this says that the odds ratio is the prior odds ratio times the ratio, under a simple π prior, of posterior to prior marginal densities for θ_2 evaluated at $\theta_2 = 0$.


dnorm(x)
1.0
0.8
0.6
0.4
0.2
0.0
−2
−1
0
1
2
3
x

How does this relate to the usual χ-squared test?

The log posterior marginal density at $\theta_2 = 0$ for the normal case (with conjugate prior) is

$$k_2 \log \sigma + \frac{1}{2} k_2 \log(2\pi) - \frac{1}{2} \log |X_1' X_1| + \frac{1}{2} \log |X'X| - \frac{1}{2}(\hat{u}_R' \hat{u}_R - \hat{u}'\hat{u}) \, .$$

With conjugate prior, there will be a similar expression for the log posterior marginal density under the prior, and the difference will be the log odds ratio.

Comparison to BIC

Note that as sample size increases, the marginal posterior density's relative size grows, while the for the prior it remains fixed. Also, if $(1/T)X'X$ converges to a constant Σ_X, the terms in X behave asymptotically like

$$-\tfrac{1}{2}\big(\log(|\Sigma_{X_1}|) - \log(|\Sigma_X|) - k_2 \log T\big)$$

The term at the right is the BIC criterion penalty on the likelihood ratio. Using it alone ignores the prior, the Σ_X matrix, and the terms in σ and 2π that don't vary with T.

Unknown σ^2

- content

Generalization to $R\theta = \gamma$