Supplement to “Model fusion and multiple testing under the laws of likelihood: Shrinkage and evidence supporting a point null hypothesis”

February 9, 2015

David R. Bickel

Ottawa Institute of Systems Biology

Department of Biochemistry, Microbiology, and Immunology

Department of Mathematics and Statistics

University of Ottawa; 451 Smyth Road; Ottawa, Ontario, K1H 8M5

Abstract

This document consists of a revision of Sections 4 and 5.2 of Bickel (2014c) with additional references.

4 Fusing models from different levels

The procedure chosen for eliminating nuisance parameters may depend on criteria outside the statistical model (Bickel, 2012). Indeed, paradoxes arise from likelihood inference under inappropriate models of the system of interest (Lindsey, 1996, §6.5). The level of abstraction in a model must be selected with the use of the model in mind. In complex problems such as large-scale inference, multiple levels of abstraction may be used together to weigh the evidence. This section applies two levels of abstraction to the problem of testing multiple hypotheses.

Considering some function φ and a parameter set $\Phi = \{\varphi(\theta) : \theta \in \Theta\}$, let $\mathfrak{I}$ denote the σ-field of subsets of Φ satisfying $\mathfrak{I} = \{\{\varphi(\theta) : \theta \in \mathcal{H}\} : \mathcal{H} \in \mathfrak{H}\}$. For every $\phi \in \Phi$, let g_ϕ be a probability density function on a set $\mathcal{Z}$ of possible values of an observable random variable Z of observed value z. The triple $(\{f_\theta : \theta \in \Theta\}, \{g_\phi : \phi \in \Phi\}, \mathfrak{Q})$ is called a *fusion* of the two parametric models, the density families $\{f_\theta : \theta \in \Theta\}$ and $\{g_\phi : \phi \in \Phi\}$, with respect to a partition $\mathfrak{Q} \subset \mathfrak{I}$ of Φ. A model used to quantify the weight of evidence without any fusion (§2) is a *pure model*.

The function $L^f : \mathfrak{H} \times \mathfrak{H} \to [0, \infty]$ is defined in accordance with equations (1)-(2) such that $L^f(\mathcal{H}|\mathcal{R}) = L(\mathcal{H}|\mathcal{R})$ for all $\mathcal{H}, \mathcal{R} \in \mathfrak{H}$. In analogy with equation (1), the function $L^g : \mathfrak{I} \times \mathfrak{I} \to [0, \infty]$ is defined such that

$$L^g(\mathcal{I}|\mathcal{S}) = \frac{\sup_{\phi \in \mathcal{I} \cap \mathcal{S}} g_\phi(z)}{\sup_{\phi \in \mathcal{S}} g_\phi(z)}$$

for all $\mathcal{I}, \mathcal{S} \in \mathfrak{I}$. For any $\mathcal{I} \in \mathfrak{I}$, define the function φ^{-1} such that

$$\varphi^{-1}(\mathcal{I}) = \{\theta \in \Theta : \varphi(\theta) \in \mathcal{I}\}.$$

Let $L^{fg} : \mathfrak{H} \times \mathfrak{I} \times \mathfrak{I} \rightarrow [0, \infty]$ denote the function satisfying

$$L^{fg} (\mathcal{H}, \mathcal{I} | \mathcal{S}) = L^{f} \left(\mathcal{H} | \varphi^{-1} (\mathcal{I} \cap \mathcal{S}) \right) L^{g} (\mathcal{I} | \mathcal{S}) , \tag{16}$$

which reduces to $L^{f} (\mathcal{H} \cap \mathcal{I} | \mathcal{S})$ according to equation (2) whenever $\varphi (\theta) = \theta$ and $g_{\theta} = f_{\theta}$ for all $\theta \in \Theta$. With $L^{fg} (\bullet, \bullet) = L^{fg} (\bullet, \bullet | \Phi)$, $L^{f} (\bullet) = L^{f} (\bullet | \Theta)$, and $L^{g} (\bullet) = L^{g} (\bullet | \Phi)$, equation (16) degenerates to

$$L^{fg} (\mathcal{H}, \mathcal{I}) = L^{f} \left(\mathcal{H} | \varphi^{-1} (\mathcal{I}) \right) L^{g} (\mathcal{I}) . \tag{17}$$

Likewise generalizing equation (7), let

$$L_{\mathfrak{Q}}^{fg} (\mathcal{H} | \mathcal{S}) = \sup_{\mathcal{I} \in \mathfrak{Q}} L^{fg} (\mathcal{H}, \mathcal{I} | \mathcal{S}) , \tag{18}$$

and let $L_{\mathfrak{Q}}^{fg} (\mathcal{H}) = L_{\mathfrak{Q}}^{fg} (\mathcal{H} | \Phi)$.

The corresponding conditional weight of evidence in the observation that $X = x$ and $Z = z$ substantiating the hypothesis that $\theta \in \mathcal{H}_1$ as opposed to the hypothesis that $\theta \in \mathcal{H}_2$ is generalized to

$$W_{\mathfrak{Q}}^{fg} (\mathcal{H}_1 ; \mathcal{H}_2 | \mathcal{S}) = \frac{L_{\mathfrak{Q}}^{fg} (\mathcal{H}_1 | \mathcal{S})}{L_{\mathfrak{Q}}^{fg} (\mathcal{H}_2 | \mathcal{S})} , \tag{19}$$

and that substantiating the hypothesis that $\theta \in \mathcal{H}$ given $\varphi (\theta) \in \mathcal{S}$ to

$$W_{\mathfrak{Q}}^{fg} (\mathcal{H} | \mathcal{S}) = W_{\mathfrak{Q}}^{fg} \left(\mathcal{H} ; \overline{\mathcal{H}} | \mathcal{S} \right) \tag{20}$$

for all $\mathcal{H}, \mathcal{H}_1, \mathcal{H}_2 \in \mathfrak{H}$ and $\mathcal{S} \in \mathfrak{I}$. The marginal counterparts are $W_{\mathfrak{Q}}^{fg} (\bullet, \bullet) = W_{\mathfrak{Q}}^{fg} (\bullet, \bullet | \Phi)$ and $W_{\mathfrak{Q}}^{fg} (\bullet) = W_{\mathfrak{Q}}^{fg} (\bullet | \Phi)$. Analogously, W^{f} is identified with the functions denoted by W in Sections 2.

Theorem 3. *For a partition $\mathfrak{Q} \subset \mathfrak{I}$ of Φ and any $\mathcal{H} \in \mathfrak{H}$,*

$$L_{\mathfrak{Q}}^{fg}(\mathcal{H}) = \sup_{\mathcal{I} \in \mathfrak{Q}} L^{f}\left(\mathcal{H}|\varphi^{-1}(\mathcal{I})\right) L^{g}(\mathcal{I}); \tag{21}$$

$$W_{\mathfrak{Q}}^{fg}(\mathcal{H}) = \frac{\sup_{\mathcal{I} \in \mathfrak{Q}} L^{f}(\mathcal{H}|\varphi^{-1}(\mathcal{I})) L^{g}(\mathcal{I})}{\sup_{\mathcal{I} \in \mathfrak{Q}} L^{f}\left(\overline{\mathcal{H}}|\varphi^{-1}(\mathcal{I})\right) L^{g}(\mathcal{I})}. \tag{22}$$

Proof. Equation (18) yields

$$L_{\mathfrak{Q}}^{fg}(\mathcal{H}) = \sup_{\mathcal{I} \in \mathfrak{Q}} L_{\mathfrak{Q}}^{fg}(\mathcal{H}, \mathcal{I}),$$

which, with equation (17), in turn yields equation (21). Equation (22) follows immediately from equations (19)-(20). □

Corollary 4. *$W_{\mathfrak{Q}}^{fg}$ is logically coherent.*

Proof. For all $\mathcal{I} \in \mathfrak{Q}$,

$$L^{f}\left(\mathcal{H}|\varphi^{-1}(\mathcal{I})\right) L^{g}(\mathcal{I}) = \sup_{\theta \in \mathcal{H}} L^{f}\left(\{\theta\}|\varphi^{-1}(\mathcal{I})\right) L^{g}(\mathcal{I})$$

$$L^{f}\left(\mathcal{H}_0|\varphi^{-1}(\mathcal{I})\right) L^{g}(\mathcal{I}) \le L^{f}\left(\mathcal{H}_1|\varphi^{-1}(\mathcal{I})\right) L^{g}(\mathcal{I}) \iff (\theta \in \mathcal{H}_0 \cap \mathcal{R} \implies \theta \in \mathcal{H}_1 \cap \mathcal{R}).$$

Thus, equation (12) is satisfied for $v = L_{\mathfrak{Q}}^{fg}$ according to equation (21) and, by implication, is also satisfied for $v = W_{\mathfrak{Q}}^{fg}$ according to equation (22). □

Properties analogous to those of Theorem 3 and Corollary 4 may be proven conditional on $\varphi(\theta) \in \mathcal{S}$.

5.2 Pure model for multiple-hypothesis evidence

Again consider some restriction set $\boldsymbol{\mathcal{R}} \in \boldsymbol{\mathfrak{H}}$. According to equation (8), the joint weight of evidence substantiating the alternative hypotheses that $\theta_j \neq \theta_0$ for all $j \in \mathcal{T}$, given that $\boldsymbol{\theta} \in \boldsymbol{\mathcal{R}}$, is

$$W\left(\left\{\theta_j \neq \theta_0 : j \in \mathcal{T}\right\} | \boldsymbol{\mathcal{R}}\right) \equiv W\left(\boldsymbol{\mathcal{H}}_{\bigwedge \mathcal{T}} | \boldsymbol{\mathcal{R}}\right) = \frac{\sup_{\boldsymbol{\theta} \in \boldsymbol{\mathcal{H}}_{\bigwedge \mathcal{T}} \cap \boldsymbol{\mathcal{R}}} f_{\boldsymbol{\theta}}\left(x\right)}{\sup_{\boldsymbol{\theta} \in \boldsymbol{\mathcal{R}} \setminus \boldsymbol{\mathcal{H}}_{\bigwedge \mathcal{T}}} f_{\boldsymbol{\theta}}\left(x\right)}; \tag{23}$$

$$\boldsymbol{\mathcal{H}}_{\bigwedge \mathcal{T}} = \left\{\left(\theta_1, \ldots, \theta_N\right) \in \Theta^N : \forall j \in \mathcal{T}\, \theta_j \neq \theta_0\right\} \tag{24}$$

(cf. Bickel (2012, §2.4.1)). Similarly, the marginal weight of evidence substantiating the alternative hypotheses that $\theta_j \neq \theta_0$ for some $j \in \mathcal{T}$, given that $\boldsymbol{\theta} \in \boldsymbol{\mathcal{R}}$, is

$$W\left(\left\{\exists j \in \mathcal{T}\, \theta_j \neq \theta_0\right\} | \boldsymbol{\mathcal{R}}\right) \equiv W\left(\boldsymbol{\mathcal{H}}_{\bigvee \mathcal{T}} | \boldsymbol{\mathcal{R}}\right) = \frac{\sup_{\boldsymbol{\theta} \in \boldsymbol{\mathcal{H}}_{\bigvee \mathcal{T}} \cap \boldsymbol{\mathcal{R}}} f_{\boldsymbol{\theta}}\left(x\right)}{\sup_{\boldsymbol{\theta} \in \boldsymbol{\mathcal{R}} \setminus \boldsymbol{\mathcal{H}}_{\bigvee \mathcal{T}}} f_{\boldsymbol{\theta}}\left(x\right)};$$

$$\boldsymbol{\mathcal{H}}_{\bigvee \mathcal{T}} = \left\{\left(\theta_1, \ldots, \theta_N\right) \in \Theta^N : \exists j \in \mathcal{T}\, \theta_j \neq \theta_0\right\}.$$

Theorem 4. *Suppose that T_i is independent of T_j for all $i \neq j$. It follows that, with probability 1, the joint weight of evidence substantiating the joint alternative hypotheses that $\theta_j \neq \theta_0$ for all $j \in \mathcal{T}$ is*

$$W\left(\boldsymbol{\mathcal{H}}_{\bigwedge \mathcal{T}}\right) = \min_{j \in \mathcal{T}} L_j^{\max} \tag{25}$$

and that the marginal weight of evidence substantiating the alternative hypotheses that $\theta_j \neq \theta_0$ for some $j \in \mathcal{T}$ is

$$W\left(\boldsymbol{\mathcal{H}}_{\bigvee \mathcal{T}}\right) = \prod_{j \in \mathcal{T}} L_j^{\max}. \tag{26}$$

Proof. Let $J\left(\mathcal{T}\right) = \arg\min_{j \in \mathcal{T}} L_j^{\max}$. The following statements hold with probability 1. By equation (23) and the fact that $L_i^{\max} > 1$ due to the continuity of the sample space and parameter space,

$$W\left(\boldsymbol{\mathcal{H}}_{\bigwedge\mathcal{T}}\right) = W\left(\boldsymbol{\mathcal{H}}_{\bigwedge\mathcal{T}}|\Theta^N\right) = \frac{\sup_{(\theta_1,\dots,\theta_N)\in\boldsymbol{\mathcal{H}}_{\bigwedge\mathcal{T}}}\prod_{i=1}^N f_{\theta_i}\left(x_i\right)}{\sup_{(\theta_1,\dots,\theta_N)\in\Theta^N\setminus\boldsymbol{\mathcal{H}}_{\bigwedge\mathcal{T}}}\prod_{i=1}^N f_{\theta_i}\left(x_i\right)} = \frac{\prod_{i\in\{1,\dots,N\}} f_{\widehat{\theta}(x_i)}\left(x_i\right)}{f_{\theta_0}\left(x_{J(\mathcal{T})}\right)\prod_{i\in\{1,\dots,N\}\setminus\{J(\mathcal{T})\}} f_{\widehat{\theta}(x_i)}\left(x_i\right)},$$

and cancellation of the products yields $W\left(\boldsymbol{\mathcal{H}}_{\bigwedge\mathcal{T}}\right) = f_{\widehat{\theta}\left(x_{J(\mathcal{T})}\right)}\left(x_{J(\mathcal{T})}\right)/f_{\theta_0}\left(x_{J(\mathcal{T})}\right)$ and thus equation (25). Similarly,

$$W\left(\boldsymbol{\mathcal{H}}_{\bigvee\mathcal{T}}\right) = \frac{\sup_{(\theta_1,\dots,\theta_N)\in\boldsymbol{\mathcal{H}}_{\bigvee\mathcal{T}}}\prod_{i=1}^N f_{\theta_i}\left(x_i\right)}{\sup_{(\theta_1,\dots,\theta_N)\in\Theta^N\setminus\boldsymbol{\mathcal{H}}_{\bigvee\mathcal{T}}}\prod_{i=1}^N f_{\theta_i}\left(x_i\right)} = \frac{\prod_{i\in\{1,\dots,N\}} f_{\widehat{\theta}(x_i)}\left(x_i\right)}{\prod_{i\in\mathcal{T}} f_{\theta_0}\left(x_{J(\mathcal{T})}\right)\prod_{i\in\{1,\dots,N\}\setminus\mathcal{T}} f_{\widehat{\theta}(x_i)}\left(x_i\right)},$$

leading to equation (26) by cancellation. □

Since Theorem 4 holds for $N = 1$ as well as $N \geq 2$, there is continuity in the evaluation of evidence from a single comparison to as many comparisons as needed. This contrasts with the sharp discontinuity seen in applications of completely different frequentist methods to problems involving very different numbers of comparisons (Bickel, 2011, §6). For example, there is a discrepancy in the practice of applying p-values to single comparisons but estimates of local false discovery rates to large numbers of comparisons (Westfall, 2010; Efron, 2010); for other solutions, see Bickel (2015, 2014a).

References

Barnard, G. A., 1967. The use of the likelihood function in statistical practice. Proc. 5th Berkeley Symp. on Math. Stat. Prob. Vol. I, 27–40.

Barndorff-Nielsen, O. E., 1994. Adjusted versions of profile likelihood and directed likelihood, and extended likelihood. Journal of the Royal Statistical Society B 56, 125–140.

Bickel, D. R., 2011. A predictive approach to measuring the strength of statistical evidence for single and multiple comparisons. Canadian Journal of Statistics 39, 610–631.

Bickel, D. R., 2012. The strength of statistical evidence for composite hypotheses: Inference to the best explanation. Statistica Sinica 22, 1147–1198.

Bickel, D. R., 2013a. Minimax-optimal strength of statistical evidence for a composite alternative hypothesis. International Statistical Review 81, 188–206.

Bickel, D. R., 2013b. Pseudo-likelihood, explanatory power, and Bayes's theorem [comment on "A likelihood paradigm for clinical trials"]. Journal of Statistical Theory and Practice 7, 178–182.

Bickel, D. R., 2014a. A fiducial continuum from confidence sets to empirical Bayes set estimates as the number of comparisons increases. Working Paper, University of Ottawa, deposited in uO Research at http://hdl.handle.net/10393/31898.

Bickel, D. R., 2014b. Inference after checking multiple Bayesian models for data conflict. Working Paper, University of Ottawa, deposited in uO Research at http://hdl.handle.net/10393/31135.

Bickel, D. R., 2014c. Model fusion and multiple testing in the likelihood paradigm: Shrinkage and evidence supporting a point null hypothesis. Working Paper, University of Ottawa, deposited in uO Research at http://hdl.handle.net/10393/31897.

Bickel, D. R., 2014d. Small-scale inference: Empirical Bayes and confidence methods for as few as a single comparison. International Statistical Review 82, 457–476.

Bickel, D. R., 2015. Blending Bayesian and frequentist methods according to the precision of prior information with applications to hypothesis testing. Statistical Methods and Applications (accepted 30 January 2015).

Bjørnstad, J. F., 1996. On the generalization of the likelihood function and the likelihood principle. Journal of the American Statistical Association 91, 791–806.

Blume, J., 2013. Likelihood and composite hypotheses [comment on "A likelihood paradigm for clinical trials"]. Journal of Statistical Theory and Practice 7 (2), 183–186.

Blume, J. D., 2011. Likelihood and its evidential framework. In: Bandyopadhyay, P. S., Forster, M. R. (Eds.), Philosophy of Statistics. North Holland, Amsterdam, pp. 493–512.

Coletti, G., Scozzafava, R., Vantaggi, B., 2009. Integrated likelihood in a finitely additive setting. In: Symbolic and Quantitative Approaches to Reasoning with Uncertainty. Vol. 5590 of Lecture Notes in Comput. Sci. Springer, Berlin, pp. 554–565.

Dubois, D., Moral, S., Prade, H., Jan. 1997. A semantics for possibility theory based on likelihoods. Journal of Mathematical Analysis and Applications 205 (2), 359–380.

Edwards, A. W. F., 1992. Likelihood. Johns Hopkins Press, Baltimore.

Efron, B., 2010. Rejoinder to comments on B. Efron, "Correlated z-values and the accuracy of large-scale statistical estimates". Journal of the American Statistical Association 105, 1067–1069.

Fisher, R. A., 1973. Statistical Methods and Scientific Inference. Hafner Press, New York.

Giang, P. H., Shenoy, P. P., 2005. Decision making on the sole basis of statistical likelihood. Artificial Intelligence 165, 137–163.

Hoch, J. S., Blume, J. D., 2008. Measuring and illustrating statistical evidence in a cost-effectiveness analysis. Journal of Health Economics 27, 476–495.

Jeffreys, H., 1948. Theory of Probability. Oxford University Press, London.

Korn, E. L., Freidlin, B., 2006. The likelihood as statistical evidence in multiple comparisons in clinical trials: No free lunch. Biometrical Journal 48, 346–355.

Lavine, M., Schervish, M. J., 1999. Bayes factors: What they are and what they are not. American Statistician 53, 119–122.

Lindsey, J., 1996. Parametric Statistical Inference. Oxford Science Publications. Clarendon Press, Oxford.

Morgenthaler, S., Staudte, R. G., 2012. Advantages of Variance Stabilization. Scandinavian Journal of Statistics 39 (4), 714–728.

Muralidharan, O., 2010. An empirical Bayes mixture method for effect size and false discovery rate estimation. Annals of Applied Statistics 4, 422–438.

Padilla, M., Bickel, D. R., 2012. Estimators of the local false discovery rate designed for small numbers of tests. Statistical Applications in Genetics and Molecular Biology 11 (5), art. 4.

Patriota, A. G., 2013. A classical measure of evidence for general null hypotheses. Fuzzy Sets and Systems 233, 74 – 88.

Pawitan, Y., 2001. In All Likelihood: Statistical Modeling and Inference Using Likelihood. Clarendon Press, Oxford.

Pawitan, Y., Murthy, K., Michiels, S., Ploner, A., 2005. Bias in the estimation of false discovery rate in microarray studies. Bioinformatics 21, 3865–3872.

Rohde, C., 2014. Pure likelihood methods. In: Introductory Statistical Inference with the Likelihood Function. Springer International Publishing, pp. 197–209.

Royall, R., 1997. Statistical Evidence: A Likelihood Paradigm. CRC Press, New York.

Royall, R., 2000. On the probability of observing misleading statistical evidence (with discussion). Journal of the American Statistical Association 95, 760–780.

Rubin, D. B., 1981. Estimation in parallel randomized experiments. Journal of Educational Statistics 6, pp. 377–401.

Schervish, M. J., 1996. P values: What they are and what they are not. American Statistician 50, 203–206.

Severini, T., 2000. Likelihood Methods in Statistics. Oxford University Press, Oxford.

Spanos, A., 2013. Revisiting the likelihoodist evidential account [comment on "A likelihood paradigm for clinical trials"]. Journal of Statistical Theory and Practice 7 (2), 187–195.

Spohn, W., 2012. The Laws of Belief: Ranking Theory and Its Philosophical Applications. Oxford University Press.

Sprott, D. A., 2000. Statistical Inference in Science. Springer, New York.

Strug, L. J., Hodge, S. E., 2006. An alternative foundation for the planning and evaluation of linkage analysis i. decoupling ’error probabilities’ from ’measures of evidence’. Human Heredity 61, 166–188.

Walley, P., Moral, S., 1999. Upper probabilities based only on the likelihood function. Journal of the Royal Statistical Society.Series B: Statistical Methodology 61, 831–847.

Westfall, P. H., 2010. Comment on B. Efron, "Correlated z-values and the accuracy of large-scale statistical estimates". Journal of the American Statistical Association 105, 1063–1066.

Yang, Y., Aghababazadeh, F. A., Bickel, D. R., 2013a. Parametric estimation of the local false discovery rate for identifying genetic associations. IEEE/ACM Transactions on Computational Biology and Bioinformatics 10, 98–108.

Yang, Z., Li, Z., Bickel, D. R., 2013b. Empirical Bayes estimation of posterior probabilities of enrichment: A comparative study of five estimators of the local false discovery rate. BMC Bioinformatics 14, art. 87.

Zhang, Z., Zhang, B., 2013a. A likelihood paradigm for clinical trials. Journal of Statistical Theory and Practice 7, 157–177.

Zhang, Z., Zhang, B., 2013b. Rejoinder [on "A likelihood paradigm for clinical trials"]. Journal of Statistical Theory and Practice 7, 196–203.