Bayesian Analysis (2024) **TBA**, Number TBA, pp. 1

Discussion of Campbell & Gustafson

Kenneth Rice*

As noted by the authors ("C&G" hereafter), putting 'spikes' in priors is contentious: the heart of the Jeffreys-Lindley paradox is arguably [2] that tests focusing on whether a point mass is 'true' can disagree totally with standard non-Bayesian two-sided tests that use p-values. But of course p-values focus not on the truth, but whether the data (or more extreme) data would be surprising enough under the null to merit saying something. For a resolution without any spikes, see Rice and Krakauer [6], reviewing work dating to at least Lehmann [3].

That said, with a spike in the prior the posterior's discontinuity may indeed prevent us defining credible intervals with some exact level of support. If 80% of the posterior lies exactly at θ_0, we can only have intervals with up to 20% support (excluding the spike) or those over 80%, which include it. C&G say this is "puzzling", but it is textbook stuff: the same behaviour crops up with frequentist tests in discrete sample spaces, with only a finite set of outcomes. I do agree that "bizarre" is the right term for C&G's artifice of randomizing the intervals – it is essentially the same workaround described by e.g. Bickel and Doksum [1, pg 224], who are clear it is only a device for proving optimality. Young and Smith [7, pg 67] add pragmatic advice on how, in practice, to avoid the embarrassment of having the analysis' conclusions rest so heavily on the toss of a coin.

Defining credible intervals when spikes are present is a detail addressed in Rice and Ye [5]'s general *regret* formulation of credible regions, where one specifies an estimation loss and forms the credible region as the least-bad 95% of the posterior with regard to the expected loss. Following longstanding non-Bayesian practice – e.g. Neyman [4, pg 112] – it defines a $1-\alpha$ credible region as being just big enough to have support *at least* $1-\alpha$, so not "undefined" as C&G would have it.

Exploring a little of how these regions behave in posteriors with spikes, we consider C&G's Normal location problem, with $\bar{Y} = 1.645/\sqrt{n}$; Figure 1 shows the posterior CDF for $n = 10$, for a 50:50 mixture of a $N(0, 1)$ prior and a spike at $\theta = 0$, and also for the $N(0, 1/2)$ continuous prior that has the same variance. To distinguish between priors with spikes at exactly $\theta = 0$ and decisions that return exactly $\theta = 0$, we consider two losses. First is familiar squared error loss, $L(d) = (\theta - d)^2$, for which we show the regret. Second, as considered in Rice and Ye [5, §4.2], the more general 'shrinkage' loss

$$L(d, h) = \gamma^{1/2} h(d - \theta)^2 + \gamma^{-1/2}(1 - h)\left((d - \theta_0)^2 + (\theta - \theta_0)^2\right), \quad (1)$$

that combines testing decision $h \in \{0, 1\}$ with a real-valued estimate d. (Following Rice and Ye [5] we use we use $\gamma = 0.207$.) The corresponding Bayes rule for d depends on the ratio of posterior mean to posterior standard deviation; based on which it is shrunk to exactly θ_0, or remains the unshrunk posterior mean. Figure 1 shows the *profile regret*, i.e. the excess expected loss for estimate d' if one optimizes out auxiliary decision h.

arXiv: 2010.00000
*Department of Biostatistics, University of Washington, kenrice@uw.edu

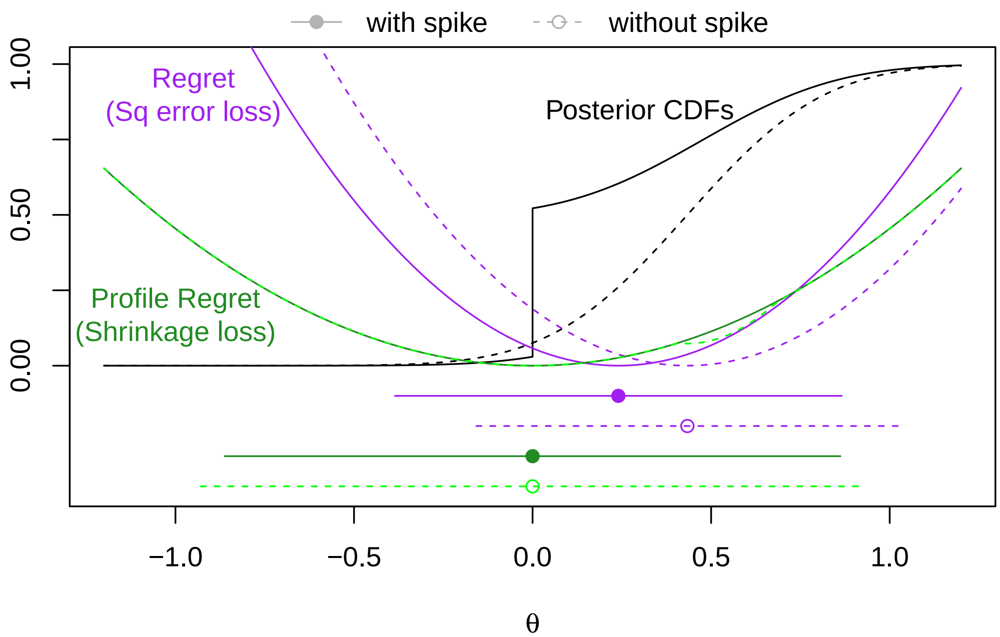

Figure 1: Posterior CDFs, regret under squared error loss and profile regret under the shrinkage loss for $z = 1.645, n = 10$. Bayes rule point estimates and 95% regret/profile regret intervals are shown below, for both priors and losses

The lower part of Figure 1 shows how the 50:50 mixture prior does not always lead to smaller posterior variance and hence narrower regret intervals; with $n = 10$, 95% intervals that are symmetric around the posterior mean have to be 5.8% wider under the spiked prior. (For larger values of n this pattern is reversed, and the ratio of widths decreases roughly linearly until near $n = 5400$, when the posterior spike reaches support 95%.) For the shrinkage loss function, the spiked posterior provides slightly (i.e. 8%) narrower intervals than the continuous one – and this pattern increases monotonically until $n = 5400$, again, when both intervals shrink to single points. The plot also shows that there is a cost (in terms of credible interval width) for using the shrinkage loss versus squared error, regardless of prior. This too persists at larger n up to $n = 5400$, where under the spiked prior using shrinkage results in an ≈15% wider interval, and for the continuous prior ≈70% wider.

We see that the impact (on intervals) of using priors with/without spikes at exactly θ_0 can differ from that of asking questions that yield estimates of exactly θ_0. I welcome C&G's insights on how analysts can be helped to describe what is known about their θ, what they want to know about θ, and how their ways of addressing those two questions may interact.

References

[1] Bickel, P. J. and Doksum, K. A. (2007). *Mathematical statistics: basic ideas and selected topics, vol I, 2nd edition*. Pearson Prentice Hall. 1

[2] Gray, J., Cherry, J. L., Wagenmakers, E.-J., and Ly, A. (2023). "The Jeffreys–Lindley paradox: an exchange." *Archive for History of Exact Sciences*, 77(4): 443–449. 1

[3] Lehmann, E. (1950). "Some Principles of the Theory of Testing Hypotheses." *The Annals of Mathematical Statistics*, 21(1): 1–26. 1

[4] Neyman, J. (1935). "On the Problem of Confidence Intervals." *The Annals of Mathematical Statistics*, 6(3): 111–116.
URL http://www.jstor.org/stable/2957537 1

[5] Rice, K. and Ye, L. (2022). "Expressing regret: a unified view of credible intervals." *The American Statistician*, 76(3): 248–256. 1

[6] Rice, K. M. and Krakauer, C. A. (2023). "Three-decision methods: A sensible formulation of significance tests—and much else." *Annual Review of Statistics and Its Application*, 10: 525–546. 1

[7] Young, G. A. and Smith, R. L. (2005). *Essentials of Statistical Inference*, volume 16. Cambridge University Press. 1