CrossMark

# Bayesian Convergence and the Fair-Balance Paradox

Bengt Autzen[1] <span>ORCID</span>

**Abstract** The paper discusses Bayesian convergence when the truth is excluded from the analysis by means of a simple coin-tossing example. In the fair-balance paradox a fair coin is tossed repeatedly. A Bayesian agent, however, holds the a priori view that the coin is either biased towards heads or towards tails. As a result the truth (i.e., the coin is fair) is ignored by the agent. In this scenario the Bayesian approach tends to confirm a false model as the data size goes to infinity. I argue that the fair-balance paradox reveals an unattractive feature of the Bayesian approach to scientific inference and explore a modification of the paradox.

## 1 Introduction

The problem of convergence to the truth in Bayesian inference has been widely discussed in the philosophical literature (e.g., Hesse 1974; Glymour 1980; Earman 1992; Kelly 1996; Hawthorne 2011; Belot 2013). Convergence to the truth results establish conditions under which the degrees of belief of Bayesian agents become more and more tightly peaked around the true hypothesis as the data accumulate. A general assumption of Bayesian convergence theorems is that the true hypothesis is included in the set of candidate hypotheses. In the discrete probability spaces containing a finite set of statistically simple hypotheses that are frequently

✉ Bengt Autzen
  b.autzen@bristol.ac.uk

[1] Department of Philosophy, University of Bristol, Cotham House, Bristol BS6 6JL, UK

considered in the philosophical literature this amounts to the requirement that the true hypothesis gets assigned non-zero prior probability.[1]

In this paper I am interested in a different problem: what happens if the truth is excluded in a Bayesian analysis? In particular, what happens if the true model is excluded and the false candidate models are equidistant from the truth in Bayesian model selection? The example I will explore looks fairly benign. Suppose a fair coin is tossed repeatedly and a Bayesian agent holds, for whatever reason, the a priori view that the coin is either biased towards heads or towards tails. As a result the truth (i.e., the coin is fair) is ignored by the agent. The question that I will address is what degrees of belief the agent will adopt in the long run as the number of coin tosses goes to infinity.

In order to study the coin-tossing example in detail, its probabilistic assumptions have to be specified and some terminology has to be introduced. It is assumed that the coin tosses are independent and identically distributed with parameter $p$ denoting the probability of the coin landing 'heads' in a single coin toss. The number of 'heads' in $n$ tosses is then described by the Binomial distribution $B(n, p)$. I will refer to a 'model' as a family of probability distributions. For instance, the family of Binomial distributions $B(n, p)$ described in terms of the parameters $n$ and $p$ qualifies as a model. Every numerical choice of $n$ and $p$ specifies a particular probability distribution describing the number of 'heads' in $n$ coin tosses. For any fixed $n$, I will consider three models: the fair-coin model $M_F$ containing only the Binomial distribution with parameter $p$ equal to $\frac{1}{2}$, the head-bias model $M_H$ containing all Binomial distributions with $p > \frac{1}{2}$ and the tail-bias model $M_T$ containing all Binomial distributions with $p < \frac{1}{2}$.[2] Since the agent is indifferent about whether the coin is biased towards heads or towards tails, she assigns equal prior probability to the two candidate models (i.e., $P(M_H) = P(M_T) = \frac{1}{2}$). As a result the true model $M_F$ is excluded, that is, $M_F$ has zero prior probability in the discrete model space.[3] Given the head-bias model $M_H$, she is indifferent with regard to the precise numerical probability $p$ and assumes that $p$ follows a uniform distribution on the interval $(\frac{1}{2}, 1)$, denoted as $U(\frac{1}{2}, 1)$. Similarly, she assumes that parameter $p$ follows a uniform distribution on the interval $(0, \frac{1}{2})$ given the tail-bias model $M_T$.

Before assessing the limiting behaviour of the model posterior probabilities in the coin-tossing example, some general comments on the approach of this paper are in order. Considering the situation in which the prior degrees of belief of an agent are

---

[1] In continuous probability spaces matters are more complicated. Here the requirement is relaxed to the effect that each open subset containing the true hypothesis has non-zero prior probability. In general, including the true hypothesis in the support of the prior is necessary but not sufficient for convergence to the truth. Freedman (1963) shows that in the case of a chance process with a countable infinity of possible outcomes, one can identify a prior with the true hypothesis in its support that can be expected to fail to converge to the truth.

[2] Note that while the truth (i.e., the coin is fair) can be represented either by the single parameter value $p = \frac{1}{2}$ in the continuous parameter set [0, 1] or by means of the trivial model $M_F$ containing only the single probability distribution $B(n, \frac{1}{2})$ in the discrete set of models, the false hypothesis that the coin is biased towards heads (tails) does not correspond to a point hypothesis in the parameter space of $p$.

[3] I will call a model 'true' if and only if it contains the true probability distribution.

not spread out over the space of possible models runs against the methodological advice generally given by philosophers with Bayesian inclinations. Since the posterior probability of a model with zero prior probability will always remain zero by Bayesian updating, Bayesian philosophers generally take a 'liberal' stance when it comes to the assignment of non-zero prior probabilities. However, even when adopting such a liberal attitude the set of candidate models does not necessarily contain the true model. Put more strongly, there are good reasons to believe that identifying a true model before analysing data is too good to be true. Indeed, Gelman and Shalizi (2013) adopt a critical stance towards the idea that in a statistical analysis, a researcher is able to identify a priori a statistical model that captures all the systematic influences among the variables of the system of interest in their correct functional form. They (2013, p. 9) comment that "[t]his could happen, but we have never seen it, and in social science we have never seen anything that comes close". These worries are, however, not exclusive to the social sciences. In climate science, for instance, it is often pointed out that all current climate models are false (e.g., Parker 2009). These considerations naturally lead to the question of what will happen in a Bayesian analysis if the true model is excluded.

The paper is structured as follows. Section 2 introduces some plausible convergence criteria for the case in which the truth is excluded in a Bayesian analysis. Section 3 presents the fair-balance paradox. Section 4 discusses some modifications of the paradox. Section 5 concludes.

## 2 Convergence Without Truth

Under the ideal scenario of an infinitely large data set an inference procedure should show certain desirable features. For instance, in Bayesian parameter estimation a reasonable requirement is that the posterior probability distribution becomes increasingly peaked around the true parameter value for any non-pathological sequence of data. Similarly, the model posterior probability distribution should become peaked on the true model as data size goes to infinity in Bayesian model selection. In our setting, however, the true model $M_F$ has zero prior probability and, hence, the posterior probability of $M_F$ will remain zero by Bayesian updating. So, what would be a reasonable requirement on an agent's degrees of beliefs as the data size goes to infinity? Lewis et al. (2005) propose that ideally the posterior probability of $M_H$ should converge in probability to the constant value 1/2 when $n$ goes to infinity (and the same applies to model $M_T$).[4] That is, the sequence of model posterior probabilities, constituting a sequence of random variables, is supposed to converge in probability to the (trivial) random variable taking only the constant value 1/2 as data accumulate.

Lewis et al.'s convergence criterion can be generalised to what might be referred to as an 'A Posteriori Indifference Principle' (APIP). Rather than considering the

---

[4] A sequence of random variables $X_n$ is said to converge in probability to the random variable $X$ if and only if for all $\epsilon > 0$ the probability $P(|X_n - X| > \epsilon)$ goes to 0 as $n$ goes to infinity.

head-bias and tail-bias models, I will phrase the principle in a slightly more general framework for reasons that will become clear in the course of the paper. Let the generalised head-bias model $M_{GH}$ contain all Binomial distributions $B(n, p)$ with parameter $p$ that lies strictly between $1/2 + c$ and 1 (i.e., $p \in (\frac{1}{2} + c, 1)$), where $c$ is a fixed value satisfying $0 \leq c < \frac{1}{2}$. It is assumed that model $M_{GH}$ has prior probability 1/2 and assigns prior probabilities to parameter $p$ based on the uniform probability distribution on the interval $(\frac{1}{2} + c, 1)$.[5] The generalised tail-bias model then contains all Binomial distributions $B(n, p)$ with parameter $p$ that lies strictly between 0 and $\frac{1}{2} - c$ (i.e., $p \in (0, \frac{1}{2} - c)$). Similarly, it is assumed that model $M_{GT}$ has prior probability 1/2 and assigns prior probabilities to parameter $p$ based on the uniform probability distribution $U(0, \frac{1}{2} - c)$. Given these assumptions APIP reads as follows:

> As the number of fair coin tosses $n$ goes to infinity, the model posterior probability distribution should converge to a probability distribution that is indifferent among the false candidate models $M_{GH}$ and $M_{GT}$.

Having introduced APIP, it is natural to ask how the principle can be motivated. A natural answer invokes Bayesian confirmation theory. According to the 'absolute notion' of Bayesian confirmation, data $D$ confirm hypothesis $H$ if and only if the posterior probability $P(H|D)$ is strictly larger than some threshold value $k$. Further, data $D$ disconfirm hypothesis $H$ if and only if $P(H|D) < k$. The threshold value $k$ is typically set at 1/2 (e.g., Achinstein 2001, p. 46). The reason for this choice of $k$ is that it assures $H$ having higher degree of belief than its negation $\neg H$ after observing $D$, if $D$ confirms $H$. It is typically assumed that an adequate account of confirmation should disconfirm false hypotheses and confirm true hypotheses as the data accumulate (e.g., Hawthorne 2011, p. 336). Applying this dictum to the coin tossing example would demand that both models $M_{GH}$ and $M_{GT}$ are to be disconfirmed as the number of fair coin tosses goes to infinity. However, this requirement violates the axioms of the probability calculus. The best one can expect is that each false model is not to be confirmed as the data size increases. This intuition leads to the requirement that the posterior probability of each model approaches 1/2 as the data size goes to infinity and is captured by APIP.

In addition, the two false models $M_{GH}$ and $M_{GT}$ are equidistant from the truth measured in terms of the Kullback-Leibler (KL) divergence. Following Dawid (1999), the KL divergence between a model $M$ and the true distribution $P$ is then understood as the infimum of the KL divergences between $P$ and the probability distributions in $M$. Given that the two false models $M_{GH}$ and $M_{GT}$ are equally distant from the truth, it should become less probable for the evidence to prefer one model to the other as the data accumulate. This requirement translates into to the condition that the posterior probability of each model approaches 1/2 as the number of coin tosses goes to infinity and is again captured in probabilistic terms by APIP.

Analogous results obtain when adopting the more prominent 'relative notion' of Bayesian confirmation, according to which data $D$ confirm hypothesis $H$ if and only

---

[5] Note that the head-bias model $M_H$ results from choosing $c = 0$ in the specification of the generalised head-bias $M_{GH}$.

if the posterior probability $P(H|D)$ is strictly larger than the prior probability of the hypothesis $H$, $P(H)$. Further, data $D$ disconfirm hypothesis $H$ if and only if $P(H|D) < P(H)$. Since the two candidate models $M_{GH}$ and $M_{GT}$ are assumed to have equal prior probability of 1/2, the intuition that the probability of one model being confirmed goes to zero as the number of coin tosses goes to infinity is again captured by APIP.

The intuition underlying APIP reflects a kind of epistemic modesty by assigning intermediate rather than extreme degrees of belief to the false candidate models in the limit.[6] One could argue, however, that the concern is not necessarily that the model posterior probabilities differ from the precise numerical value 1/2 in the limit but that the model posterior probabilities converge in probability to random variables taking either very large or very small values. As such APIP is to be seen as a stronger version of the following requirement, which might be called a 'Bayesian Modesty Principle' (BMP):

> As the number of fair coin tosses $n$ goes to infinity, the probability that the posterior probability of $M_{GH}$ is larger than, say, 0.9 should converge to 0. The same applies to the posterior probability of model $M_{GT}$.

Again, Bayesian confirmation theory helps to motivate this principle. Suppose we assume the relative notion of confirmation. In contrast to APIP, BMP does not demand that a false model, say, $M_{GH}$ is not confirmed as the data size goes to infinity. As a result BMP cannot be motivated by focusing exclusively on qualitative confirmation statements. In order to illustrate the intuition underlying BMP, we have to consider a quantitative account of confirmation. Quantitative accounts of confirmation involve the concept of a degree of confirmation, which indicates how strongly data $D$ confirm hypothesis $H$. Let us, for instance, consider the difference measure made popular by Carnap (1962)[7]: $d(D, H) = P(H|D) - P(D)$. Suppose $D$ confirms $H$. Then, the larger the value of $d(D, H)$, the stronger the inductive support for $H$ provided by the data $D$. Now, if BMP holds, then the probability of $M_{GH}$ being strongly confirmed goes to zero as the data size goes to infinity (here, 'strongly confirmed' means that the difference measure takes a value that is larger than the arbitrary threshold value 0.4).

## 3 Fair-Balance Paradox

While the previous section provided some arguments for the desirability of APIP and BMP, the question remains whether these principles are, in fact, satisfied. In order to address the empirical validity of these principles, let us focus on the behaviour of the model posterior probability of the head-bias model $M_H$ for the sake of simplicity. Yang (2007) demonstrates that if the the truth is that the coin is fair, the posterior probability of $M_H$ converges in probability to a random variable with

---

[6] An alternative way of looking at APIP is to view this principle as a strengthening or extension of statistical consistency.

[7] More recent proponents of the difference measure include Eells (1982) and Jeffrey (1992).

the uniform distribution $U(0, 1)$ for $n$ going to infinity. That is, the posterior probability of the false model $M_H$ converges, but not to a constant value. Phrased differently, the posterior probability of model $M_H$ is drawn 'randomly' from the interval $(0, 1)$ when the data sets become infinitely large based on tossing a fair coin. These analytic results are in accordance with simulation studies showing that for data sets of size $n = 10^6$ the posterior probability distribution of $M_H$ mirrors the uniform distribution $U(0, 1)$ (Yang 2007). That is, if you simulate the fair-coin experiment a million times, then the empirical distribution of the posterior probability of $M_H$ approximates the uniform distribution on the interval $(0, 1)$. The phenomenon that the posterior probability of $M_H$ fails to converge to the single numerical value $1/2$ for $n$ going to infinity has been labelled the 'fair-balance paradox' in the biological literature.

The fair-balance paradox reveals an undesirable feature of the Bayesian approach to scientific inference as it violates both APIP and BMP. Consider APIP first. Rather than converging in probability to a random variable with the single value $1/2$ as required by APIP, the posterior probability of $M_H$ converges in probability to a random variable with the uniform distribution $U(0, 1)$ if the coin is fair. Hence, the Bayesian approach tends to confirm one of the false candidate models as data accumulate. Further, the model $M_H$ will be strongly confirmed with probability 0.1 in the limit. So, as the posterior probability of $M_H$ converges in probability to a random variable with the uniform distribution $U(0, 1)$, there exists, in violation of BMP, a non-vanishing probability that this model posterior probability is larger than 0.9 in the limit.

It is important to stress that even though the true fair-coin model $M_F$ has zero prior probability and, hence, the prior probability distribution on the discrete space of models (including the fair-coin model, the head-bias model and the tail-bias model) does not have full support, the entire prior probability distribution on parameter $p$ is of full support in the sense that it assigns positive probability to every open neighbourhood of every point hypothesis regarding the probability of 'heads' of the coin. Phrased differently, in the model selection problem the truth is excluded since the true model $M_F$ has zero prior probability in the discrete model space. In contrast, the truth is in the support of the prior when focusing on the entire prior probability distribution on parameter $p$ in the continuous parameter space. An alternative way of describing the relationship between the model prior and the prior on parameter $p$ is to state that while the prior on parameter $p$ is indifferent over all possible values of $p$, the model prior is not indifferent over the three possible models $M_F, M_H$ and $M_T$.

Since the fair-balance paradox is based on a chance process (i.e., coin tossing) with a finite number of possible outcomes, the prior on parameter $p$ is consistent in the statistical sense of the term (Freedman 1963).[8] This becomes apparent when mapping the posterior probability distribution of parameter $p$: as the data size increases the posterior probability distribution of $p$ becomes more and more

---

[8] A prior probability distribution $P_0$ is consistent at $\theta \in \Theta$ if given hypothesis $\theta$ the probability for observing a sequence of outcomes that gives rise to a sequence of posterior probability distributions $(P_1, P_2, ...)$ that does not become more and more tightly peaked around parameter value $\theta$ is zero. A prior probability distribution $P_0$ is called consistent if it is consistent at every $\theta \in \Theta$.

concentrated around the true parameter value $p = \frac{1}{2}$ (see figure 2 in Lewis et al. (2005)). So, focusing exclusively on the posterior probabilities of the models $M_H$ and $M_T$ in the fair-balance paradox does not provide a comprehensive picture of the underlying chance process.

An agent who thinks that all information necessary for Bayesian model selection is contained in the model posterior probabilities and that these posterior quantities indicate the relative plausibilities of the candidate models is referred to as an 'overconfident' Bayesian by Morey et al. (2013). The fair-balance paradox reinforces the view that an exclusive focus on model posterior probabilities does not provide a satisfactory account of inference as the model posteriors fail to adequately report the relative plausibilities of the two candidate models. In contrast, Morey et al. refer to a 'humble' Bayesian as an agent who questions the models used for inference and invokes a variety of Bayesian tools, including posterior distributions, model odds and Bayes factors for model checking. In a simple example such as the fair-balance paradox already using both the posterior probability distribution on parameter $p$ and the model posteriors suffices to indicate problems with the initial choice of candidate models and, hence, serves the need of the humble Bayesian.

## 4 Modifying the Paradox

One essential characteristic of the fair-balance paradox is its symmetry: the candidate models are equidistant from the truth. Furthermore, the parameter $p$ in the false models $M_H$ and $M_T$ gets infinitely close to the true parameter value $p = 1/2$. While the second feature follows naturally from identifying the hypothesis 'The coin is biased towards heads' with model $M_H$ (and, similarly, identifying the hypothesis 'The coin is biased towards tails' with model $M_T$) and does not affect the example's function to put APIP and BMP to the test, a natural question to ask is what happens in cases where the false candidate models are still equidistant from the truth but do not come arbitrarily close to the true parameter value. One might suspect that the paradox disappears in such a setting.

In order to address this question, I will consider the following two models: the strong head-bias model $M_{SH}$ contains all Binomial distributions $B(n, p)$ with parameter $p$ located strictly between $1/2 + c$ and 1 (i.e., $p \in (\frac{1}{2} + c, 1)$) with a fixed value $c$ satisfying $0 < c < \frac{1}{2}$. As a result the parameter denoting the probability of 'heads' of the candidate model $M_{SH}$ does not get infinitely close to the true parameter value $p = \frac{1}{2}$. Again, it is assumed that model $M_{SH}$ has prior probability 1/2 and assigns prior probabilities to parameter $p$ based on the uniform probability distribution $U(\frac{1}{2} + c, 1)$.[9] The strong tail-bias model $M_{ST}$ then contains all Binomial distributions $B(n, p)$ with parameter $p$ located strictly between 0 and $\frac{1}{2} - c$ (i.e., $p \in (0, \frac{1}{2} - c)$). Similarly, it is assumed that model $M_{ST}$ has prior probability 1/2 and

---

[9] Note that the strong head-bias model $M_{SH}$ results from the generalised head-bias model $M_{GH}$ by excluding the choice of constant $c$ being equal to 0.

assigns prior probabilities to parameter $p$ based on the uniform probability distribution $U(0, \frac{1}{2} - c)$.

In both the fair-balance paradox and the modified coin tossing example the true model has zero prior probability. As a result the model prior is not indifferent over all possible models in both examples. In contrast to the fair-balance paradox where the prior on parameter $p$ does have full support, the truth is not in the support of the prior on parameter $p$ in the modified coin tossing problem. Phrased differently, while the prior on parameter $p$ is indifferent over all possible values of $p$ in the fair-balance paradox, it is not indifferent in the modified coin tossing problem.

The posterior probability of $M_{SH}$ converges in probability to a random variable that takes the value 0 with probability 1/2 and the value 1 with probability 1/2 as the number of coin tosses goes to infinity (see Theorem 1, "Appendix").[10] Given the symmetry of the problem the same applies to the posterior probability of $M_{ST}$. It follows that one of the two false models will, with probability 1, be strongly confirmed in the limit. Even though the resulting limiting behaviour differs between the head-bias model $M_H$ and the strong head-bias model $M_{SH}$, the fair-balance paradox persists since both APIP and BMP are again violated. There is a sense, however, in which the move towards the models $M_{SH}$ and $M_{ST}$ aggravates the problem as the probability of a candidate model being strongly confirmed in the limit increases significantly.

The discussion shows that two plausible constraints on Bayesian convergence, referred to as APIP and BMP, do not hold. Both the original fair-balance paradox involving the head-bias and the tail-bias models and the modified fair-balance paradox involving the strong head-bias and the strong tail-bias models violate these two principles. Indeed, the modified coin tossing problem increases the probability of confirming a false model with a high degree of confirmation.

Before concluding a final comment is in order. Both the fair-balance paradox and its modification consider false models with equal distance from the truth due to the symmetry of the set-up. This approach differs from a situation in which the truth is excluded from the set of candidate models but these models have different distances from the truth. In the latter scenario Bayesian inference typically shows a much more benign face. To illustrate, consider the following two candidate models: The asymmetric head-bias model $M_{AH}$ contains all Binomial distributions $B(n, p)$ with parameter $p$ that lies strictly between $1/2 + c_1$ and 1 (i.e., $p \in (\frac{1}{2} + c_1, 1)$) with a fixed value $c_1$ satisfying $0 < c_1 < \frac{1}{2}$. The asymmetric tail-bias model $M_{AT}$ then contains all Binomial distributions $B(n, p)$ with parameter $p$ that lies strictly between 0 and $\frac{1}{2} - c_2$ (i.e., $p \in (0, \frac{1}{2} - c_2)$) with $0 < c_2 < \frac{1}{2}$ and $c_1 \neq c_2$. Again, it is assumed that the two models $M_{AH}$ and $M_{AT}$ have equal prior probability and assign a uniform prior to parameter $p$ over the relevant intervals. Suppose model $M_{AH}$ is closer to the truth than model $M_{AT}$ (i.e., $c_1 < c_2$). It follows from general results on Bayesian convergence (Dawid 1999) that the posterior probability of the false

---

[10] This result on the asymptotic behaviour of model posterior probabilities sits well with work on Bayesian convergence under a misspecified model (Berk 1966). Generally speaking, if there is no probability distribution that is uniquely closest to the truth, the posterior probability distribution will alternate between concentrating around several minima.

model with the closest distance to the truth (as measured by KL divergence) converges in probability to 1 as the data size goes to infinity.

## 5 Conclusion

Good methods of scientific inference are expected to have desirable limiting features as the data size goes to infinity. The fair-balance paradox and its modification reveal an unattractive feature of the Bayesian approach to scientific inference. When choosing between two false candidate models that are equidistant from the truth, the Bayesian approach tends to confirm a false candidate model when the data size grows infinitely. As such, Bayesian inference violates two desirable principles, the A Posteriori Indifference Principle and the Bayesian Modesty Principle, set out in this paper.

## Appendix

**Theorem 1** *The posterior probability of $M_{SH}$ converges in probability to a random variable that takes the value 0 with probability $\frac{1}{2}$ and the value 1 with probability $\frac{1}{2}$ as the number of coin tosses n goes to infinity. The same applies to the posterior probability of $M_{ST}$.*

*Proof* Let $\bar{p}$ denote the proportion of heads in $n$ fair coin tosses. Further, let $\alpha$ be a real number that lies strictly between 0 and 0.5. Then, by the Berry-Esseen Theorem, the probability of event $E_{-}$ that $\bar{p}$ lies between $\frac{1}{2} - n^{-0.5-\alpha}$ and $\frac{1}{2} - c$ converges to $\frac{1}{2}$ as $n \to \infty$. Similarly, the probability of the event $E_{+}$ that $\bar{p}$ lies between $\frac{1}{2} + n^{-0.5-\alpha}$ and $\frac{1}{2} + c$ converges to $\frac{1}{2}$ as $n \to \infty$.

By definition we have

$$P(\bar{p}|M_{SH}) = \int_{\frac{1}{2}+c}^{1} P(\bar{p}|p = x) \frac{1}{\frac{1}{2} - c} dx.$$

Now consider what happens when $E_{-}$ occurs. In that case the following inequality holds

$$P(\bar{p}|M_{SH}) \leq P\left(\bar{p}|p = \frac{1}{2} + c\right).$$

The Central Limit Theorem yields

$$P\left(\bar{p}\Big|p = \frac{1}{2} + c\right) \sim \frac{1}{\sigma_n\sqrt{2\pi}}\exp\left(-\frac{(\bar{p} - (\frac{1}{2} + c))^2}{2\sigma_n^2}\right),$$

where $\sigma_n^2 = \frac{p(1-p)}{n} = (\frac{1}{4} - c^2)/n$.

Conditional on $E_-$, the quantity $\bar{p} - (\frac{1}{2} + c)$ lies between $-2c$ and $-c - n^{-0.5-\alpha}$. Hence, we have the following asymptotic inequality

$$P(\bar{p}|M_{SH}) \le B_1\sqrt{n}\exp\left(-An(c + n^{-0.5-\alpha})^2\right), \qquad (1)$$

with $A = \frac{1}{2(\frac{1}{4}-c^2)}$ and $B_1 = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\frac{1}{4}-c^2}}$.

Let us now turn to model $M_{ST}$. By definition we have

$$P(\bar{p}|M_{ST}) = \int_0^{\frac{1}{2}-c} P(\bar{p}|p = x)\frac{1}{\frac{1}{2}-c}dx.$$

Again, consider what happens when $E_-$ occurs. In that case the following inequality holds (when considering an interval of size $\frac{1}{n^2}$ to the left of $\frac{1}{2} - c$)

$$P(\bar{p}|M_{ST}) \ge \frac{1}{(\frac{1}{2}-c)}\frac{1}{n^2}P\left(\bar{p}\Big|p = \frac{1}{2} - c - \frac{1}{n^2}\right).$$

The Central Limit Theorem tells us that

$$P\left(\bar{p}\Big|p = \frac{1}{2} - c - \frac{1}{n^2}\right) \sim \frac{1}{\sigma_n\sqrt{2\pi}}\exp\left(-\frac{(\bar{p} - (\frac{1}{2} - c - \frac{1}{n^2}))^2}{2\sigma_n^2}\right).$$

Conditional on $E_-$, the quantity $\bar{p} - (\frac{1}{2} - c - \frac{1}{n^2})$ lies between $\frac{1}{n^2}$ and $c - n^{-0.5-\alpha} + \frac{1}{n^2}$. Hence, we have the following asymptotic inequality

$$P(\bar{p}|M_{ST}) \ge B_2\sqrt{n}\exp\left(-An(c - n^{-0.5-\alpha} + n^{-2})^2\right)\frac{1}{n^2}, \qquad (2)$$

with $B_2 = \frac{1}{\frac{1}{2}-c}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{\frac{1}{4}-c^2}}$.

Then, conditional on $E_-$, we get the following inequality by combining (1) and (2)

$$\frac{P(\bar{p}|M_{SH})}{P(\bar{p}|M_{ST})} \le \frac{B_1\sqrt{n}\exp\left(-An(c + n^{-0.5-\alpha})^2\right)}{B_2\sqrt{n}\exp\left(-An(c - n^{-0.5-\alpha} + n^{-2})^2\right)\frac{1}{n^2}}.$$

The right hand side reduces to

$$\frac{B_1}{B_2}n^2\exp\left(-A\left(4cn^{0.5-\alpha} + 2cn^{-1} - 2n^{-1.5-\alpha} + n^{-3}\right)\right),$$

which converges to 0 as $n \to \infty$. As a result we have

$$\frac{P(\bar{p}|M_{SH})}{P(\bar{p}|M_{ST})} \to 0$$

as $n \to \infty$. In other words, conditional on $E_-$, the likelihood ratio of $M_{SH}$ to $M_{ST}$ converges to 0. By a similar argument, conditional on $E_+$, the likelihood ratio of $M_{ST}$ to $M_{SH}$ converges to 0.

By applying Bayes's theorem we get the following expression for the posterior probability of $M_{ST}$:

$$P(M_{ST}|\bar{p}) = \frac{1}{(1+R)},$$

where

$$R = \frac{\frac{1}{2}P(\bar{p}|M_{SH})}{\frac{1}{2}P(\bar{p}|M_{ST})}.$$

And so if $P(\bar{p}|M_{SH})/P(\bar{p}|M_{ST})$ converges to 0 as $n \to \infty$, then so does $P(M_{SH}|\bar{p})$. In summary, as $n \to \infty$, $P(M_{ST}|\bar{p})$ converges to 0 with probability 0.5, and converges to 1 with probability 0.5. $\square$

# References

Achinstein, P. (2001). *The book of evidence*. Oxford: Oxford University Press.

Belot, G. (2013). Bayesian orgulity. *Philosophy of Science*, *80*, 483–503.

Berk, R. H. (1966). Limiting behaviour of posterior distributions when the model is incorrect. *Annals of Mathematical Statistics*, *37*, 51–58.

Carnap, R. (1962). *The logical foundations of probability*. Chicago: Chicago University Press.

Dawid, A. P. (1999). *The trouble with Bayes factors*. Research Report 202. Department of Statistical Science. University College London.

Earman, J. (1992). *Bayes or bust: A critical examination of Bayesian confirmation theory*. Cambridge, MA: MIT Press.

Eells, E. (1982). *Rational decision and causality*. Cambridge: Cambridge University Press.

Freedman, D. (1963). On the asymptotic behavior of Bayes' estimates in the discrete case. *Annals of Mathematical Statistics*, *34*, 1386–1403.

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

Glymour, C. (1980). *Theory and evidence*. Princeton: Princeton University Press.

Hawthorne, J. (2011). Confirmation theory. In P. S. Bandyopadhyay, & M. R. Forster (Eds.) *Philosophy of statistics: Handbook of the philosophy of science* (vol. 7, pp. 333–389).

Hesse, M. (1974). *The structure of scientific inference*. Berkeley: University of California Press.

Jeffrey, R. (1992). *Probability and the art of judgement*. Cambridge: Cambridge University Press.

Kelly, K. T. (1996). *The logic of reliable inquiry*. Oxford: Oxford University Press.

Lewis, P. O., Holder, M. T., & Holsinger, K. E. (2005). Polytomies and Bayesian phylogenetic inference. *Systematic Biology*, *54*, 241–253.

Morey, R. D., Romeijn, J.-W., & Rouder, J. N. (2013). The humble Bayesian: Model checking from a fully Bayesian perspective. *British Journal of Mathematical and Statistical Psychology*, *66*, 68–75.

Parker, W. (2009). Confirmation and adequacy-for-purpose in climate modelling. *Proceedings of the Aristotelian Society Supplementary*, *83*, 233–249.

Yang, Z. (2007). Fair-balance paradox, star-tree paradox, and Bayesian phylogenetics. *Molecular Biology and Evolution*, *24*, 1639–1655.