

Bayesian parametric testing for independent distributions with nuisance parameters

*Verifica di ipotesi parametrica Bayesiana per distribuzioni indipendenti in
presenza di parametri di disturbo*

Francesco Bertolino and Walter Racugno
Dipartimento di Matematica, Università di Cagliari
e-mail: bertolin@unica.it

Riassunto: Nel confronto tra parametri di due modelli campionari tra loro indipendenti, l'uso del fattore di Bayes come misura di evidenza dà luogo a problemi di diversa natura. In questo lavoro si evidenziano in particolare quei problemi che sorgono dalla riparametrizzazione del sistema di ipotesi e dall'uso di distribuzioni a priori non informative, mostrando come certe difficoltà possano essere affrontate utilizzando opportune pseudoverosimiglianze.

Keywords: Bayes factor, proper and improper priors, pseudo-likelihood.

1. Problem setting

Consider two independent populations with sampling models $X_i \sim f(\cdot|\theta_i)$, $X_i \in \mathcal{X}$, $\theta_i \in \Theta_i \subseteq \mathbb{R}^k$, assumed for simplicity belonging to the exponential family, with natural parameter θ_i , $i = 1, 2$. It is of interest to test the system of composite hypotheses $H_0 : \theta_1 = \theta_2$ vs $H_1 : \theta_1 \neq \theta_2$, using the Bayes factor (BF) $B_{10}(\mathbf{x}_1, \mathbf{x}_2) = m_1(\mathbf{x}_1, \mathbf{x}_2)/m_0(\mathbf{x}_1, \mathbf{x}_2)$, where the $m_j(\mathbf{x}_1, \mathbf{x}_2)$ are the predictive distributions of the i.i.d. data $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{in_i})$, $i = 1, 2$, conditional to the hypothesis H_j , $j = 0, 1$. The computation of the BF can be burdensome also when $k = 1$ and can be carried both in the presence and in the absence of prior informations. When prior informations are available, sufficient to elicitate proper priors $g_i(\theta_i)$, $i = 1, 2$, under H_1 the predictive distribution $m_1(\mathbf{x}_1, \mathbf{x}_2) = \prod_{i=1}^2 \int f(\mathbf{x}_i|\theta_i) g_i(\theta_i) d\theta_i$ can be computed and it is independent on the assumed parameterization. On the contrary, under the null hypothesis H_0 , which in the parameter space $\Theta_1 \times \Theta_2$ is represented by a set of null measures, the computation of $m_0(\mathbf{x}_1, \mathbf{x}_2)$ depends on the parameterization used and hence different inferential conclusions can arise. This ambiguity is known as the Borel-Kolmogorov paradox (Lindley, 1982).

Let us consider two transformations. In particular, let (a): $\{\delta = \theta_1 - \theta_2, \phi = \theta_2\}$, with $|J|^{(a)} = 1$, giving the system of hypotheses $H_0 : \delta = \mathbf{0}$ vs $H_1 : \delta \neq \mathbf{0}$, $\forall \phi \in \mathbb{R}^k$, where $\delta \in \mathbb{R}^k$ is the parameter of interest and $\phi \in \mathbb{R}^k$ is the nuisance parameter. Let (b): $\{\rho_h = \theta_{1h}/\theta_{2h}, h = 1, 2, \dots, k, \phi = \theta_2\}$, with $|J|^{(b)} = \prod_{h=1}^k |\phi_h|$, giving the system of hypotheses $H_0 : \rho = \mathbf{1}$ vs $H_1 : \rho \neq \mathbf{1}$, $\forall \phi \in \mathbb{R}^k$, where $\rho \in \mathbb{R}^k$ and ϕ are the interest and nuisance parameters, respectively. Under the transformation (a), the prior turns into $(\delta, \phi) \sim g_1(\delta + \phi)g_2(\phi)$, while under (b) we have $(\rho, \phi) \sim g_1(\rho_1\phi_1, \rho_2\phi_2, \dots, \rho_k\phi_k)g_2(\phi)\prod_{h=1}^k |\phi_h|$. In view of this, the priors of the nuisance parameter conditional to H_0 are $\phi|\delta = \mathbf{0} \sim h_a(\phi) \propto \prod_{i=1}^2 g_i(\phi)$ and $\phi|\rho = \mathbf{1} \sim h_b(\phi) \propto \prod_{i=1}^2 g_i(\phi) \prod_{h=1}^k |\phi_h|$, and, under H_0 , we have the predictive distributions $m_0^{(j)}(\mathbf{x}_1, \mathbf{x}_2) = \int \prod_{i=1}^2 f(\mathbf{x}_i | \phi)$

$h_j(\phi) d\phi$, $j = a, b$, which depend on the parameterization. Hence the ambiguity which can produce different inferential conclusions.

In order to use the BF when there are no sufficient prior informations, the *reference priors* $g_i^N(\theta_i) = c_i h_i(\theta_i)$, where the c_i are normalizing *pseudo*-constants, can be used. Since the BF depends on the ratio c_1/c_2 , any inferential conclusion is allowed. In this case, it is possible to resort to the intrinsic and fractional procedures and to compute the corresponding Bayes factors, IBF and FBF, respectively (O'Hagan, 1994, Berger and Pericchi, 1996, Moreno *et al.*, 1998). The FBF is simple to compute, but it depends on the assumed parameterization. On the other hand, the IBF does not depend on the transformation but it can be more burdensome to compute.

To alleviate some of the problems related to the computation of the intrinsic procedure, we suggest to replace the likelihood functions $f(\mathbf{x}_i|\theta_i)$, $i = 1, 2$, with suitable *pseudo*-likelihoods, such as the marginal, conditional or profile, which do not depend on the nuisance parameters. A well known example on the use of a pseudo-likelihood in the frequentist setting is the exact Fisher's test to compare two models belonging to the exponential family, with natural parameters θ_i , $i = 1, 2$. Under the parameterization (a) the conditional distribution of U_1 given $T = t$, with $U_i = \sum_{r=1}^{n_i} X_{ir}$, $i = 1, 2$ and $T = U_1 + U_2$, is given by $f_{U_1|T}(u|t, \delta)$ and it depends only on the parameter of interest. In this case, the composite null hypothesis simplifies into the simple one $H_0 : \delta = 0$ (see Cox-Hinkley, 1974, Section 5.2). Using the conditional likelihood $f_{U_1|T}(u|t, \delta)$, the conditional Bayes factor (CBF) can be computed. The proper prior of δ can be derived by marginalization of the proper priors g_i . By construction, the CBF $B_{10}^C(u_1|t) = m_1(u_1|t)/m_0(u_1|t)$ is unique.

2. Example: Comparison of normal distributions

Given two normal independent populations $X_i \sim N(x|\mu_i, 1)$, $X_i \in \mathbb{R}$, $\mu_i \in \mathbb{R}$, $i = 1, 2$, let us assume for the natural parameters μ_i the conjugate priors $\mu_i \sim N(\mu_i|\eta_i, \lambda_i^{-1})$, $i = 1, 2$, where (η_i, λ_i) are known hyperparameters. It is of interest to verify the system of hypotheses $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$. Let us assume the data $\mathbf{x}_i$ be i.i.d., and let $u_i = \sum_{r=1}^{n_i} x_{ir}$, $\bar{x}_i = n_i^{-1}u_i$, $t = u_1 + u_2$, $\bar{\bar{x}} = (n_1\bar{x}_1 + n_2\bar{x}_2)/(n_1 + n_2)$. The sufficient statistics U_i are such that $U_i \sim N(u_i|n_i\mu_i, n_i)$, $i = 1, 2$. Under H_1 , the predictive distribution of (u_1, u_2) is $m_1(u_1, u_2) = \prod_{i=1}^2 n_i^{-1} N(\bar{x}_i|\eta_i, (\lambda_i + n_i)/(n_i \lambda_i))$. Using the parameterization $\{\delta = \mu_1 - \mu_2; \psi = \mu_2\}$, with $|J| = 1$, the conditional prior of ψ given $H_0 : \delta = 0$ can be derived, giving $\psi|\delta = 0 \sim \frac{N(\psi|\eta_1, \lambda_1^{-1})N(\psi|\eta_2, \lambda_2^{-1})}{N(0|\eta_1 - \eta_2, \tilde{\lambda}^{-1})}$, with $\tilde{\lambda} = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2}$. In this case, the predictive distribution is given by $m_0(u_1, u_2) = (n_1 n_2)^{-1} N(\bar{\bar{x}}|\tilde{\eta}_0, (\lambda_1 + \lambda_2)^{-1} + (n_1 + n_2)^{-1}) N(\bar{x}_1 - \bar{x}_2|0, \tilde{n}^{-1})$, with $\tilde{\eta}_0 = \frac{\eta_1 \lambda_1 + \eta_2 \lambda_2}{\lambda_1 + \lambda_2}$ and $\tilde{n} = \frac{n_1 n_2}{n_1 + n_2}$. The genuine FB is then $B_{10}(u_1, u_2) = \frac{N(\bar{x}_1|\eta_1, \frac{\lambda_1 + n_1}{n_1 \lambda_1}) N(\bar{x}_2|\eta_2, \frac{\lambda_2 + n_2}{n_2 \lambda_2})}{N(\bar{\bar{x}}|\tilde{\eta}_0, (\lambda_1 + \lambda_2)^{-1} + (n_1 + n_2)^{-1}) N(\bar{x}_1 - \bar{x}_2|0, \tilde{n}^{-1})}$.

Focusing attention on a pseudo-likelihood for the parameter of interest $\delta = \mu_1 - \mu_2$, both the marginal and the conditional likelihoods can be used. From the elicited priors the prior of δ can be derived, that is $\delta \sim N(\delta|\eta_1 - \eta_2, \tilde{\lambda}^{-1})$. Since $\bar{X}_1 - \bar{X}_2 \sim N(\bar{x}_1 - \bar{x}_2|\delta, \tilde{n}^{-1})$, the marginal Bayes factor (MBF) of H_1 against the simple hypothesis

$$H_0 \text{ is } B_{10}^M(\bar{x}_1 - \bar{x}_2) = \frac{N(\bar{x}_1 - \bar{x}_2 | \eta_1 - \eta_2, \tilde{\lambda}^{-1} + \tilde{n}^{-1})}{N(\bar{x}_1 - \bar{x}_2 | 0, \tilde{n}^{-1})}.$$

The computation of the CBF is simple. Indeed, we have $U_1 | T = t \sim N(u_1 | \tilde{n} \delta + n_1 \bar{x}, \tilde{n})$ and, under H_1 , the predictive distribution of U_1 given $T = t$ is $m_1(u_1 | t) = N(u_1 | \tilde{n}(\eta_1 - \eta_2) + n_1 \bar{x}, \tilde{n}(\tilde{\lambda}^{-1} + 1))$. In view of this, $B_{10}^C(u_1 | t) = \frac{N(u_1 | \tilde{n}(\eta_1 - \eta_2) + n_1 \bar{x}, \tilde{n}(\tilde{\lambda}^{-1} + 1))}{N(u_1 | n_1 \bar{x}, \tilde{n})}$, and it is easy to show that $B_{10}^M(\bar{x}_1 - \bar{x}_2) = B_{10}^C(u_1 | t)$.

In the absence of informative priors, the *reference priors* $g_i^N(\mu_i) = c_i \mathbf{1}_{\mathbb{R}}(\mu_i)$, with c_i *pseudo-constants*, can be assumed and the intrinsic procedure can be used. Moreover, it is possible to reformulate the test as a comparison of nested models $\mathcal{M}_0^k : \{X_1, X_2 \sim N(x_1 | \theta, 1)N(x_2 | \theta, 1), g_0^k(\theta)\}$, $\mathcal{M}_1 : \{X_1, X_2 \sim N(x_1 | \mu_1, 1)N(x_2 | \mu_2, 1), g^N(\mu_1, \mu_2)\}$, where $g_0^k(\theta) = \frac{1}{2k} \mathbf{1}_{(-k, k)}(\theta)$ and $g^N(\mu_1, \mu_2) = c \mathbf{1}_{\mathbb{R} \times \mathbb{R}}(\mu_1, \mu_2)$ are the proper and improper *default priors*, with $k > 0$ arbitrary (Moreno *et al.*, 1998). Since the *minimal training sample* is (x_1, x_2) , it follows that $m_1^N(x_1, x_2) = c$ and that $m_0^k(x_1, x_2) = \frac{1}{4k\pi} \int_{-k}^k \exp\{-\frac{1}{2}[(x_1 - \theta)^2 + (x_2 - \theta)^2]\} d\theta$. Let $B_{01}^{k,N}(x_1, x_2) = c^{-1} m_0^k(x_1, x_2)$ be the BF corresponding to the *training sample*. Under $\mathcal{M}_1$, it has mean $E_{\mu_1, \mu_2}^{\mathcal{M}_1} B_{01}^{k,N}(X_1, X_2)$ and the proper intrinsic prior can be derived for $\mathcal{M}_1$, given by $g^{k,I}(\mu_1, \mu_2) = g^N(\mu_1, \mu_2) E_{\mu_1, \mu_2}^{\mathcal{M}_1} B_{01}^{k,N}(X_1, X_2)$, which does not depend on the *pseudo-constant* c . Fixed k , the BF associated to (u_1, u_2) is $B_{10}^{k,I}(u_1, u_2) = m_1^{k,I}(u_1, u_2) / m_0^k(u_1, u_2)$ (see Moreno *et al.*, 1998). The IBF can be derived considering $B_{10}^I(u_1, u_2) = \lim_{k \rightarrow \infty} B_{10}^{k,I}(u, v) = \lim_{k \rightarrow \infty} \frac{m_1^{k,I}(u, v)}{m_0^k(u, v)} = \sqrt{\frac{1}{4\tilde{n} + 1}} \exp\{\frac{2\tilde{n}^2}{4\tilde{n} + 1}(\bar{x}_1 - \bar{x}_2)^2\}$.

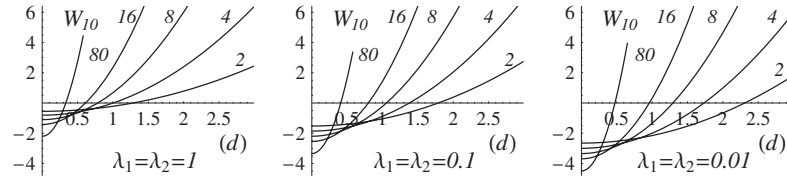


Figure 1: $W_{10}(u_1, u_2)$ with respect to $d = \bar{x}_1 - \bar{x}_2$, for $n = 2, 4, 8, 16, 80$ and some λ_i .

The figures focus on the weight of evidence $W_{10} = \log B_{10}$. Figure 1 shows the weights of evidence $W_{10}(u_1, u_2)$ with respect to the difference $d = \bar{x}_1 - \bar{x}_2$ (with $\bar{x}_2 = \frac{1}{2}$ and $n_1 = n_2 = n$), with the choice $\eta_1 = \eta_2 = \frac{1}{2}$ and different values of $\lambda_1 = \lambda_2$. As the λ_i decrease, from concentrated priors to flat priors, the test appears more conservative.

Figure 2 shows the behavior of $W_{10}(u_1, u_2)$ for n increasing, with $n_1 = n_2 = n$, and for some values of $d = \bar{x}_1 - \bar{x}_2$, with $\bar{x}_2 = \frac{1}{2}$, and for fixed $\eta_1 = \eta_2 = \frac{1}{2}$ and $\lambda_1 = \lambda_2 = \frac{1}{10}$. It can be noted that for $d = 0.35$ there is more evidence for the hypothesis H_0 also for large sample sizes. On the contrary, for $d = 1.05$ there is more evidence for the hypothesis H_1 also for small values of n . The weights of evidence $W_{10}^M(\bar{x}_1 - \bar{x}_2)$, $W_{10}^C(u_1 | t)$ and $W_{10}(u_1, u_2)$ agree.

Finally, Figure 3 shows the intrinsic weight of evidence $W_{10}^I(u_1, u_2)$ with respect to $d = \bar{x}_1 - \bar{x}_2$ for fixed $\bar{x}_2 = \frac{1}{2}$. It is few conservative and it differs slightly from $W_{10}(u_1, u_2)$

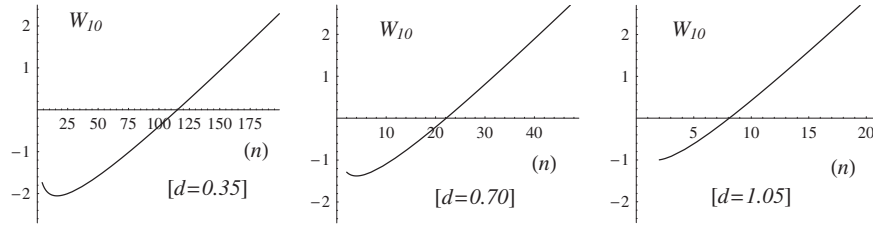


Figure 2: Weights of evidence $W_{10}(u_1, u_2)$ with respect to n and for $d = 0.35, 0.70, 1.05$.

derived with proper prior few concentrated. The curves correspond to $n_1 = 2, 4, 8, 16, 32$; in all we have $n_2 = 3 n_1$.

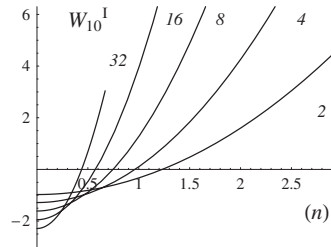


Figure 3: Weight of evidence $W_{10}^I(u_1, u_2)$, with respect to $d = \bar{x}_1 - \bar{x}_2$, for some n_1 .

As a final comment, we remark that in hypothesis testing problems for the natural parameters of two independent populations, the marginal, the conditional and the genuine BF's provide the same evidence when the parameterization (a) is used. Moreover, in the presence of proper flat priors, a good agreement with the intrinsic BF can be observed. These results can be generalized to other different kind of problems, when the natural parameterization is considered. The use of different parameterizations leads to the Borel-Kolmogorov paradox, to whom appears to be insensible the IBF.

References

- Berger, J.O., Pericchi, L.R. (1996) The intrinsic Bayes factor for model selection and prediction, *Journal of the American Statistical Association*, 91, 109–122.
- Bertolino, F., Ratto, A. (2003) On the Bayesian inference of the Hardy-Weinberg equilibrium model, in *Applied Bayesian Statistical Studies in Biology and Medicine*, Kluwer Academic Publishers, Norwell.
- Cox, D.R., Hinkley, D.V. (1974) *Theoretical Statistics*, Chapman & Hall, London.
- Lindley, D.V. (1982) The Bayesian approach to statistics, in *Some Recent Advances in Statistics*, eds. T. de Oliveira and B. Epstein, 65–87, Academic Press, New York.
- Moreno, E., Bertolino, F., Racugno, W. (1998) An intrinsic limiting procedure for model selection and hypotheses testing, *Journal of the American Statistical Association*, 93, 1451–1460.
- O'Hagan, A. (1994) *Kendall's Theory of Advanced Statistics*, 2B, Arnold, London.