Econ. 556b Spring 1999 Chris Sims

Response to some questions

```
a student wrote:

> 1. In  your notes on 'testing restrictions and comparing models', you
> said g and h are conditional prior pdf's. what do you mean by
> 'conditional'? is that they are come up with without even knowing data?
```

Priors are *always* supposed to represent the uncertainty about the parameters before seeing – that is, "without even knowing" — data. I said g and h are conditional priors because each is by itself a pdf — g is the density function of the prior distribution *conditional* on θ lying in Θ, for example. The unconditional density of the prior is $\pi(\Theta)g$ within Θ and $\pi(\Phi)h$ within Φ. The unconditional density has to integrate to one over the entire parameter space, $\Theta \cup \Phi$, and so integrates to less than one within each piece of the parameter space.

```
> 2. In deriving (4) and (5), is the power to det(SIGMA) 1/2 instead of
> -1/2, by (3)?
```

You're right. They should be to the power plus 1/2. This sign error affects also equation (7), but not (for some reason) equation (9), which is the correct formula for the Schwarz criterion (but appears wrong if you try to derive it from (7).

```
> 3. I can't see why (9) converges to infinity or -infinity as you
> claimed. for a special case where Phi is a subspace of Theta (m>n), as T
> goes to infinity, the last term of (9) goes to -infinity , this requires
> the second term goes to infinity faster than it if Theta contains the
> true model. is this guaranteed?
```

If Θ contains the true model and Φ does not, it is easy to demonstrate the conclusion when the data are i.i.d. Then

$$\log(p(X \mid \hat{\theta})) = \sum_{t=1}^{T} \log(f(X_t \mid \hat{\theta}))$$

and under mild regularity conditions we will have

$$\frac{1}{T} \sum_{t=1}^{T} \log(f(X_t \mid \hat{\theta})) \xrightarrow[T \to \infty]{P} E[\log(f(X_t \mid \theta_0))] \, ,$$

where θ_0 is the true value of θ. It is a well known result that when $f(X_t \mid \theta_0)$ is the true pdf of X_t,

$$E[\log(w(X_t))] = \int \log(w(X_t)) f(X_t \mid \theta_0)\, dX_t$$

is maximized over all w's that are pdf's by setting $w(x) = f(x \mid \theta_0)$. From this it is fairly easy to see that the first term in (9) in the notes increases linearly in T when Θ contains the true value and Φ does not, which is enough to offset the second term even when it is going to $-\infty$ because $m > n$.

`4. another question about terminology in Bayesian analysis: when you say 'the integrals of the posterior pdf' (p2 just over (4)), i find it is indeed the integral of joint pdf of data and parameters. is it right?`

The sentence you are asking about is mistaken. What is displayed in (4) and (5) is the integrals of the joint pdf of data and parameters over the two pieces of the parameter space, Θ and Φ, with the data held fixed at the observed value. The ratio of these two integrals is the posterior odds ratio and the posterior probabilities on the two spaces are proportional to these two integrals. The integrals of the posterior pdf over the two spaces would also give the posterior odds ratios, but would give the posterior probabilities directly, instead of giving weights that have to be normalized to sum to one.