September 21, 2007

Bayesian comparison of alternative distribution models for the same data

by Ove Ditlevsen

Let a sample $x_1, \ldots, x_N$ from some distribution be given and consider two different probability density models $f_0(x; \alpha_1, \ldots, \alpha_m)$ with parameters $\alpha_1, \ldots, \alpha_m$ and $f_1(x; \beta_1, \ldots, \beta_n)$ with parameters $\beta_1, \ldots, \beta_n$, $n \geq m$, as possible distribution models for the sample. Then the first observation x_1 can be considered as drawn from the mixture $p f_0(x; \alpha_1, \ldots, \alpha_m) + (1 - p) f_1(x; \beta_1, \ldots, \beta_n)$, where p is the probability that x_1 comes from the first distribution. However, all the following observations are not drawn independently from the mixture, but independently from the distribution selected by the first observation. To keep track of this we introduce the indicator random variable I for which $P(I = 0) = p$ and $P(I = 1) = 1 - p$. Thus we have the conditional likelihood functions

$$L(\alpha_1, \ldots, \alpha_m \,|\, I = 0) = \prod_{i=1}^{N} f_0(x_i; \alpha_1, \ldots, \alpha_m) \tag{1}$$

$$L(\beta_1, \ldots, \beta_n \,|\, I = 1) = \prod_{i=1}^{N} f_1(x_i; \beta_1, \ldots, \beta_n) \tag{2}$$

The total likelihood function is then

$$L(p, \alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_n) = p L(\alpha_1, \ldots, \alpha_m \,|\, I = 0) + (1 - p) L(\beta_1, \ldots, \beta_n \,|\, I = 1) \tag{3}$$

The total likelihood is maximal at the parameter values where each of the two conditional likelihoods are maximal. If the two maximal values are equal, then p is undetermined in the interval $[0, 1]$. If the first likelihood is larger than the second, then total maximum is for $p = 1$, otherwise for $p = 0$.

However, it is not reasonable to let the α-parameters and the β-parameters vary independently of each other. A natural requirement is that the moments up to some order should be the same for the two distributions. Given that there is a one-one relation between the first m moments and the parameters $\alpha_1, \ldots, \alpha_m$ and also between the first m moments and m of the parameters $\beta_1, \ldots, \beta_n$, we can shape the problem such that $(\alpha_1, \ldots, \alpha_m) = (\beta_1, \ldots, \beta_m)$. This restriction changes the total likelihood (3) to

$$L(p, \alpha_1, \ldots, \alpha_m, \beta_{m+1}, \ldots, \beta_n)$$
$$= p L(\alpha_1, \ldots, \alpha_m \,|\, I = 0) + (1 - p) L(\alpha_1, \ldots, \alpha_m, \beta_{m+1}, \ldots, \beta_n \,|\, I = 1) \tag{4}$$

Maximum is obtained for

$$L(\alpha_1,\dots,\alpha_m\,|\,I=0)-L(\alpha_1,\dots,\alpha_m,\beta_{m+1},\dots,\beta_n\,|\,I=1)=0 \tag{5}$$

$$p\frac{\partial}{\partial\alpha_i}L(\alpha_1,\dots,\alpha_m\,|\,I=0)+(1-p)\frac{\partial}{\partial\alpha_i}L(\alpha_1,\dots,\alpha_m,\beta_{m+1},\dots,\beta_n\,|\,I=1)=0 \tag{6}$$

$$\frac{\partial}{\partial\beta_j}L(\alpha_1,\dots,\alpha_m,\beta_{m+1},\dots,\beta_n\,|\,I=1)=0 \tag{7}$$

$$i=1,\dots,m;\,j=m+1,\dots,n$$

given that there is a solution for p between 0 and 1. If $p=0$ or $p=1$ the first equation (5) will not be satisfied.

Instead of looking for maximum likelihood estimates let us consider the posterior distribution of the Bayesian random variable P (corresponding to p) in the case of the likelihood function (4) and with the prior distribution be given as the probability density

$$f_{P,\mathbf{A},\mathbf{B},\mathrm{prior}}(p,\alpha_1,\dots,\alpha_m,\beta_{m+1},\dots,\beta_n)$$
$$= f_{P,\mathrm{prior}}(p)\,f_{\mathbf{A},\mathrm{prior}}(\alpha_1,\dots,\alpha_m)\,f_{\mathbf{B},\mathrm{prior}}(\beta_{m+1},\dots,\beta_n) \tag{8}$$

using evident notation. Then

$$f_{P,\mathrm{posterior}}(p)/f_{P,\mathrm{prior}}(p)$$

$$\propto p\int_{\alpha_1=-\infty}^{\infty}\cdots\int_{\alpha_m=-\infty}^{\infty} L(\alpha_1,\dots,\alpha_m\,|\,I=0)\,f_{\mathbf{A},\mathrm{prior}}(\alpha_1,\dots,\alpha_m)\,\mathrm{d}\alpha_1\dots\mathrm{d}\alpha_m$$

$$+(1-p)\int_{\alpha_1=-\infty}^{\infty}\cdots\int_{\alpha_m=-\infty}^{\infty}\int_{\beta_{m+1}=-\infty}^{\infty}\cdots\int_{\beta_n=-\infty}^{\infty} L(\alpha_1,\dots,\alpha_m,\beta_{m+1},\dots,\beta_n\,|\,I=1)$$

$$\times f_{\mathbf{A},\mathrm{prior}}(\alpha_1,\dots,\alpha_m)\,f_{\mathbf{B},\mathrm{prior}}(\beta_{m+1},\dots,\beta_n)\,\mathrm{d}\alpha_1\dots\mathrm{d}\alpha_m\,\mathrm{d}\beta_{m+1}\dots\mathrm{d}\beta_n$$

$$= p\,E_{\mathbf{A},\mathrm{prior}}[L(\mathbf{A}\,|\,I=0)]+(1-p)\,E_{\mathbf{A},\mathbf{B},\mathrm{prior}}[L(\mathbf{A},\mathbf{B}\,|\,I=1)] \tag{9}$$

($\propto$ means "is proportional to") with the normalizing factor C given by

$$\frac{1}{C}=E_{P,\mathrm{prior}}[P]\,E_{\mathbf{A},\mathrm{prior}}[L(\mathbf{A}\,|\,I=0)]+(1-E_{P,\mathrm{prior}}[P])\,E_{\mathbf{A},\mathbf{B},\mathrm{prior}}[L(\mathbf{A},\mathbf{B}\,|\,I=1)] \tag{10}$$

Thus the posterior density of P is a trapez, implying that the largest posterior density is for $p=1$ if $E_{\mathbf{A},\mathrm{prior}}[L(\mathbf{A}\,|\,I=0)]>E_{\mathbf{A},\mathbf{B},\mathrm{prior}}[L(\mathbf{A},\mathbf{B}\,|\,I=1)]$ and for $p=0$, otherwise, except if the two expectations are equal.

Let us study what happens if the priors of $\mathbf{A}$ and $\mathbf{B}$ are chosen as non-informative priors. For this purpose it sufficient to study the situation for $m=1$ and $n=2$, and adopt a uniform prior over the interval $[-a/2,a/2]$ for A_1 and a uniform

prior over the interval $[-b/2, b/2]$ for B_2. Correspondingly assume that we have the regular case $J_0 = \int_{-\infty}^{\infty} L(\alpha \mid I = 0)\, d\alpha < \infty$, $J_1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} L(\alpha, \beta \mid I = 1)\, d\alpha\, d\beta < \infty$ such that the non-informative priors are allowable in the limit as $a \to \infty$, $b \to \infty$. Then

$$f_{P,\text{posterior}}(p) \approx \frac{p J_0 + (1-p) J_1/b}{E_{P,\text{prior}}[P] J_0 + (1 - E_{P,\text{prior}}[P]) J_1/b} f_{P,\text{prior}}(p) \qquad (11)$$

asymptotically as $b \to \infty$. Thus the posterior density depends on the arbitrary choice of b, and in the limit $b \to \infty$ we get

$$f_{P,\text{posterior}}(p) \approx \frac{p}{E_{P,\text{prior}}[P]} f_{P,\text{prior}}(p) \qquad (12)$$

i.e., a posterior density that only depends on the prior density of P and not on the sample. Obviously this result is unreasonable. This non-informative prior problem is not present if $m = n$, that is, if the parameters are the same in the two distribution models. Then (11) reads

$$f_{P,\text{posterior}}(p) = \frac{p J_0 + (1-p) J_1}{E_{P,\text{prior}}[P] J_0 + (1 - E_{P,\text{prior}}[P]) J_1} f_{P,\text{prior}}(p) \qquad (13)$$

where $J_0 = \int_{-\infty}^{\infty} L(\alpha \mid I = 1)\, d\alpha$, $J_1 = \int_{-\infty}^{\infty} L(\alpha \mid I = 0)\, d\alpha$. In particular taking $f_{P,\text{prior}}(p) = 1$ gives

$$f_{P,\text{posterior}}(p) = \frac{2[p J_0 + (1-p) J_1]}{J_0 + J_1} \qquad (14)$$

The maximal posterior density is obtained for $p = 1$ if $J_0/J_1 > 1$, and for $p = 0$ if $J_0/J_1 < 1$. If $J_0/J_1 = 1$ the posterior distribution of P is uniform, and the two models may be stated to fit the data equally well.