Departing from Bayesian inference toward minimaxity to the extent that the posterior distribution is unreliable

David R. Bickel

December 31, 2017

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

+01 (613) 562-5800, ext. 8670

dbickel@uottawa.ca

Abstract

A Bayesian model may be relied on to the extent of its adequacy by minimizing the posterior expected loss raised to the power of a discounting exponent. The resulting action is minimax under broad conditions when the sample size is held fixed and the discounting exponent is infinite. On the other hand, for any finite discounting exponent, the action is Bayes when the sample size is sufficiently large. Thus, the action is Bayes when there is enough reliable information in the posterior distribution, is minimax when the posterior distribution is completely unreliable, and is a continuous blend of the two extremes otherwise.

Keywords: Bayes action; blended inference; discounting coefficient; L_p norm; minimax action; relative risk aversion

1 Introduction

According to the foundations of Bayesian decision making, an agent should choose the action that minimizes posterior expected loss. That prescription has the advantage of operationally defining probability but often cannot be freely carried out in applications of Bayesian statistics. That is because the prior distribution, if not other aspects of the statistical model, is to some extent unreliable, for a number of reasons. For example, the prior may be improper, at best approximating a probability measure and subject to the marginalization paradox (Dawid et al., 1973). Or the prior, chosen for mathematical or computational tractability, may have highly arbitrary elements. It is rarely feasible to explicitly account for all relevant sources of uncertainty. As a result, prior distributions are considered tentative, subject to revision in light of model checking on the basis of observed data (Gelman and Shalizi, 2013). What do the resulting posterior probabilities mean, seeing that they are not necessarily reliable enough to use for taking the actions that minimize posterior loss?

To address the problem of the unreliable prior distribution, various robust Bayesian solutions weaken or discount the prior distribution to an extent determined by a discounting coefficient by replacing the prior with a set of prior distributions in some neighborhood around it, with larger neighborhoods for larger values of the coefficient (Augustin et al., 2014, §4.7). Since such solutions are complex and leave open the question of how to report an estimate or prediction or to make another decision on the basis of the resulting set of posterior distributions, Section 2 introduces a simple approach to decision making on the basis of an unreliable prior. It prescribes the Bayes action when the prior is completely reliable, the minimax action when the prior is completely unreliable, and otherwise an intermediate action that is discounted to the extent that the prior is unreliable, with the Bayes and minimax actions at the ends of a continuum (cf. Bickel, 2015). Some hypothesis-based examples illustrate the ideas and their applicability to set estimation (Section 3).

The last part of the paper focuses on asymptotics with illustrative examples. The minimin act as well as the minimax act is a limiting special case as the degree of discounting becomes infinite (Section 4). Section 5 proves that the discounted action eventually converges to the Bayes action as the sample size increases.

2 Discounted expected loss and discounted action

Consider a family $\mathcal{F} = \{f(\bullet|\theta) : \theta \in \Theta\}$ of sampling distributions such that $f(\bullet|\theta)$ is a probability density function on a sample space $\mathcal{X}(n)$ for a sample size of n observations and conditional on a parameter value θ in some set Θ. The observed sample $x \in \mathcal{X}(n)$ is modeled as a realization of the random element $X \sim f(\bullet|\theta)$. Let $L : \Theta \times \mathcal{A} \to [0, \infty[$ be a loss function on Θ and a set $\mathcal{A}$ of possible actions.

A Bayesian model $(\mathcal{F}, P)$, consisting of $\mathcal{F}$ and a prior distribution P of θ, is discounted to degree $\Delta \in \mathbb{R}$, the *discounting exponent*, as follows. For a measure ν that dominates P, the Radon-Nikodym derivative $p = dP/d\nu$ is the *prior density*. The corresponding *posterior density* is

$$p(\theta|x) := \frac{f(x|\theta)\, p(\theta)}{\int f(x|\theta)\, p(\theta)\, d\nu(\theta)}$$

where $d\nu(\theta) = d\theta$ if ν is the Lebesgue measure, and the posterior distribution $P(\bullet|x)$ satisfies $p(\bullet|x) = dP(\bullet|x)/d\nu$. For any $\Delta \in \{-\infty, \infty\} \cup \mathbb{R} \setminus \{0\}$, the Δ-*discounted posterior expected loss* and Δ-*discounted posterior action* are

$$E^p_{\Delta,a}(L|x) = \begin{cases} \left(\int L^\Delta(\theta, a)\, p(\theta|x)\, d\nu(\theta)\right)^{1/\Delta} & \text{if } -\infty < \Delta < \infty, \Delta \neq 0 \\ \operatorname{ess\,sup}^{P(\bullet|x)}_{\theta\in\Theta} L(\theta, a) & \text{if } \Delta = \infty \\ \operatorname{ess\,inf}^{P(\bullet|x)}_{\theta\in\Theta} L(\theta, a) & \text{if } \Delta = -\infty \end{cases}$$

$$\widehat{a}^p_\Delta = \arg\inf_{a\in\mathcal{A}} E^p_{\Delta,a}(L|x),$$

where $L^\Delta(\theta, a) = (L(\theta, a))^\Delta$, and $\operatorname{ess\,sup}^{P(\bullet|x)}_{\theta\in\Theta} h(\theta)$ is the essential supremum of a nonnegative-valued, $P(\bullet|x)$-measurable function h over Θ with respect to $P(\bullet|x)$, that is,

$$\operatorname{ess\,sup}^{P(\bullet|x)}_{\theta\in\Theta} h(\theta) = \inf\{H > 0 : P(\{\theta \in \Theta : h(\theta) \le H\}|x) = 1\}$$

with $\inf \emptyset = \infty$. Similarly, $\operatorname{ess\,inf}^{P(\bullet|x)}_{\theta\in\Theta} h(\theta)$ is the essential infimum of h.

In the case of no discounting $(\Delta = 1)$, $\widehat{a}^p_1 = \arg\inf_{a\in\mathcal{A}} E^p_{1,a}(L|x)$ is called the *Bayes action* since it minimizes $E^p_{1,a}(L|x)$, the posterior expected loss. Under complete discounting $(\Delta = \infty)$, $\widehat{a}^p_\infty = \arg\inf_{a\in\mathcal{A}} \operatorname{ess\,sup}^{P(\bullet|x)}_{\theta\in\Theta} L(\theta, a$ is the *essential minimax action*. Between them $(1 < \Delta < \infty)$, the effect of the model $(\mathcal{F}, P)$ on the action is reduced by amount Δ. That suggests 2, the harmonic mean of 1 and ∞, as a moderate default value of Δ. At the opposite extreme, that of maximal inverse discounting ($\Delta = -\infty$), $\widehat{a}_{\infty}^{p} = \arg\inf_{a\in\mathcal{A}} \operatorname{ess\,inf}_{\theta\in\Theta}^{P(\bullet|x)} L(\theta, a)$ is the *essential minimin action*, which is of less practical interest.

3 Examples involving hypotheses

Example 1. In hypothesis testing, one infers whether ($a = 1$) or not ($a = 0$) the unknown value of θ is in $\mathcal{H}$, a measurable subset of Θ. A relevant loss function is

$$L(\theta, a) = \begin{cases} 1 & \text{if } \theta \in \mathcal{H}, a = 0 \\ c & \text{if } \theta \notin \mathcal{H}, a = 1 \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

where the *cost ratio* $c > 0$, is the penalty for incorrectly inferring that $\theta \in \mathcal{H}$ relative to that for incorrectly inferring that $\theta \notin \mathcal{H}$. Let $\chi_{\{0\}}$ and $\chi_{\{1\}}$ denote characteristic functions on $\{0, 1\}$. For any finite $\Delta \neq 0$,

$$\begin{aligned} E_{\Delta,a}^{p}(L|x) &= \left(\int_{\mathcal{H}} \chi_{\{0\}}^{\Delta}(a)\, p(\theta|x)\, d\nu(\theta) + \int_{\Theta\setminus\mathcal{H}} c^{\Delta}\chi_{\{1\}}^{\Delta}(a)\, p(\theta|x)\, d\nu(\theta)\right)^{1/\Delta} \\ &= \left(P(\mathcal{H}|x)\,\chi_{\{0\}}(a) + c^{\Delta}(1 - P(\mathcal{H}|x))\,\chi_{\{1\}}(a)\right)^{1/\Delta} \\ &= P^{1/\Delta}(\mathcal{H}|x)\,\chi_{\{0\}}(a) + c\,(1 - P(\mathcal{H}|x))^{1/\Delta}\,\chi_{\{1\}}(a) \end{aligned}$$

$$\widehat{a}_{\Delta}^{p} = \arg\inf_{a\in\{0,1\}} E_{\Delta,a}^{p}(L|x) = \begin{cases} 0 & \text{if } \left(\frac{P(\mathcal{H}|x)}{1-P(\mathcal{H}|x)}\right)^{1/\Delta} < c \\ 1 & \text{if } \left(\frac{P(\mathcal{H}|x)}{1-P(\mathcal{H}|x)}\right)^{1/\Delta} > c \\ \{0,1\} & \text{if } \left(\frac{P(\mathcal{H}|x)}{1-P(\mathcal{H}|x)}\right)^{1/\Delta} = c \end{cases},$$

suggesting that $\omega_{\Delta}(\mathcal{H}|x) \coloneqq (P(\mathcal{H}|x)/(1 - P(\mathcal{H}|x)))^{1/\Delta}$ be reported as the Δ-*discounted posterior odds* of $\theta \in \mathcal{H}$. The corresponding Δ-*discounted posterior probability* that $\theta \in \mathcal{H}$ would then be $P_{\Delta}(\mathcal{H}|x) \coloneqq 1/(1 + 1/\omega_{\Delta}(\mathcal{H}|x))$, resulting in $P_1(\mathcal{H}|x) = P(\mathcal{H}|x)$ and $\lim_{\Delta\to\infty} P_{\Delta}(\mathcal{H}|x) = 1/2$. At $\Delta = 2$, lower posterior probabilities are discounted by more orders of magnitude than higher ones (Figure 1, left plot), as is desirable given the unreliability of extremely small probabilities.

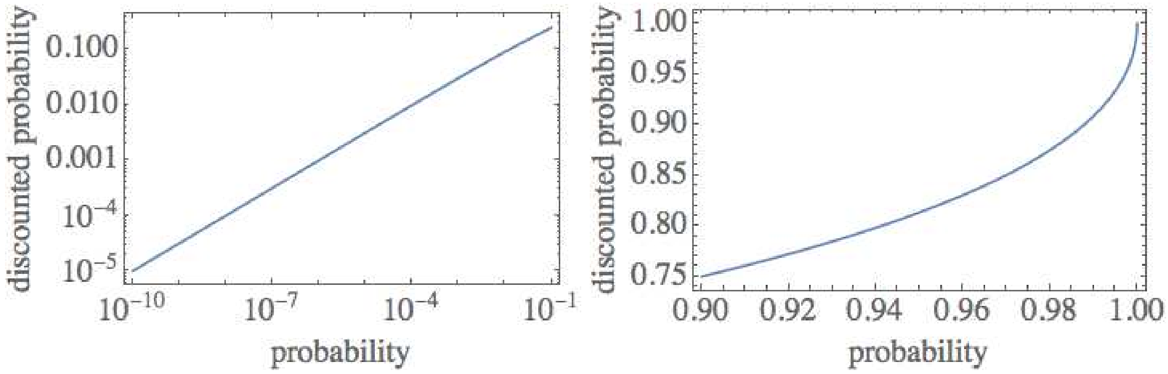

Figure 1: The 2-discounted posterior probability $P_2(\mathcal{H}|x)$ as a function of the posterior probability $P(\mathcal{H}|x)$. Left plot: probabilities small enough to be interesting as posterior probabilities of a null hypothesis. Right plot: probabilities large enough for use as the posterior probabilities of credible sets.

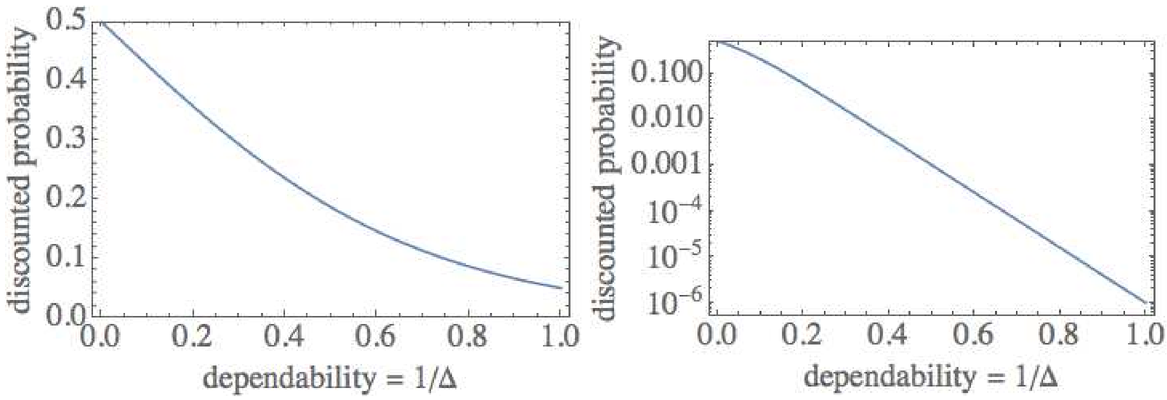

Figure 2: The Δ-discounted posterior probability $P_\Delta(\mathcal{H}|x)$ as a function of the model's dependability, the reciprocal of its discounting exponent Δ. Left plot: $P(\mathcal{H}|x) = 5\%$. Right plot: $P(\mathcal{H}|x) = 10^{-6}$.

That is applicable not only to hypothesis testing but also to set estimation. A set estimate $\mathcal{H}(x)$ is a $(1-\alpha)\,100\%$ Δ-*discounted credible set* if $P_\Delta(\mathcal{H}(x)\,|x) = (1-\alpha)\,100\%$. For example, a 95% 2-discounted credible set would be an approximately 80% 1-discounted credible set, a standard 80% credible set without discounting (Figure 1, right plot), whereas a 95% ∞-discounted credible set would be a 50% credible set.

Figure 2 displays discounted probabilities at other values of Δ. There, the *dependability* $D(\Delta)$ of a model is the reciprocal of the extent that it requires discounting: $D(\Delta) = 1/\Delta$. Thus, $D(1) = 1$, $D(2) = 1/2$, $D(4) = 1/4$, and $D(\infty) = 0$. ▲

Example 2. For $\mathcal{H}$, a measurable subset of Θ, let $\chi_\mathcal{H}$ denote a characteristic function on Θ. When estimating

$\chi_{\mathcal{H}}(\theta)$, the indicator of whether θ is in $\mathcal{H}$, with an action in $\mathcal{A} = [0,1]$ as opposed to the $\mathcal{A} = \{0,1\}$ of Example 1, squared-error loss is appropriate:

$$L(\theta, a) = (a - \chi_{\mathcal{H}}(\theta))^2 .$$

For any finite $\Delta \neq 0$,

$$\begin{aligned} E^p_{\Delta,a}(L|x) &= \left(\int (a - \chi_{\mathcal{H}}(\theta))^{2\Delta} p(\theta|x)\, d\nu(\theta) \right)^{1/\Delta} \\ &= \left(\int_{\mathcal{H}} (1-a)^{2\Delta} p(\theta|x)\, d\nu(\theta) + \int_{\Theta\setminus\mathcal{H}} a^{2\Delta} p(\theta|x)\, d\nu(\theta) \right)^{1/\Delta} \\ &= \left(P(\mathcal{H}|x)(1-a)^{2\Delta} + (1 - P(\mathcal{H}|x))\, a^{2\Delta} \right)^{1/\Delta} \end{aligned}$$

$$\begin{aligned} \widehat{a}^p_\Delta &= \arg \inf_{a\in[0,1]} P(\mathcal{H}|x)(1-a)^{2\Delta} + (1 - P(\mathcal{H}|x))\, a^{2\Delta} \\ &= \left(1 + \left(\frac{P(\mathcal{H}|x)}{1 - P(\mathcal{H}|x)} \right)^{-1/(2\Delta-1)} \right)^{-1} . \end{aligned}$$

Although $\widehat{a}^p_1$ might be considered as an elicited posterior probability that $\theta \in \mathcal{H}$ (cf. Lad, 1996, §6.6.3), the $\widehat{a}^p_\Delta$ of this example differs in general from the $P_\Delta(\mathcal{H}|x)$ of Example 1. That may be illustrated using the default $\Delta = 2$:

$$\frac{P_2(\mathcal{H}|x)}{1 - P_2(\mathcal{H}|x)} = \sqrt{\frac{P(\mathcal{H}|x)}{1 - P(\mathcal{H}|x)}} \neq \left(\frac{P(\mathcal{H}|x)}{1 - P(\mathcal{H}|x)} \right)^{\frac{1}{3}} = \widehat{a}^p_2 .$$

Nonetheless, $\widehat{a}^p_1 = P(\mathcal{H}|x) = P_1(\mathcal{H}|x)$ and $\lim_{\Delta\to\infty} \widehat{a}^p_\Delta = 1/2 = \lim_{\Delta\to\infty} P_\Delta(\mathcal{H}|x)$. ▲

4 Limiting cases of the discounting exponent

With n held fixed, the discounted posterior expected losses of finite Δ approach those of $\Delta = \pm\infty$ under the conditions of the following proposition, which needs Lemma 1 for its proof.

Lemma 1. *If h is a nonnegative-valued, $P(\bullet|x)$-measurable function on Θ such that* $\operatorname{ess\,sup}^{P(\bullet|x)}_{\theta\in\Theta} h(\theta) < \infty$, *then*

$$\lim_{\partial\to\infty}\left(\int h^{\partial}(\theta)\, d\,P(\theta|x)\right)^{1/\partial} = \text{ess sup}_{\theta\in\Theta}^{P(\bullet|x)}\, h(\theta)\,.$$

Proof. That directly follows from Aliprantis and Border (2006, Lemma 13.1). □

Proposition 1. *Suppose* $L(\bullet, a)$ *is nonnegative-valued and* $P(\bullet|x)$*-measurable and that it satisfies*

$$\text{ess sup}_{\theta\in\Theta}^{P(\bullet|x)}\, L(\theta, a) < \infty$$

$$\text{ess sup}_{\theta\in\Theta}^{P(\bullet|x)}\, 1/\,L(\theta, a) < \infty$$

for some $a \in \mathcal{A}$*. Then* $\lim_{\Delta\to\pm\infty} E_{\Delta,a}^{p}(L\,|x) = E_{\pm\infty,a}^{p}(L\,|x)\,.$

Proof. By Lemma 1, $\lim_{\Delta\to\infty} E_{\Delta,a}^{p}(L\,|x) = E_{\infty,a}^{p}(L\,|x)$, and

$$\begin{aligned}
\lim_{\Delta\to-\infty} E_{\Delta,a}^{p}(L\,|x) &= \lim_{\Delta\to-\infty}\left(\int L^{-|\Delta|}(\theta, a)\, p(\theta|x)\, d\nu(\theta)\right)^{-1/|\Delta|} \\
&= \lim_{\partial\to\infty}\left(\int L^{-\partial}(\theta, a)\, p(\theta|x)\, d\nu(\theta)\right)^{-1/\partial} \\
&= \left(\lim_{\partial\to\infty}\left(\int (1/\,L(\theta, a))^{\partial}\, P(\theta|x)\right)^{1/\partial}\right)^{-1} \\
&= \left(\text{ess sup}_{\theta\in\Theta}^{P(\bullet|x)}\, 1/\,L(\theta, a)\right)^{-1} \\
&= \text{ess inf}_{\theta\in\Theta}^{P(\bullet|x)}\, L(\theta, a) = E_{-\infty,a}^{p}(L\,|x)\,.
\end{aligned}$$

□

The next example compares that discounted-model approach to minimaxity with one based on a degree of risk aversion.

Example 3. Suppose $W(\theta, a) \geq 0$ would be the amount of future wealth in a portfolio if $\theta \in \Theta$ is the future state of the world and $a \in \mathcal{A}$ is the investment strategy. The loss corresponding to the power utility function

(Campbell and Viceira, 2002, §2.1.2) is

$$L(\theta, a) = \begin{cases} L_0 + \frac{1}{\rho - 1}\left(\frac{1}{W^{\rho-1}(\theta, a)} - 1\right) & \text{if } \rho \neq 1 \\ L_0 - \ln W(\theta, a) & \text{if } \rho = 1 \end{cases},$$

where $\rho \in \mathbb{R}$ is the *relative risk aversion*, and $L_0 \in \mathbb{R}$ such that $L(\theta, a) \geq 0$ for all $\theta \in \Theta$ and $a \in \mathcal{A}$. If $W(\theta, a) < 1$ and if $W(\bullet, a)$ satisfies the conditions of Lemma 1, then the Bayes action in the limit of increasing relative risk aversion is

$$\begin{aligned} \lim_{\rho\to\infty} \widehat{a}_1^p &= \arg \inf_{a\in\mathcal{A}} \lim_{\rho\to\infty} E_{1,a}^p (L|x) \\ &= \arg \inf_{a\in\mathcal{A}} \lim_{\rho\to\infty} \left(\int \left(\frac{1}{W^{\rho-1}(\theta, a)} - 1\right) p(\theta|x)\, d\nu(\theta)\right)^{1/\rho} \\ &= \arg \inf_{a\in\mathcal{A}} \text{ess sup}_{\theta\in\Theta}^{P(\bullet|x)} \frac{1}{W(\theta, a)} = \arg \sup_{a\in\mathcal{A}} \text{ess inf}_{\theta\in\Theta}^{P(\bullet|x)} W(\theta, a), \end{aligned}$$

which reduces to the special case of minimax reciprocal wealth or maximin wealth (cf. Arrow, 1973, pp. 256-257). On the other hand, Proposition 1's limit of maximal discounting ($\Delta \to \infty$) gives, with a fixed $\rho \neq 1$,

$$\begin{aligned} \widehat{a}_\infty^p &= \arg \inf_{a\in\mathcal{A}} \text{ess sup}_{\theta\in\Theta}^{P(\bullet|x)} \frac{1}{\rho - 1}\left(\frac{1}{W^{\rho-1}(\theta, a)} - 1\right) \\ &= \arg \sup_{a\in\mathcal{A}} \text{ess inf}_{\theta\in\Theta}^{P(\bullet|x)} W(\theta, a), \end{aligned}$$

implying that $\lim_{\rho\to\infty} \widehat{a}_1^p = \widehat{a}_\infty^p$. The same holds for $\rho = 1$. ▲

5 Convergence to the Bayes action as data arrive

Discounting the model leads to estimates and other actions that tend to be closer to the minimax action for smaller samples but closer to the Bayes action for larger samples, once there are enough data to overwhelm questionable aspects of the model.

Proposition 2. *Let δ_{θ_0} denote the Dirac measure with its mass at a $\theta_0 \in \Theta$, and let $\xrightarrow{weak}$ denote weak convergence as $n \to \infty$. If $X \sim f(\bullet|\theta_0)$, if $P(\bullet|X) \xrightarrow{weak} \delta_{\theta_0}$, and if $L(\bullet, a)$ is continuous on Θ for some*

$a \in \mathcal{A}$, then, for any finite $\Delta \neq 0$,

$$E^p_{\Delta,a}(L|X) - E^p_{1,a}(L|X) \xrightarrow{weak} 0.$$

Proof. For any finite $\Delta \neq 0$, the mapping theorem yields

$$\begin{aligned} E^p_{\Delta,a}(L|X) &= \left(\int L^\Delta(\theta, a)\, d\,P(\theta|X)\right)^{1/\Delta} \\ &\xrightarrow{\text{weak}} \left(\int L^\Delta(\theta, a)\, d\,\delta_{\theta_0}(\theta)\right)^{1/\Delta} \\ &= \left(L^\Delta(\theta_0, a)\right)^{1/\Delta} = L(\theta_0, a)\,. \end{aligned}$$

Thus, $E^p_{\Delta,a}(L|X) - E^p_{1,a}(L|X) \xrightarrow{\text{weak}} L(\theta_0, a) - L(\theta_0, a)$. □

The next example illustrates the result.

Example 4. Example 2, continued. Let $p(\theta|\mathcal{H})$ and $p(\theta|\Theta\backslash\mathcal{H})$ denote the prior probability densities conditional on a null hypothesis that $\theta \in \mathcal{H}$ and the alternative hypothesis that $\theta \notin \mathcal{H}$, respectively. The posterior odds of the null hypothesis is

$$\frac{P(\mathcal{H}|x)}{1 - P(\mathcal{H}|x)} = \frac{P(\mathcal{H})}{1 - P(\mathcal{H})} \frac{\int f(x|\theta)\, p(\theta|\mathcal{H})\, d\nu(\theta)}{\int f(x|\theta)\, p(\theta|\Theta\backslash\mathcal{H})\, d\nu(\theta)},$$

the right-hand side of which is the product of the prior odds and the Bayes factor. Both quantities typically involve unverified assumptions that require some kind of discounting before decision making on the basis of the resulting $P(\mathcal{H}|x)$. As a concrete form of Proposition 2, Figure 3 displays the convergence of the discounted and Bayes estimates to the same value, 0 if the null hypothesis is false or 1 if it is true, for, as $n \to \infty$, the Bayes factor tends to decrease if the null hypothesis is false or increase if it is true. The curves for $\widehat{a}^p_2$ and $\widehat{a}^p_4$ reflect the gradation of estimates between the extremes of the Bayes action $\widehat{a}^p_1$ and the minimax action $\widehat{a}^p_\infty = 1/2$. ▲

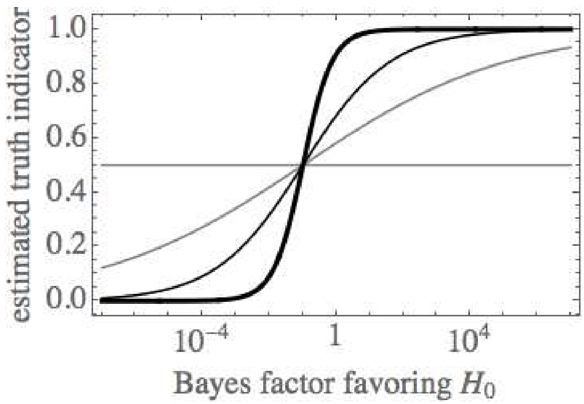

Figure 3: Each $\widehat{a}^{p}_{\Delta}$ as an estimate of $\chi_{\mathcal{H}}(\theta)$ for $\Delta = \infty$ (the horizontal line at 1/2), $\Delta = 4$ (the lighter thin curve), $\Delta = 2$ (the darker thin curve), and $\Delta = 1$ (the thick curve). Following Benjamin et al. (2017), the prior odds is 10:1.

Acknowledgments

This research was partially supported by the Natural Sciences and Engineering Research Council of Canada and by the Faculty of Medicine of the University of Ottawa.

References

Aliprantis, C., Border, K., 2006. Infinite Dimensional Analysis: A Hitchhiker's Guide. Springer-Verlag, Berlin.

Arrow, K. J., 1973. Some ordinalist-utilitarian notes on rawls's theory of justice. The Journal of Philosophy 70 (9), 245–263.

Augustin, T., Coolen, F., de Cooman, G., Troffaes, M. (Eds.), 2014. Introduction to Imprecise Probabilities. Wiley Series in Probability and Statistics. Wiley.

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C. D., Clyde, M., Cook, T. D., De Boeck, P., Dienes, Z., Dreber, A., Easwaran, K., Efferson, C., Fehr, E., Fidler, F., Field, A. P., Forster, M., George, E. I., Gonzalez, R., Goodman, S., Green, E., Green, D. P., Greenwald, A. G., Hadfield, J. D., Hedges, L. V., Held, L., Hua Ho, T., Hoijtink, H., Hruschka, D. J., Imai, K., Imbens, G., Ioannidis, J. P. A., Jeon, M., Jones, J. H., Kirchler, M., Laibson, D., List, J., Little, R., Lupia, A., Machery, E., Maxwell, S. E., McCarthy, M., Moore, D. A., Morgan, S. L., Munafó, M., Nakagawa, S., Nyhan, B., Parker, T. H., Pericchi, L., Perugini, M., Rouder, J., Rousseau, J., Savalei, V., Schönbrodt, F. D., Sellke, T., Sinclair, B., Tingley, D., Van Zandt, T., Vazire, S., Watts, D. J., Winship, C., Wolpert, R. L., Xie, Y., Young, C., Zinman, J., Johnson, V. E., 9 2017. Redefine statistical significance. Nature Human Behaviour, 1.

Bickel, D. R., 2015. Blending Bayesian and frequentist methods according to the precision of prior information with applications to hypothesis testing. Statistical Methods & Applications 24, 523–546.

Campbell, J., Viceira, L., 2002. Strategic Asset Allocation: Portfolio Choice for Long-term Investors. Oxford University Press, New York.

Dawid, A. P., Stone, M., Zidek, J. V., 1973. Marginalization paradoxes in Bayesian and structural inference. Journal of the Royal Statistical Society B 35, 189–213.

Gelman, A., Shalizi, C., 2013. Philosophy and the practice of bayesian statistics. British Journal of Mathematical and Statistical Psychology 66 (1), 8–38.

Lad, F., 1996. Operational Subjective Statistical Methods: A Mathematical, Philosophical, and Historical Introduction. Wiley-Interscience.