A GAME-THEORETIC FRAMEWORK FOR BLENDING BAYESIAN AND FREQUENTIST METHODS OF STATISTICAL INFERENCE

David R. Bickel

Ottawa Institute of Systems Biology
Department of Biochemistry, Microbiology, and Immunology
University of Ottawa; dbickel@uottawa.ca

1. MOTIVATION

Papers compiled in Good (1983) made first attempts at combining attractive aspects of Bayesian and frequentist approaches to statistical inference. While the hybrid inference approach of Yuan (2009) succeeded in leveraging Bayesian point estimators with maximum likelihood estimates, hybrid inference does not yet cover the case of a parameter of interest that has a partially known prior. Since such partial knowledge of a prior occurs in many scientific inference situations, it calls for a theoretical framework for method development that appropriately blends Bayesian and frequentist methods by meeting these criteria:

1. **Complete knowledge of the prior.** If the prior is known, the corresponding posterior is used for inference. Among statisticians, this principle is almost universally acknowledged. However, it is rarely the case of the prior is essentially known.

2. **Negligible knowledge of the prior.** If there is no reliable knowledge of a prior, inference is based on methods that do not require such knowledge. This principle motivates not only the development of confidence intervals and p-values but also Bayesian posteriors derived from improper and data-dependent priors. Accordingly, blended inference must allow the use of such methods when applicable.

3. **Continuum between extremes.** Inference relies on the prior to the extent that it is known while relying on the other methods to the extent that it is not known. Thus, there is a gradation of methodology between the above two extremes. This intermediate scenario calls for a careful balance between pure Bayesian methods on one hand and impure Bayesian or non-Bayesian methods on the other hand.

Instead of framing the knowledge of a prior in terms of confidence intervals, as in pure empirical Bayes approaches, the full version of this extended abstract (Bickel, 2011a) frames it more generally in terms of a set of plausible priors, as in interval probability (Weichselberger, 2000; Augustin, 2002, 2004) and robust Bayesian (Berger, 1984) approaches. Whereas the concept of an unknown prior cannot arise in strict Bayesian statistics, it does arise in robust Bayesian statistics when the levels of belief of an intelligent agent have not been fully assessed (Berger, 1984). Unknown priors also occur in many more objective contexts involving purely frequentist interpretations of probability in terms of variability in the observable world rather than the uncertainty

This research was partially supported by the Canada Foundation for Innovation, by the Ministry of Research and Innovation of Ontario, and by the Faculty of Medicine of the University of Ottawa.

in the mind of an agent. For example, frequency-based priors are routinely estimated under random effects and empirical Bayes models; see, e.g., Efron (2010). (Bickel (2011a) comments further on interpretations of probability and relaxes the assumption of a true prior.)

With respect to the problem at hand, the most relevant robust Bayesian approaches are the *minimax Bayes risk* ("Γ-minimax") practice of minimizing the maximum Bayes risk (Robbins, 1951; Berger, 1985; Vidakovic, 2000) and the *maxmin expected utility* ("conditional Γ-minimax") practice of maximizing the minimum posterior expected payoff or, equivalently, minimizing the maximum posterior expected loss (Gilboa and Schmeidler, 1989; DasGupta and Studden, 1989; Vidakovic, 2000; Augustin, 2002, 2004). Augustin (2004) reviews both methods in terms of interval probabilities that need not be subjective. With typical loss functions, the former method meets the above criteria for classical minimax alternatives to Bayesian methods but does not apply to other attractive alternatives. For example, several confidence intervals, p-values, and objective-Bayes posteriors routinely used in biostatistics are not minimax optimal. (Fraser and Reid (1990) and Fraser (2004) argued that requiring the optimality of frequentist procedures can lead to trade-offs between hypothetical samples that potentially mislead scientists or yield pathological procedures.) Optimality in the classical sense is not required of the alternative procedures under the framework outlined below, which can be understood in terms of maxmin expected utility with a payoff function that incorporates the alternative procedures to be used as a benchmark for the Bayesian posteriors.

2. HEURISTIC OVERVIEW

To define a general theory of blended inference that meets a formal statement of the three criteria, Bickel (2011a) introduced a variation of a zero-sum game of Topsøe (1979), Harremoës and Topsøe (2001), and Topsøe (2007). (The discrete version of the game also appeared in Pfaffelhuber (1977); Grünwald and Philip Dawid (2004) interpreted it as the codelength special case of the maxmin expected utility problem. Alternative versions of minimax optimal codelength, including recent generalizations of normalized maximum likelihood (NML) (e.g., Rissanen and Roos, 2007; Bickel, 2011b), have also been applied to statistical inference.) The "nature" opponent selects a prior consistent with the available knowledge as the "statistician" player selects a posterior distribution with the aim of maximizing the minimum information gained relative to one or more alternative methods. Such benchmark methods may be confidence interval procedures, frequentist hypothesis tests, or other techniques that are not necessarily Bayesian.

From that game, Bickel (2011a) derived a widely applicable framework for testing hypotheses. For concreteness, the motivating results are heuristically summarized here. Consider the problem of testing $H_0 : \theta_* = 0$, the hypothesis that a real-valued parameter θ_* of interest is equal to the point 0 on the real line $\mathbb{R}$. The observed data vector x is modeled as a realization of a random variable denoted by X. Let $p(x)$ denote the p-value resulting from a statistical test.

The p-value for a simple (point) null hypothesis is often smaller than Bayesian posterior probabilities of the hypothesis (Lindley, 1957; Berger and Sellke, 1987). Suppose θ_* has an unknown prior distribution according to which the prior probability of H_0 is π_0. While π_0 is unknown, it is assumed to be no less than some known lower bound denoted by $\underline{\pi}_0$.

Following the methodology of Berger et al. (1994), Sellke et al. (2001) found a generally applicable lower bound on the Bayes factor. As Bickel (2011a) explains, that bound immediately leads to

$$\underline{\Pr}\left(H_0 | p(X) = p(x)\right) = \left(1 - \left(\frac{1 - \underline{\pi}_0}{\underline{\pi}_0 e p(x) \log p(x)}\right)\right)^{-1}$$

as a lower bound on the posterior probability of the null hypothesis for $p(x) < 1/e$ and to $\underline{\pi}_0$ as a lower bound on the probability if $p(x) \geq 1/e$.

In addition to $\Pr\left(H_0 | p(X) = p(x)\right)$, the unknown Bayesian posterior probability of H_0, there is a frequentist posterior probability of H_0 that will guide selection of a posterior probability for inference based on $\pi_0 \geq \underline{\pi}_0$ and other constraints summarized by

$$\Pr(H_0|p(X) = p(x)) \geq \underline{\Pr}(H_0|p(X) = p(x)).$$

While it is incorrect to interpret the p-value $p(x)$ as a *Bayesian* posterior probability, it is seen in Bickel (2011a) that $p(x)$ is a *confidence* posterior probability that H_0 is true.

With the confidence posterior as the benchmark, the solution to the optimization problem described above gives the blended posterior probability that the null hypothesis is true. It is simply the maximum of the p-value and the lower bound on the Bayesian posterior probability:

$$\Pr(H_0; p(x)) = p(x) \vee \underline{\Pr}(H_0|p(X) = p(x)).$$

By plotting $\Pr(H_0; p(x))$ as a function of $p(x)$ and $\underline{\pi}_0$, figures in Bickel (2011a) illustrate each of the above criteria for blended inference:

1. **Complete knowledge of the prior.** In this example, the prior is only known when $\underline{\pi}_0 = 1$, in which case

   $$\Pr(H_0; p(x)) = \underline{\Pr}(H_0|p(X) = p(x)) = 1$$

   for all $p(x)$. Thus, the p-value is ignored in the presence of a known prior.

2. **Negligible knowledge of the prior.** There is no knowledge of the prior when $\underline{\pi}_0 = 0$ and negligible knowledge when $\underline{\pi}_0$ is so low that $\underline{\Pr}(H_0|p(X) = p(x)) \leq p(x)$. In such cases, $\Pr(H_0; p(x)) = p(x)$, and the Bayesian posteriors are ignored.

3. **Continuum between extremes.** When $\underline{\pi}_0$ is of intermediate value in the sense that $\underline{\Pr}(H_0|p(X) = p(x))$ is exclusively between $p(x)$ and 1,

   $$\Pr(H_0; p(x)) = \underline{\Pr}(H_0|p(X) = p(x)) < 1.$$

   Consequently, $\Pr(H_0; p(x))$ increases gradually from $p(x)$ to 1 as $\underline{\pi}_0$ increases (Bickel, 2011a). In this case, the blended posterior lies in the set of allowed Bayesian posteriors but is on the boundary of that set that is the closest to the p-value. Thus, both the p-value and the Bayesian posteriors influence the blended posterior and thus the inferences made on its basis.

The plotted parameter distribution is presented in Bickel (2011a) as a widely applicable blended posterior.

Bickel (2011a) offers additional details and generalizations.

3. REFERENCES

Augustin, T., 2002. Expected utility within a generalized concept of probability - a comprehensive framework for decision making under ambiguity. Statistical Papers 43 (1), 5–22.

Augustin, T., 2004. Optimal decisions under complex uncertainty - basic notions and a general algorithm for data-based decision making with partial prior knowledge described by interval probability. Zeitschrift fur Angewandte Mathematik und Mechanik 84 (10-11), 678–687.

Berger, J., Brown, L., Wolpert, R., 1994. A unified conditional frequentist and Bayesian test for fixed and sequential simple hypothesis-testing. Annals of Statistics 22 (4), 1787–1807.

Berger, J. O., 1984. Robustness of Bayesian analyses. Studies in Bayesian econometrics. North-Holland.

Berger, J. O., 1985. Statistical Decision Theory and Bayesian Analysis. Springer, New York.

Berger, J. O., Sellke, T., 1987. Testing a point null hypothesis: The irreconcilability of p values and evidence. Journal of the American Statistical Association 82 (397), 112–122.

Bickel, D. R., 2011a. Blending Bayesian and frequentist methods according to the precision of prior information with an application to hypothesis testing. Technical Report, Ottawa Institute of Systems Biology, arXiv:1107.2353.

Bickel, D. R., 2011b. A predictive approach to measuring the strength of statistical evidence for single and multiple comparisons. Canadian Journal of Statistics, DOI: 10.1002/cjs.10109.

DasGupta, A., Studden, W., 1989. Frequentist behavior of robust Bayes estimates of normal means. Statist. Dec. 7, 333–361.

Efron, B., 2010. Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction. Cambridge University Press.

Fraser, D. A. S., 2004. Ancillaries and conditional inference. Statistical Science 19 (2), 333–351.

Fraser, D. A. S., Reid, N., 1990. Discussion: An ancillarity paradox which appears in multiple linear regression. The Annals of Statistics 18 (2), 503–507.

Gilboa, I., Schmeidler, D., 1989. Maxmin expected utility with non-unique prior. Journal of Mathematical Economics 18 (2), 141–153.

Good, I., 1983. Good Thinking: the Foundations of Probability and Its Applications. G - Reference, Information and Interdisciplinary Subjects Series. University of Minnesota Press.

Grünwald, P., Philip Dawid, A., 2004. Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. Annals of Statistics 32 (4), 1367–1433.

Harremoës, P., Topsøe, F., 2001. Maximum entropy fundamentals. Entropy 3 (3), 191–226.

Lindley, D. V., 1957. A statistical paradox. Biometrika 44 (1/2), pp. 187–192.

Pfaffelhuber, E., 1977. Minimax Information Gain and Minimum Discrimination Principle. In: Csiszár, I., Elias, P. (Eds.), Topics in Information Theory. Vol. 16 of Colloquia Mathematica Societatis János Bolyai. János Bolyai Mathematical Society and North-Holland, pp. 493–519.

Rissanen, J., Roos, T., 2007. Conditional NML universal models. Information Theory and Applications Workshop, 2007 (ITA-07), 3337–3341.

Robbins, H., 1951. Asymptotically subminimax solutions of compound statistical decision problems. Proc. Second Berkeley Symp. Math. Statist. Probab. 1, 131–148.

Sellke, T., Bayarri, M. J., Berger, J. O., 2001. Calibration of p values for testing precise null hypotheses. American Statistician 55 (1), 62–71.

Topsøe, F., 1979. Information theoretical optimization techniques. Kybernetika 15 (1), 8–27.

Topsøe, F., 2007. Information theory at the service of science. In: Tóth, G. F., Katona, G. O. H., Lovász, L., Pálfy, P. P., Recski, A., Stipsicz, A., Szász, D., Miklós, D., Csiszár, I., Katona, G. O. H., Tardos, G., Wiener, G. (Eds.), Entropy, Search, Complexity. Vol. 16 of Bolyai Society Mathematical Studies. Springer Berlin Heidelberg, pp. 179–207.

Vidakovic, B., 2000. Gamma-minimax: A paradigm for conservative robust Bayesians. Robust Bayesian Analysis. Springer, New York, pp. 241–260.

Weichselberger, K., 2000. The theory of interval-probability as a unifying concept for uncertainty. International Journal of Approximate Reasoning 24 (2-3), 149–170.

Yuan, B., 2009. Bayesian frequentist hybrid inference. Annals of Statistics 37, 2458–2501.