DOI: 10.1017/psa.2026.10266
This is a manuscript accepted for publication in *Philosophy of Science*.
This version may be subject to change during the production process.

Exploring the Maximally Sensitive Priors

Nikodem Lewandowski[1*] and Rafał Urbaniak[2]

[1]Institute of Philosophy, University of Gdańsk, Gdańsk, Poland
[2]Basis Research Institute, New York, USA
*Corresponding author: Nikodem Lewandowski; nikodem.lewandowski@phdstud.ug.edu.pl

Abstract
We explore the Maximum Sensitivity method (MaxSen) of selecting priors for Bayesian inference in the face of severe lack of evidence (Konek, 2013). MaxSen minimizes the need to rely on the epistemic luck of true parameter values falling closely to their prior estimate. We explore the impact of the choice of the scoring method on the recommendations made by MaxSen. It turns out that if we use Kullback-Leibler Divergence (KLD) as a distance measure, the recommendations are the same as that of MaxEnt. We estimate how quickly the recommended prior distribution concentration increases with the planned sample size.

1. Introduction

We investigate the Maximum Sensitivity method (MaxSen) of selecting priors for Bayesian inference in the face of severe lack of evidence, proposed by (Konek, 2013, 2016). Intuitively, MaxSen minimizes the need to rely on the epistemic luck of true parameter values (such as chances) falling close to their prior estimate. Somewhat more formally, imagine a scientist intends to adjudicate between pairwise incompatible and jointly exhaustive hypotheses $H_1, \ldots, H_n$, each associated with a chance function ch_i, based on observations $\mathbf{D} = \{D_1, \ldots, D_m\}$. In this context, $ch_i(\mathbf{D})$ is the chance of $\mathbf{D}$ if H_i is true. The method recommends the prior that minimizes the difference between the range of the objective expected posterior inaccuracy, i.e., the quantity $max_i \mathbb{E}_{ch_i}[I(c')] - min_i \mathbb{E}_{ch_i}[I(c')]$, where I maps credence functions c' and worlds w ("omniscient probability functions") to nonnegative real numbers (parametrization suppressed in the original notation), measuring how inaccurate c' is if w is actual, and $\mathbb{E}_{ch_i}[I(c')]$ takes expectation w.r.t. to which worlds are factual and which datapoints arise.

In his proposal, Konek makes some design decisions: as an inaccuracy measure, he takes Cramer-von-Mises, and as a measure of variability of accuracy, he focuses on the range of expected inaccuracies instead of some integral that incorporates more information about the curvature. Inspired to investigate to what extent the results depend on those design choices, we somewhat generalize this approach, and while we still look at distances to "omniscient" world distributions and expectations over them, we use two different inaccuracy measures, Cramer-von-Mises (CvM) and Kullback-Leibler Divergence (KLD), and aside from looking at the difference between maximum and

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press

minimum as a measure of the expected inaccuracy curvature (CurvK), we use an entropy-based curvature measure, which is a function of the overall shape thereof (CurvE). In this somewhat more general setting, we can, to some extent, isolate the role played by the intuition that we should maximize sensitivity from the particular choices of how the idea is implemented. It turns out that those choices have non-trivial consequences.

In a simulation study, in the context of a symmetric prior Beta distribution selection, we inspect the behavior of four variations of the MaxSen method: CvM CurvK (which is mathematically equivalent to Konek's original method, with some notational changes), CvM CurvE, KLD CurvK, and KLD CurvE. The two key observations result:

(1) With the exception of KLD CurvK, all the methods recommend priors in which the recommended $\alpha = \beta$ grows with the sample size n, rather quickly, leading to recommended priors of surprisingly high levels of concentration.[1]
(2) KLD CurvK, which we think is better in (i) using a more information-theoretically grounded distance measure, and (ii) integrating over the domain instead of comparing the maxima and the minima when measuring variability, consistently recommends uniform priors, suggesting that in a certain rather plausible configuration, MaxSen, a method that considers priors only as instrumental, recommends a prior that is also minimally informative, agreeing with the entropy-recommended priors.

Our analysis reveals that the choice of configuration is far from trivial, as different scoring rules lead to significantly divergent recommendations. Notably, we find that under the KLD CurvK configuration, arguably the most principled approach, MaxSen aligns with the uniform prior. This offers a compelling conceptual bridge, suggesting that the Maximum Entropy Principle might itself be understood as a sensitivity maximizing rule. Furthermore, among the variations we tested, this is the only configuration that avoids the counter-intuitive outcome of prior concentration scaling rapidly with sample size.

We conclude by outlining the necessary steps to generalize MaxSen beyond the binomial constraints used in this study. Specifically, we highlight the need for a formal theoretical proof to validate our simulation based finding that KLD CurvK consistently favors uniformity, a result that opens new avenues for complex modeling choices.

We start with a brief outline of MaxEnt and an explanation of Konek's anti-luck approach to priors. Next, we explore his perspective on the theoretical role of priors and illustrate how this viewpoint informs the formulation of MaxSen. Then, we explore the impact of the choice of the scoring method on the recommendations made by MaxSen.

2. Bayesian inference and maximum entropy

The paper aims to investigate a novel approach to selecting priors in Bayesian inference, particularly in scenarios of severe uncertainty. This approach, introduced by Konek (Konek, 2013, 2016), suggests that rather than solely representing the state of total uncertainty of the rational agent, the focus should be on ensuring a maximally luck-independent posterior.

[1] Prototype simulations suggest that asymmetric betas are not recommended anyway, and as we are dealing with an unbiased prior in a situation of extreme uncertainty, there is conceptually little to recommend investigating asymmetric priors for the task at hand.

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press

The whole discussion is based on the ground of Bayesian inference. It is composed of three essential elements: the prior, the likelihood, and the posterior. The prior represents our belief about the parameters of interest before conditioning on data. The likelihood is the probability of observing the data given particular values of the parameters. The posterior represents the result of updating the prior with the new observations.

In the context of this work, we will focus on the scenario of a scientist and an experiment that she is about to conduct, set to adjudicate between some competing hypotheses $H_1, \ldots, H_n$. The prior and the posterior are probability distributions over those hypotheses. The prior is the scientist's initial belief about the hypotheses, and the posterior is the scientist's updated belief about the hypotheses after she has observed the data.

The relation between these elements is given by Bayes' theorem, below in a simple version in which H stands for hypothesis and E for evidence:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \qquad \text{(Bayes' Theorem)}$$

Our interest lies in how the prior ought to be represented in a scenario of total uncertainty when a scientist has no good evidence that she could use to represent her prior with.[2] There are many approaches to this problem: classical strategies presume that the theoretical role of priors is to represent the epistemic state of an agent. We will explore those propositions to contrast them in the next sections with the approach of Konek, who believes that the theoretical role of priors is securing a maximally luck-independent posterior.

Typically, one distinguishes between subjective and objective Bayesianism. In the state of total uncertainty, the subjective approach advocates suggest that any prior is equally reasonable as the agent has no evidence to support any particular prior (De Finetti, 1979). Therefore, an agent can support her prior with a hunch or intuition. An objectivist, however, would argue that the prior should reflect the state of epistemic uncertainty; therefore, an advocate of that approach will argue that the prior should minimize the amount of information it contains.

In the context of this work, we will stick to the objectivist approach in search of a principled way of selecting rational priors. Advocates of minimizing the amount of information included in the prior represent the non-informative prior approach. This view is contrasted with that of the advocates of weakly informed priors, who argue that minimally informed priors also contain assumptions that can lead to overly informed inferences. Also, in some contexts, the uniform probability distribution is sometimes argued to be sub-optimal, as it seems that naturally, epistemic agents adjudicate extreme values (close to the edges) to be less probable (considerations of extreme values are even more important when we reflect on weakly-informed priors, such as Jeffreys priors, which assign higher density to extreme values) (Gelman, 2009). The proponents of non-informative priors, on the other hand, literally focus on minimizing the amount of information, as measured using an information-theoretic metric, and so they opt for the Maximum Entropy (MaxEnt) method. We will describe this method more thoroughly now, as it is widely used in the literature and stands in contrast to Konek's proposed method.

[2]Eliciting probabilities even from informed experts is still a challenge (O'Hagan et al., 2006), but we do not discuss it in this paper.

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press

If all we know is that the random variable has values from a given interval, the corresponding MaxEnt prior is a uniform distribution.[3] In what follows, however, we will be mostly working with finite sets of hypotheses (grid-approximating a range of parameter values) and distributions over them. In this context, we can state the recommendation given by MaxEnt a bit more clearly: in severe uncertainty about the hypotheses $H_1, \ldots, H_n$ about the population frequency, the prior should be the distribution with maximal entropy, that is, it should be the uniform distribution assigning equal probability to each value of the random variable.[4]

There are two main classes of arguments for adopting MaxEnt: theoretical and pragmatic. The theoretical arguments are based on the premise that the central role of priors is to represent a researcher's epistemic state. If she is completely uncertain about the distribution over the hypotheses, she ought to adopt a minimally informative prior that mirrors this uncertainty. The pragmatic argument is simple and modest in nature: the use of MaxEnt recommended distributions is abundant in many empirical sciences (for instance, it is often assumed that the distribution of residuals is normal), and it works quite well.[5]

As the theoretical argument for MaxEnt is somewhat similar to Laplace's Principle of Insufficient Reason, they face the same challenge. The uniformity of a distribution is not preserved under arbitrary transformations of random variables. For instance, a uniform distribution over the values of $X \in [0, 1]$ will not lead to a uniform distribution over the values of X^2, so already some information has to be incorporated at the point when the researchers decide which variables they want to accept uniform priors about. This seems to disagree with the idea that minimally informative priors somehow represent a pristine, epistemic *tabula rasa*.

Another problem with taking a uniform prior is that a uniform distribution is susceptible to over-fitting. The model tends to be over-fitted when it is too flexible and too sensitive to data, in which case the noise can have too much impact on it. Such a model might adjust itself to irrelevant patterns in data because of how easily it responds to observations. This is closely related to what will be discussed in the paper, and we will get back to the issue later on.

Konek argues that we need a principle that governs our choice of priors, but he also claims that MaxEnt is not the right one. He re-frames the theoretical role of priors and proposes a new principle, which is centered on maximizing the sensitivity of the prior distribution—hence the name MaxSen. Both, the method and Konek's rationale for it will be discussed in the next section.

The goal of our paper is to study this proposal. We start with a brief outline of MaxEnt and an explanation of Konek's anti-luck approach to priors. Next, we explore his perspective on the theoretical role of priors and illustrate how this viewpoint informs the formulation of MaxSen. Then, we explore the impact of the choice of the scoring method on the recommendations made by MaxSen. It turns out that if we use Kullback-Leibler

[3] If, for instance, we know that we are dealing with a continuous variable with finite variance, MaxEnt recommends the normal distribution instead. If we know that we are dealing with a discrete distribution of counts with two unordered possible events at each observation and a constant expected value, the recommended distribution is binomial.

[4] In the context of a continuous random variable, the uniform continuous distribution maximizes differential entropy as well if all we know is that it takes value from a specified range.

[5] For more sophisticated arguments in favor of MaxEnt see (Williamson, 2010)

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press

Divergence (KLD) as a distance measure (instead of Cramer-von-Mises, which Konek originally used), the recommendations are the same as those of MaxEnt. Next, we estimate how quickly the recommended prior distribution concentration increases with the planned sample size.

Some caveats before we move on. Throughout the paper, for computational ease, we will focus on a discrete hypotheses grid-approximating a continuous parameter space. All the conceptual points we will make apply when this restriction is removed.[6]

3. The anti-luck approach to priors

Konek proposes an *instrumental account of priors*. He argues that priors have a more important role than just being used as a representation of epistemic states: "I suggest that the central role of priors is to help us secure accurate posterior beliefs and to minimize our need for epistemic luck in securing those beliefs" (Konek, 2013, 3).

The theoretical role of priors, on this account, is to optimize for accurate posterior beliefs while minimizing the impact of epistemic luck, if possible. Thus, Konek rejects the view that the main role of priors is to represent an agent's epistemic state in radical uncertainty.

Let's analyze Konek's stance. We will start with the notion of **epistemic luck**. Think of accuracy as the proximity between a posterior distribution, reached by Bayesian conditioning, and the true hypothesis.[7] Having accurate posteriors can be accidental. I could have randomly favored one hypothesis without any research whatsoever and still be lucky enough to be right about it, so I can be lucky to have an accurate belief. Yet, according to the rather plausible *incompatibility thesis* (Hetherington, 2019), epistemic luck is incompatible with knowledge. The view stems from an extensive debate surrounding Gettier-style cases in epistemology (Gettier, 1963), in which the subject, due to some coincidence, seems to have justified true beliefs, but fails to have knowledge. One of the explanations on the market is that in such cases, the subject is just epistemically lucky in being right (Engel Jr., 1992), which, arguably, undermines their claim to knowledge.

Two kinds of epistemic luck ought to be distinguished: *environmental* and *intervening*. Environmental luck has a passive influence on an agent's success. The success of a pro football player in scoring a goal is explained by her skill, but of course, she was lucky in that e.g. her shoe did not fall off or that she was not hit by the water bottle thrown by a fan of the opposing team. Intervening luck, however, destroys the simplicity of the causal link between skill and success. It is the kind of luck in which, for instance, a ball kicked by a pro football player was hit by a water bottle mid-air, losing its trajectory, but luckily happened to hit the referee in the head and finally landed in the

[6]Of course, there is a sense in which entropy can be applied to continuous hypothesis spaces, although it is not a fully straightforward extension of Shannon's entropy. A straightforward extension would assign infinity to any continuous distribution. While conceptually this makes sense— with infinite precision comes infinite informativeness—this approach would make a comparison of such distributions uninteresting. If we ignore the part of the continuous version of the entropy formula which tends to infinity, we are left with differential entropy. We will prefer to discretize continuous distributions with fixed levels of precision so that we can make sensible comparisons between grid-approximated continuous and other discrete distributions in a convenient manner.

[7]We take Bayesian conditioning to be a correct updating method and do not concern ourselves with accuracy-centered arguments for it.

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press

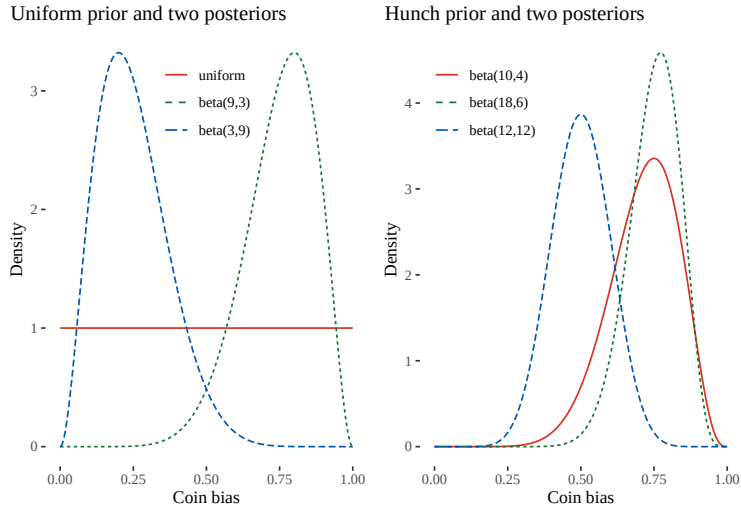


Figure 1. The continuous red line on the first plot is the distribution that MaxEnt recommends in the no-evidence scenario. It is the uniform beta distribution ($\alpha = \beta = 1$). The prior on the second plot, represented with a continuous red line, is a lucky guess, a beta distribution with $\alpha = 10$, $\beta = 4$. The short-dashed green lines represent the posteriors constructed upon observing 10 tosses resulting in 8 heads and 2 tails. The long-dashed blue lines represent alternative posteriors obtained after observing 2 heads and 8 tails.

gate. In such a situation, success cannot be explained by the skill of a football player. The kind of epistemic luck that Konek suggests epistemic agents should strive to avoid is intervening luck.

Crucially, priors might be subject to intervening epistemic luck: the true chances, the true probability, or the truth of a proposition might simply happen to fall in an appropriate sense close to the prior distribution. Priors that are more susceptible to such luck than others seem less reasonable. Therefore, the intuition is that we should strive to minimize the impact of intervening luck on our beliefs.

What makes priors susceptible to this kind of luck? This susceptibility is linked to the distribution's *resiliency* to data. A prior that is more resilient, with respect to a fairly wide range of data, tends to be more luck-dependent. If a prior distribution p is resilient to a datum D, then p conditioned on D is close to p. A resilient prior is less susceptible to over-fitting and changes in its shape in the face of new data. For instance, concentrated prior distributions are far less flexible and more resilient than the uniform distribution, and so they are far more dependent on the luck that real chances happen to fall close to the posteriors reached from them. Figure 1 illustrates the difference between the resiliency of the uniform distribution on the left and of a more concentrated distribution on the right.

The first observation is that the uniform prior changes dramatically and is thus very susceptible to new data. On the other hand, the hunch-based prior, represented with a continuous red line on the right-hand side, did not move much, as the dashed green line that represents the posterior is not too far from the prior. The hunch-based prior is much more resilient to this data.

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press

The second observation is more apparent when we compare the posterior on the second plot with the alternative posterior represented by the long-dashed blue line. This line represents the interaction of the concentrated prior with the "opposite" data (2 Heads and 8 Tails). The resulting high accuracy of the first posterior, in an example with the biased prior, comes from a lucky guess. By luck, the prior bias was close to the observed data. The resilience of a prior is correlated with its susceptibility to intervening luck, the luck that chances happened to be close to the prior distribution.

Konek argues that in adopting a prior distribution, a rational agent ought to choose the prior that best meets the primary theoretical role of priors, that is, optimizes for accuracy while avoiding susceptibility to luck. So far, we talked about luck; let us turn to **accuracy** now.

To make a comparison of two different distributions, Konek uses the Cramer-von Mises distance (CvM). It is, roughly speaking, calculated by taking the squared distance between two cumulative distributions. It's noteworthy that while Konek doesn't strongly argue for the exclusive use of CvM, it is a widely recognized metric commonly employed for making such distribution comparisons.

$$\mathscr{C}(p, p_D) = \int_0^1 [P(x) - P_D(x)]^2 \, dx \qquad \text{(CvM)}$$

where $P(x)$ (P_D) is the cumulative distribution corresponding to p (p_D). The measure is similar to the Brier score. Both methods rely on squared distance; the main differences are that the Brier score works with squared differences of probabilities and takes a sum, while the Cramer-von Mises works with squared differences in cumulative probabilities and integrates. As we want to facilitate computation without hindering insight and to compare distributions of various shapes, including ones that assign probability 1 to a single parameter value (preferably without using Dirac-Delta), we will often follow the practice of using grid approximation with fixed precision levels even in the calculation of CvM, simply making sure that the precision level remains the same for all the comparisons, and using summation instead of integration in such approximations. For now, we will stick to integration to closely follow Konek's work first.

Using (CvM) we can measure the resiliency of our two prior distributions. If we calculate these using the official definition and integration, the resulting values are $\mathscr{C}(u, u_D) = 0.093$, and $\mathscr{C}(l, l_D) = 0.001$.[8] The outcome is a distance between the prior and its posterior, which depends on two factors: the prior's shape and the data. For resilient distributions, this distance will be small, because their spiky prior shifts reluctantly, and the uniform prior will move significantly when updated by the same data. Thus, the distance between the prior and the posterior can be used to gauge the resiliency of a given prior to the data.

A posterior can be also scored by its distance to the true hypothesis (Konek uses CvM here as well)—the inaccuracy of a distribution. If we calculate this value for each particular possible chance hypothesis, thinking about the result as a function of the true

[8]If we use grid approximation of 1001 evenly spaced values, we obtain $\mathscr{C}_{disc}(u, u_D) = 92.832$, and $\mathscr{C}_{disc}(l, l_D) = 3.429$. While the exact values are much larger (as expected when summing a thousand squared distances), the proportional differences are similar (around 22 vs. 27), and moving to the approximation preserves ordering. Assuming that in a given calculation we stick to one of the methods, the exact values do not matter that much for the purposes of comparison.

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press

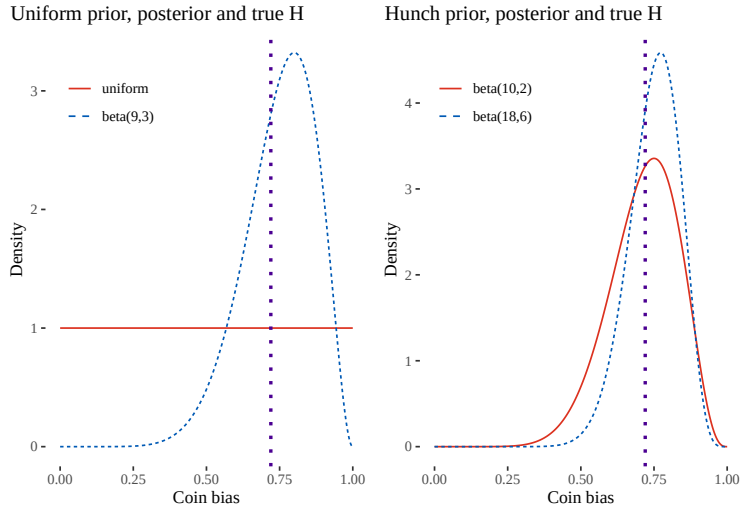


Figure 2. The plot on the left represents the uniform prior, its posterior, and a true hypothsis; the plot on the right starts with the biased prior. The red continuous lines represent priors, and the blue short-dashed ones represent the posteriors. The straight purple dotted lines represent the true hypothesis *h*.

chance hypothesis, we obtain what Konek calls the *objective expected inaccuracy* (OEI). OEI is, according to Konek, a preferable way of representing the prior's accuracy. In the experimental setting, we do not possess knowledge about true chances; therefore, we should ensure arriving at an accurate posterior, regardless of what the world would be. OEI's shape is explained by the two factors:

- The internal factor—the prior's resiliency, that is, its responsiveness to new data. A prior that does not eagerly move will not adjust to new evidence, which affects its accuracy.
- The external factors, such as the proximity of the coin's real bias to the prior distribution.

These two factors are connected. Resilient priors, if they are postulated without any evidential ground, strongly rely on a special kind of intervening luck, which qualifies as an external factor. According to Konek, priors are resilient when their CvM distance to the updated distribution ($\mathscr{C}(p, p_D)$) is close to zero.

However, in the selection of priors, it is not directly resiliency, but rather (expected objective) inaccuracy that plays the key role in Konek's proposal, as he aims at minimizing luck dependency. Let us now come back to the plotted examples with coins.

The plots in Figure 2 illustrate our two priors, lucky and uniform, and the posteriors they lead to after observing the same data (8 Tails, and 2 Heads). The purple, dotted, straight line represents the assumed true hypothesis h from the set of all possible hypotheses H, according to which *the bias of the coin is equal to 0.72*. We now use (CvM) to measure the distance between the hypotheses h and cumulative posterior distributions u_D and l_D: $\mathscr{C}(u_D, h) = 0.035$ and $\mathscr{C}(l_D, h) = 0.026$. The smaller the score, the better, and so the lucky posterior is closer to the true probability of the coin. However, this does not mean that the lucky prior is more reasonable than the uniform one. This

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press

proximity is explainable by intervening luck. In an alternative scenario, an alternative and not-so-lucky outcome would result in a much less accurate posterior.

For this reason, the entire spectrum of possible distances to the alternative potential true hypotheses should be taken into account. Therefore, we should not only calculate the inaccuracy for every possible true state of the world but also consider all possible evidence scenarios. This will result in a distribution of accuracy scores for all the possible combinations for a given prior. What Konek suggests is exactly this: Using the binomial distribution[9] the formula for the (CvM) distance between a posterior distribution u_D and a true hypothesis H (distribution) conditional on the true chance being h, for D consisting of k successes in n observations is as follows:

$$\mathsf{OEI}_{\mathsf{CvM}}(u,h) = \sum_{k=0}^{n} \binom{n}{k} \times h^k \times (1-h)^{n-k} \times \mathscr{C}(u_D, H) \tag{1}$$

That is, to calculate $\mathsf{OEI}_{\mathsf{CvM}}(u,h)$ for each possible outcome D:

- calculate its probability using the binomial formula, assuming the true chance is h, obtaining a weight,
- update u with D, obtaining u_D and calculate the (CvM) distance of u_D from the cumulative distribution corresponding to the true chance being h,
- multiply the distances by their corresponding weights and sum them up. The shape of the objective expected posterior inaccuracy—Konek claims—helps in assessing the epistemic value of a prior, as it does not require any knowledge about what the true hypothesis actually is, and allows us to gauge the sensitivity of (in)accuracy to what the true chance is.

For instance, the OEIs for the priors we used in our running example are illustrated in Figure 3. The first plot represents the expected inaccuracy of a uniform prior in a scenario of 14 coin tosses (trials). The plot on the right represents the OEI of the hunch-based prior $\alpha = 10, \beta = 2$ in the same scenario. Keep in mind that these measurements represent the expected inaccuracy of posteriors built upon those priors and not the inaccuracy of the priors themselves.

OEI for the uniform distribution stays fairly constant across possible true hypotheses and is symmetrical. For the lucky prior, however, we observe a high degree of inaccuracy for low values of the true probability, and inaccuracy similar to the one for the uniform prior as the true hypothesis gets closer to $h = 0.72$. This illustrates how it was pure luck that the hunch-based prior led to a slightly less inaccurate posterior. Indeed, this is possible if the true bias happens to be close to the value around which the prior is centered, but the price is that inaccuracy grows much faster than for the posteriors learned from a uniform prior if the true chance is further from the center of the hunch-based prior distribution.

A general observation is that the MaxEnt-recommended uniform prior is less susceptible to intervening luck when compared to lucky spiky beta priors. However, note that the OEI for the uniform prior is not flat. The question is whether a choice of priors is possible that would make OEI even flatter. Then, from the perspective of taking the

[9]The formal method provided by Konek applies to the binomial distribution only. While perhaps it is possible to develop analogous methods for other distributions, we restrict our attention to the context set up by the proponent of this method.

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press

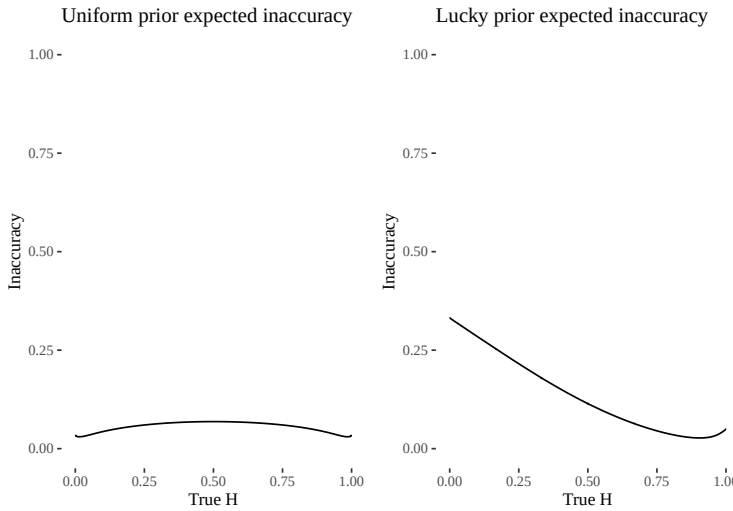


Figure 3. Two plots representing objective expected inaccuracies of: a uniform prior on the left, and a hunch-based prior on the right.

primary role of priors to be the improvement of accuracy, such a prior would minimize the impact of intervening luck, and would come recommended. We arrived at a method of representing the prior's luck dependency with OEI, but it is not clear what exactly a preferable shape of OEI is. In the next section, we will present Konek's approach on how to choose a prior using this tool that will be minimally luck-dependent.

4. Maximum sensitivity

The approach proposed by Konek that ought to recommend a prior minimally susceptible to intervening luck is MaxSen. However, following Konek's methods requires some computation to identify the recommended prior. For the sake of the paper, we will constrain ourselves to beta distributions. The method for identifying the recommended prior for beta distributions is as follows:

- For any combination from a selection of values of α and β and a given planned sample size,[10] compute the corresponding OEI.
- For each such an OEI, compute its curvature (using some sensible curvature measure).
- Use the quadratic approximation of curvature as a function of α and β.
- Use the quadratic approximation to find the values of α and β that minimize this curvature.

As for each particular choice of α and β the calculation of OEI is computationally demanding, the approximation step is practically important.[11] Later on, we will argue that the choice of a quadratic function is sub-optimal once one takes into account sample sizes that go above 100.

[10] Which prior is recommended by MaxSen depends on the sample size, we will come back to this issue.

[11] An analytic approach to the whole problem has not been developed. Perhaps trying to find an analytic method is worthwhile. However we think that a simulation-based study should precede such a search.

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press

alpha	beta	n	CurvK
1.2	1.2	5	0.0169
1.5	1.5	10	0.0140
1.8	1.8	15	0.0103
2.0	2.0	20	0.0096

Table 1. MaxSen recommended beta distributions for sample size n, assuming CvM as a scoring rule and curvature measure CurvK. The grid step size is 0.1.

So here are the key moves that we make in our simulation-based approach, which we first use to successfully replicate Konek's claims, and are about to deploy in further investigations. First, we discretize the range of hypotheses, that is, we look at 1001 evenly spaced chance hypotheses between 0 and 1. Second, we discretize the options for shape parameters (α and β)—on some occasions, the results will be good enough to guide us to decent values, on some occasions, we may still further prefer to further approximate with a quadratic function. Third, as which prior is recommended by MaxSen depends on the expected sample size, we only consider some values of n for the sake of illustration.

Moreover, even once we restrict ourselves to this somewhat simplified setup, there are important choices to be made. First, we need to decide which scoring rule to use. For now, we will stick to Konek's rule of choice based on Cramer-von-Mises distance. Later on, we will introduce and use another rule to investigate how sensitive the recommendations are to such a choice. Second, we need to decide how to measure how uneven the OEI is. Konek's choice is rather crude, as he uses the difference between the maximum and the minimum of a given function:

$$\mathsf{curv}_k(\mathsf{prior}) = \max\left[\mathsf{OEI}(\mathsf{prior})\right] - \min\left[\mathsf{OEI}(\mathsf{prior})\right] \qquad \text{(CurvK)}$$

Konek does not firmly motivate this particular method of scoring curvature, and given that it is only based on two values, a more fine-grained approach might be desirable. We will propose such a measure, which will moreover be based on information-theoretic considerations. For now, we will stick to CurvK.

Figure 4 illustrates a set of beta distributions and their curvature scores. This plot exemplifies how MaxSen works in practice. These distributions are candidates for a maximally sensitive prior, curvature indicates CurvK value of their OEIs (for $n = 5$ and CvM used as a scoring rule). The prior recommended by Konek's method is $\alpha = \beta = 1.2$. In fact, sensitivity is always maximized by a distribution where $\alpha = \beta$, which is also illustrated on the plot by the fact that the values are arranged in a symmetrical pattern and minimal values fall in the middle.

Quite importantly, which $\alpha = \beta$ comes recommended depends on the planned sample size. In Table 1 we illustrate how the recommendations change with $n \in \{5, 10, 15, 20\}$.

For now, we work with small samples, but later on, we will more extensively test two hypotheses: (1) As n increases, so do the MaxSen-recommended $\alpha = \beta$. The functional form of this relation is not clear. (2) As n increases, sensitivity decreases.

Let us reflect for a moment here. Why should the prior with the lowest OEI curvature be considered maximally sensitive and rational? Konek argues, relying on his theoretical

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press

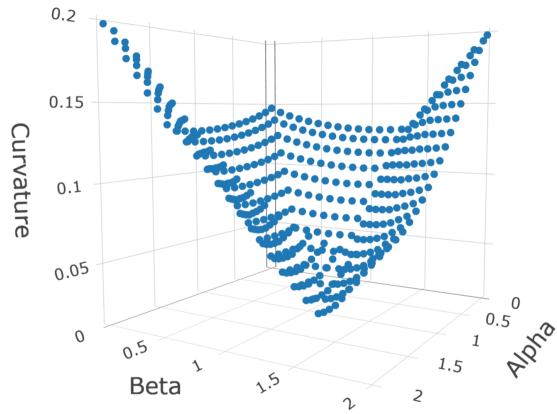


Figure 4. Curvature for a discretized range of possible values of α and β, restricted to (0, 2) with grid step=0.1. These values determine prior distributions, which with sample size (here, $n = 5$) and the choice of a distance measure (here, CvM), determine OEI's. These, with a curvature measure (here, CurvK) determine curvature scores, which are on the vertical axis. The OEI with the lowest curvature is the one that maximizes sensitivity in Konek's sense. In this example, it is $\alpha = \beta = 1.2$.

role of the priors, that we want the data to explain the posterior distribution's accuracy, and not the external factors related to a researcher's hunch. Low OEI curvature indicates that there is no such bias towards any particular chance hypothesis and indicates that the prior is more or less equally sensitive to new data whatever the actual observations will be. For this reason, arguably, it is resistant to intervening luck.

Quite interestingly, uniform priors do not maximize sensitivity when it's measured using Konek's combination of CvM-CurvK, which one might find somewhat surprising—as this suggests that the distribution that is more **equally sensitive to evidence, whatever evidence will be** is not the one that is maximally sensitive to evidence in general. In the next section, we investigate if this fact holds for a different scoring rule and a different curvature measure, though.

Before we move on, two more issues with MaxSen that we mentioned in passing. One is that due to the lack of a known analytic solution, this method requires quite a bit of computation, especially when we consider continuous beta distribution parameters and significantly larger sample sizes, and so systematic use thereof would require either analytic solutions or better approximation methods. Another is that one might think that at least in some contexts, the difference between the priors recommended by MaxEnt and those recommended by MaxSen might not be too important, as the washing out theorem (Joyce, 2004) suggests that if the priors are relatively flat, as soon as the datasets are interestingly large, the difference between the posteriors are not going to matter. This

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press

would suggest that the choice of priors is insignificant to start with. This would be true only under the conditions needed for the theorem to hold, where the problem is not too multi-dimensional and the dataset available is large enough; in other contexts, the choice of priors and regularization methods matters quite a lot.

5. Choice of measures and their impact

5.1. Scoring with KLD

OEI calculations for a given prior require a distance measure. While Konek's method of choice is CvM, other mathematical tools can be used as well. One such measure, compelling due to its information-theoretic motivations, is Kullback-Leibler divergence (KLD) (Kullback and Leibler, 1951). In this section, we will investigate the consequences of using KLD instead of CvM. As it turns out, the end result is importantly different, as **MaxSen in this configuration recommends the same prior as MaxEnt.**

The Kullback-Leibler divergence (KLD), also known as relative entropy, is a commonly used tool for comparing a model distribution to a true, unknown distribution, as it measures the information lost when approximating the true distribution with the model distribution. Since we are working with grid approximation, we will continue to use summation instead of integration.[12] For two distributions P and Q over the same hypothesis space, where i ranges over all possible hypotheses, the KLD from Q to P is:

$$\mathscr{KLD}(P \parallel Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \qquad \text{(KLD)}$$

Arguably, given that KLD is a more theoretically grounded measure than CvM because it is rooted in information theory and has a clear statistical interpretation, KLD provides a more robust and better motivated solution for comparing distributions.

It is essential to note that from this point on, we will assume that the beta distribution that maximizes sensitivity is always symmetrical, i.e., $\alpha = \beta$. We will not include skewed beta distributions since we did not observe any unsymmetrical beta distribution that would be suggested prior to using any of the methods we have used. However, please be aware that this is an assumption, and we do not have proof of this at the moment.

Let us run some examples to compare what happens when we replace CvM with KLD in the deployment of MaxSen. Figure 5 illustrates some OEI curvature measurements (with CurvK) for a range of sample sizes. Since computations become cumbersome for larger values of n, we selected a few values not exceeding 100. For each of those values, we will identify the $\alpha = \beta$ parameters that minimize the curvature.

Figure 5 illustrates an intriguing phenomenon: MaxSen with KLD recommends the uniform priors ($\alpha = \beta = 1$) as they maximize sensitivity, at least for $n \in \{5, 10, 15, 20, 25, 30, 100\}$. While we do not have a general proof that MaxSen with KLD recommends a uniform prior for every sample size, it holds true for every sample size we tested.[13] This suggests that MaxSen's recommendations are not stably (in the

[12] To apply the KLD measure, we have used the `kullback_leibler_distance()` function from the `philentropy` package.

[13] We tested a sequence of values ranging from 0 to 100 with an increment of 5, as well as additional values of 150, 200, and 500.

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press

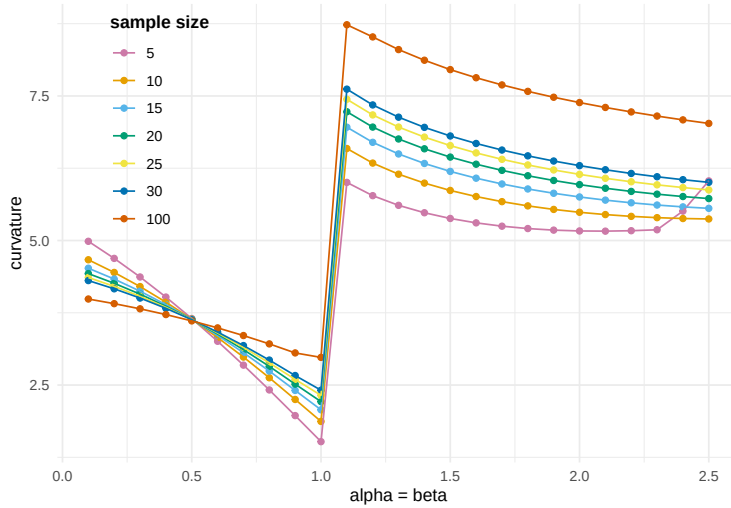


Figure 5. The plot shows seven different MaxSen searches for given sample sizes (n) with a grid approximation of 0.1. For all sample sizes, the recommended prior is $\alpha = \beta = 1$ since it consistently exhibits the lowest OEI curvature. The points represent real measurements, while the lines are included only for enhanced visibility.

sense of being preserved under a sensible range of scoring rules) different from those of MaxEnt.

5.2. Scoring-rule/curvature configurations

In the previous section, we observed that the selection of a scoring rule significantly impacts whether MaxSen recommends regularization instead of following MaxEnt's recommendations. Now, let's explore another parameter: the curvature measure. Konek utilizes CurvK, which captures the difference between the maximum and minimum values of OEI. However, this measure is not very fine-grained as it relies on only two data points of OEI. To address this limitation, we examine the implications of replacing CurvK with an information-theoretic curvature measure, denoted as CurvE. The objective of this section is to compare the results obtained by combining the choice of curvature measure with the selection of a scoring rule in all configurations: KLD/CvM with CurvK/CurvE.

CurvE measures how much less entropy the given distribution has compared to a uniform distribution:

$$\mathsf{curv}_e = 1 - \frac{H[\mathsf{OEI}(\mathsf{prior})]}{H(\mathsf{uniform})} \qquad \text{(CurvE)}$$

where $H[\mathsf{OEI}(\mathsf{prior})]$ is the entropy of normalized OEI corresponding to the prior in question, and $H(\mathsf{uniform})$ is the entropy of a uniform distribution as calculated on the same discretized grid. CurvE ranges from 0 to 1, with higher values indicating a greater deviation from the uniform distribution and, therefore, greater curvature. CurvE, being

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press

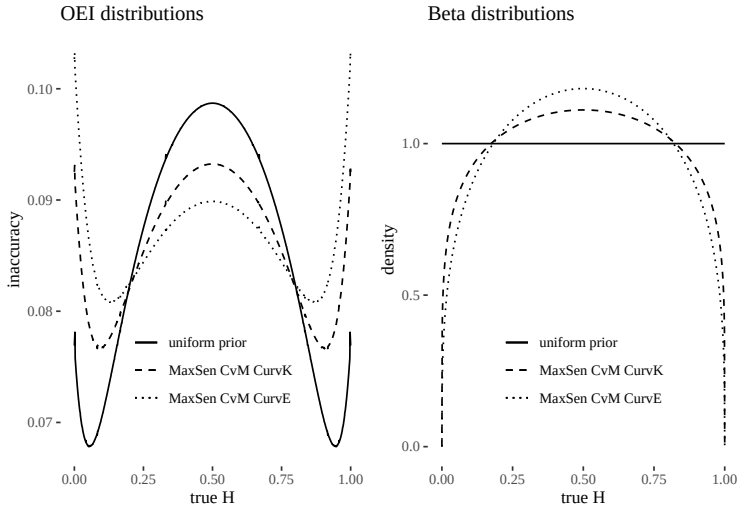


Figure 6. Shape comparison of three OEIs is shown on the left. The dashed line represents MaxSen CurvK ($\alpha = \beta = 1.19$), the dotted line represents MaxSen CurvE ($\alpha = \beta = 1.32$), and the continuous line represents the uniform prior OEI recommended by MaxEnt. The OEI was obtained with an approximation to 0.01, and the number of trials, n, is equal to 5. The plot on the right represents the beta distributions of those priors. The dashed line corresponds to Konek's MaxSen, the continuous line represents the uniform distribution, and the dotted line represents the CurvE MaxSen prior.

sensitive to the entire shape of the distribution appears to be a more justifiable measure of curvature than CurvK.

Figure 6 illustrates three priors recommended by different methods for a sample size of $n = 5$. Two variations of MaxSen, one utilizing CurvK and the other incorporating CurvE, employ CvM as a distance measure. The third prior is recommended by MaxEnt. CurvK recommends a prior of $\mathsf{beta}(1.19, 1.19)$, CurvE recommends $\mathsf{beta}(1.32, 1.32)$, and MaxEnt recommends a uniform distribution. On the left side of the figure, we can see the distributions of their OEIs while on the right side, the beta distributions of those priors are displayed.

The overall shape of the OEIs for these priors is relatively similar. One observation is that the values of OEI of the uniform prior are much lower when compared to MaxSen CurvK for extreme values of the assumed true hypothesis but higher near the center. Another is that prima facie it is not obvious that the OEI curve for the uniform distribution is the flattest.

On the beta distribution plot in Figure 6, the priors are markedly different. The shape of the uniform prior is a straight line parallel to the x-axis. In contrast, the MaxSen priors exhibit a bell-like shape, with near-edge values at 0 and the majority of values slightly above the uniform distribution. The CurvK and CurvE distributions share a similar shape. The bell-like shape of the MaxSen prior may be the most optimal for adjusting its position based on new data, which could explain why the MaxSen method does not recommend the uniform prior.

Now, let's consider a similar test, but this time we will not use the CvM as a measure of distance. Instead, we will utilize the Kullback-Leibler divergence (KLD). In the visualizations shown in Figure 7, once again, the left side displays the distribution of OEIs,

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press

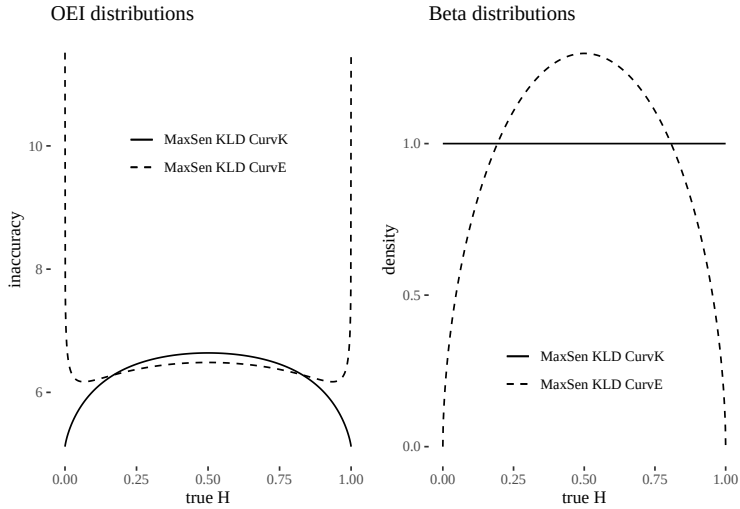


Figure 7. Shape comparison of two OEIs is shown on the left. The dashed line represents MaxSen KLD CurvK ($\alpha = \beta = 1$), and the continuous line represents MaxSen KLD CurvE ($\alpha = \beta = 1.55$) expected inaccuracy distribution. The OEI was obtained by rounding to two decimal places, and the sample size is 5. The plot on the right represents the beta distributions of those priors.

while the right side showcases the beta distributions of the respective priors. Here, you can observe two priors recommended for $n = 5$ using CurvK and CurvE, both of which employ KLD. It is important to note that there are only two priors this time, as we discovered in the previous section that MaxSen KLD with CurvK recommends uniform priors (at least for the tested values of n).

The CurvE OEI exhibits an unusual shape where its marginal values are significantly distant from the rest. All priors generated by MaxSen KLD likely have these outlier-like marginal values, which is why CurvK, being sensitive to disparities between the maximum and minimum values, recommends the uniform prior. The beta distributions plot is less surprising. The CurvE prior is similar to the priors we observed before, with the only difference being that it appears to be more peaked, and more values are above 1 (which is indicated by the uniform prior) compared to the previous examples.

To see how the different approaches to MaxSen change the recommendations, refer to Figure 8, which shows the various n size outcomes for all the configurations mentioned in this section. One striking observation is that the recommended α and β values seem to increase with the number of observations, n. This is a problem that will be explored in the next section.

5.3. Impact of sample size on MaxSen priors

One problem of the MaxSen setting is that the recommended priors $\alpha = \beta$ parameters increase with the planned sample size. The setting that is free of this issue is Curvk KLD, as it's always recommending a uniform prior. This similarity to classical hypothesis testing methods, which are sensitive to stopping intentions, may raise concerns. It is worth mentioning that MaxSen is not alone as the dependency on sample size is a known problem of some prior acquiring principles (Darnieder, 2011).

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press

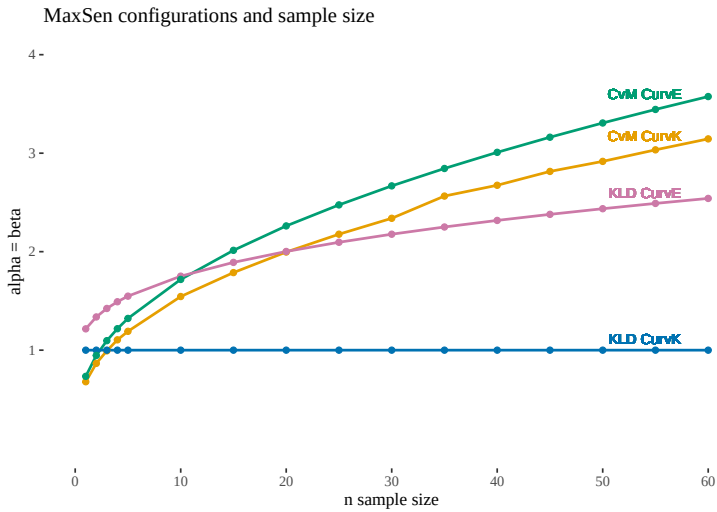


Figure 8. The four lines represent 4 MaxSen methods and the recommended $\alpha = \beta$ depending on the n sample size. Except for KLD CurvK, all the methods recommend priors whose parameters increase with the sample size. The smoothed lines go through the point values obtained for $n = \{1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60\}$ a (non-analytic) gradient descent.

Figure 9 illustrates some MaxSen priors (using CvM and CurvK) and provides information on the relationship between sample size and beta distribution parameters. Those calculations are represented with points. These calculations were performed using gradient descent, introducing some degree of uncertainty to the values. At least in some proximity to the values for which approximations have been obtained, the relationship between the shape parameters and the sample sizes is close to exponential (models with a few formulaic shapes, such as logarithmic or polynomials, were also tested, but displayed worse performance). A Bayesian model that predicts the $\alpha = \beta$ based on the sample size n is visualized as a line, where the orange line represents the mean prediction and the gray area represents the 89% highest posterior density interval (HPDI) of the predictions.

The approximating learned function that best describes the growth of the recommended α and β parameters with the sample size (the exponent $1/2.5$ is the result of an optimization) n is:

$$f(x) = 0.61 \cdot n^{\frac{1}{2.5}}$$

6. Experimental design applicability

In many cases, the sample size is not known prior to the experiment and can be difficult to predict. Some experimental designs (Kruschke, 2015) might make it even in principle impossible to predict the sample size prior to the experiment (a simple example is flipping a coin until observing 30 heads—you can't predict how many tosses you will need to make). In such situations, where the stopping intention influences the data, researchers following Konek's recommendation, it seems, would be forced to declare

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press

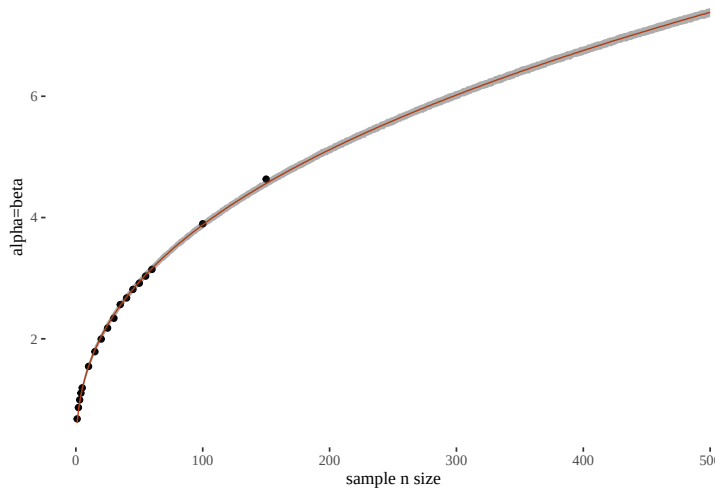


Figure 9. MaxSen (CvM, CurvK) results for different sample sizes are depicted in the plot. Each point represents an optimized value of $\alpha = \beta$, ranging up to 150 with irregular breaks. The line represents a learned exponential model predicting the observed trend. The gray area represents the 89% highest posterior density interval (HPDI), while the continuous line represents the mean prediction.

the sample size to be able to choose a prior. Konek himself admits some of these limitations (Konek, 2016): My aim now is to use this statistic to help us secure nice posterior credences—credences that are not only reasonably likely to be accurate but whose accuracy is the product of cognitive ability—in a wide range of inference problems. Inference problems, of the sort familiar from classical statistics, arise in structured but nonetheless fairly ubiquitous contexts of inquiry [. . . in which . . .] we know enough about the data-generating process (the experiment) to know which data items could be produced, and what the chances of receiving each of the various possible data items would be, were these different theoretical hypotheses true (typically as a result of careful experimental design)

[. . .] What does this tell us about the more complicated inference problems that arise in the sciences, and so forth? [These tools. . .] will also help us secure such knowledge in any inference problem that involves performing a fixed number of identical and independent trials of an experiment with only finitely many possible outcomes. Quite a lot of inference problems in the sciences, engineering, and so forth, are like this. [p. 538]

[. . .] in some contexts of inquiry, we simply do not know enough about the data-generating process to know which data items our experiment might produce (what statisticians call the sample space of the experiment) [. . .] in still other contexts of inquiry, new scientific data is simply not generated by running IID trials of an experiment, or anything of the sort, for example, in paleontology, geology, and so forth. In such contexts, it is not obvious how to generalize MaxSen to help secure probabilistic knowledge. Let us isolate at least two issues here. One question is: as the method is given only for cases in which the number of observations is fixed, how to generalize to context it which it isn't? Another is, with the same observations in the future, can the method recommend different priors depending on what experimental design intentions we start with?

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press

The answer to the first question is not completely disappointing. after all, assuming the experimenter has some probability distribution over the number of observations they might be making, this distribution itself can be used in the computation of double expectations while preserving the spirit of the proposal. Sure, some challenges remain: computations of double expectations are computationally tricky, and it is not quite clear how the experimenter should have chosen their priors for guessing the number of planned observations, but perhaps all hope is not lost.

More crucially, however, even if one decides on some stopping method and comes up with a distribution over observations resulting, for any number of particular actual observations n made in the future, this approach will recommend a different prior than the one in which the experimenter simply intends to collect n observations. This means that for exactly the same number of observations—for the same observations even—the stopping intention paired with MaxSen will recommend different priors, and lead to different posteriors, based on the experimenter's intentions. This seems to suggest that, insofar as sensitivity to stopping intentions goes, the proposal is maybe not in the same wagon but at least on the same train as the classical approaches involving p-value computations (see ch. 11 of (Kruschke, 2015) for a more detailed criticism of classical statistics from this perspective).

Another significant concern arises when the method leads to a high concentration of the prior distribution for large sample sizes. Many real-life applications of Bayesian methods involve much larger sample sizes than those we have discussed, making it crucial to investigate what MaxSen (and its variations) recommends for large values of n— whether practitioners will be convinced to accept extremely strong priors for real-life problems with large datasets, remains unclear.

7. Conclusions

We have discussed the Maximum Sensitivity principle for choosing priors under radical uncertainty. The method focuses on the primary theoretical role of priors: mitigating reliance on epistemic luck. Konek argues that a rational agent should adopt priors that are maximally sensitive to achieve minimally luck-dependent priors. This sensitivity is measured by first calculating the objective expected posterior inaccuracy OEI of a prior and selecting the prior with the flattest OEI curvature—where distances are measured using the Cramér-von Mises (CvM) and the range of possible values (CurvK) is the curvature measure.

Having explored Konek's view, we tested four variations of the MaxSen method. We introduced the Kullback-Leibler Divergence (KLD) as the distance measure and CurvE as a curvature measure with better theoretical support. The four configurations are: CvM CurvK, CvM CurvE, KLD CurvK, and KLD CurvE. Two key observations result:

(1) With the exception of KLD CurvK, all the methods recommend priors in which the sample size n grows with $\alpha = \beta$. The concentration of the prior, influenced directly by the $\alpha = \beta$ parameter, increases with n.
(2) KLD CurvK consistently recommends uniform priors, suggesting that in a certain configuration, MaxSen, a method that considers priors only as instrumental, recommends a prior that is also minimally informative.

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press

In general, it is not obvious which configuration of MaxSen makes the most sense, and decisions in this dimension lead to different recommendations. The fact that MaxSen in one fairly sensible configuration (KLD CurvK) recommends uniform priors suggests that MaxEnt can be interpreted as a sensitivity-maximizing rule to start with. Moreover, from the options we considered, it is the only one that is resilient to the problem of growing $\alpha = \beta$.

Further research is required to expand the MaxSen method to encompass more intricate modeling choices, extending beyond the constraints of binomial distributions. Of particular interest is the validation of a claim that KLD CurvK consistently endorses a uniform prior. This validation should not solely rely on example calculations but rather be substantiated with formal proof.

Acknowledgments

The authors thank the anonymous reviewers for their helpful comments.

Funding Statement

This work was supported by the National Science Centre, Poland (Narodowe Centrum Nauki) under grant number 2021/41/B/HS1/01814.

Declarations

Competing interests: the authors declare none.

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press

References

Darnieder, W. (2011) *Bayesian Methods for Data-dependent Priors*. Ohio State University.

De Finetti, B. (1979) *Theory of Probability. A critical introductory treatment*. John Wiley & Sons.

Engel Jr., M. (1992) Is epistemic luck compatible with knowledge? *The Southern Journal of Philosophy*. *30*(2), 59–75. https://doi.org/10.1111/j.2041-6962.1992.tb01715.x.

Gelman, A. (2009) Bayes, jeffreys, prior distributions and the philosophy of statistics. *Statistical Science*. *24*(2), 176–178. https://doi.org/10.1214/09-STS284D.

Gettier, E. L. (1963) Is justified true belief knowledge? *Analysis*. *23*(6), 121–123. https://doi.org/10.1093/analys/23.6.121.

Hetherington, S. (2019) chapter26 in *The Luck/Knowledge Incompatibility Thesis*. Routledge. https://doi.org/10.4324/9781351258760-27.

Joyce, J. M. (2004) Bayesianism . In *The Oxford Handbook of Rationality*. Oxford University Press. Available at: https://doi.org/10.1093/oxfordhb/9780195145397.003.0008.

Konek, J. (2016) Probabilistic knowledge and cognitive ability. *Philosophical Review*. *125*(4), 509–587. https://doi.org/10.1215/00318108-3624754.

Konek, J. P. (2013) *New Foundations for Imprecise Bayesianism.* Ph.D. thesis.

Kruschke, J. (2015) *Doing Bayesian Data Analysis (Second Edition)*. Academic Press. Boston.

Kullback, S. and Leibler, R. A. (1951) On information and sufficiency. *The annals of mathematical statistics*. *22*(1), 79–86. https://doi.org/10.1214/aoms/1177729694.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E. and Rakow, T. (2006) *Uncertain judgements: eliciting experts' probabilities*. https://doi.org/10.1002/0470033312.

Williamson, J. (2010) *In defence of objective Bayesianism*. Oxford University Press Oxford. https://doi.org/10.1093/acprof:oso/9780199228003.001.0001.

https://doi.org/10.1017/psa.2026.10266 Published online by Cambridge University Press