Bayes factors for research workers

Ly, A.

Chapter 6

A Limit-Consistent Bayes Factor for Testing the Equality of Two Poisson Rates

Abstract

To facilitate the selection of prior distributions with good properties we introduce the desideratum of *limit-consistency*. This desideratum is relevant for tests of equality between two processes, and it concerns the hypothetical scenario where data acquisition for one process is terminated early whereas data acquisition of the second process continues indefinitely. In such cases, the Bayes factor ought to approach a finite limit. We rederive Jeffreys's 1939 Bayes factor for the comparison between two Poisson rates and prove that it is not limit-consistent: as sample size for the uninterrupted process increases, support in favour of the null hypothesis eventually grows without bound. We generalise Jeffreys's approach by centring the alternative hypothesis around the value specified by the null hypothesis. We prove that the generalised version of Jeffreys's test is limit-consistent.

Keywords: Bayes factor, hypergeometric functions, statistical evidence, two-sample test.

6.1 Introduction

A homogenous Poisson process $Y_i(t_i)$ has rate λ_i if, for given $\lambda_i > 0$, the chance of observing $Y_i(t_i) = y_i$ after t_i units of time is

$$f(y_i \mid \lambda_i, t_i) = \frac{(\lambda_i t_i)^{y_i}}{y_i!} e^{-\lambda_i t_i}, \tag{6.1.1}$$

This chapter is under preparation as Ly, A., Raj, A., Marsman, M. & Wagenmakers, E.–J. (2017). A limit-consistent Bayes factor for testing the equality of two Poisson rates.

where y_i is any non-negative integer. A famous use of the Poisson distribution is in the detection of radioactivity using a Geiger counter (Rutherford et al., 1910; Stirzaker, 2000).

Here we study a two-sample problem and focus on evaluating the hypothesis that two homogenous Poisson processes have equal rate with exposure times t_1 and t_2 not necessarily the same.

Throughout the text t_1 and t_2 represent time, but they could as well relate to measurement from different areas (e.g., Haight, 1967) or in the case of radioactivity refer to the different number of atoms in two specimens of rock. Indeed, measurements of Poisson processes with $t_1 = t_2$ are rare and, more often than not, $t_1 \neq t_2$.

The frequentist test for the equality of two Poisson rates has received considerable attention (e.g., Haight, 1967; Krishnamoorthy and Thomson, 2004; Przyborowski and Wilenski, 1940 for the $t_1 = t_2$ case, and Ng et al., 2007; Ractliffe, 1964; Shiue and Bain, 1982 for the $t_1 \neq t_2$ case). Here we focus on the Bayesian hypothesis test known as the Bayes factor.

The purpose of this paper is three-fold: Firstly, we introduce the desideratum of limit-consistency, relevant for the behaviour of any test that involves a comparison between two or more processes. Secondly, we rederive the Bayes factor proposed by Jeffreys for the two-sample Poisson problem (Jeffreys, 1939, pp. 211-212)[1] and prove that it violates limit-consistency. Thirdly, we propose a generalisation of Jeffreys's test that is limit-consistent.

6.1.1 Desiderata that facilitate the selection of prior distributions

Let $\mathcal{M}_1$ denote the model in which the rates λ_1 and λ_2 of the two Poisson processes are free to vary, and $\mathcal{M}_0$ the restriction of $\mathcal{M}_1$ such that $\lambda_1 = \lambda_2$. In the Bayesian setting the models are assigned prior model probabilities $0 < P(\mathcal{M}_0), P(\mathcal{M}_1) < 1$, which in light of the observed data d can be updated to posterior model probabilities using Bayes' theorem. Doing this for both models and subsequently taking the ratio of the result leads to the key expression

$$\underbrace{\frac{P(\mathcal{M}_1 \,|\, d)}{P(\mathcal{M}_0 \,|\, d)}}_{\text{Posterior odds}} = \underbrace{\frac{p(d \,|\, \mathcal{M}_1)}{p(d \,|\, \mathcal{M}_0)}}_{\text{BF}_{10}(d)} \underbrace{\frac{P(\mathcal{M}_1)}{P(\mathcal{M}_0)}}_{\text{Prior odds}} . \tag{6.1.2}$$

The term $\text{BF}_{10}(d)$ is known as the Bayes factor and equals the change from prior to posterior model odds brought about by the observed data d (Etz and Wagenmakers, 2017; Jeffreys, 1935; Kass and Raftery, 1995; Ly et al., 2016a).

Note that the Bayes factor $\text{BF}_{10}(d)$ does not depend on the prior model probabilities $P(\mathcal{M}_1)$ and $P(\mathcal{M}_0)$. However, the Bayes factor is the ratio of marginal likelihoods:

$$p(d \,|\, \mathcal{M}_i) = \int f(d \,|\, \theta_i, \mathcal{M}_i) \pi_i(\theta_i) \mathrm{d}\theta_i, \tag{6.1.3}$$

[1] The widely available third edition contains the same text (i.e., Jeffreys, 1961, pp. 267–268).

which shows that the Bayes factor does depend on the priors $\pi_1(\theta_1)$ *and* $\pi_2(\theta_2)$ that are assigned to the parameters within the two models.

For this reason, the selection of prior distributions demands careful consideration. Fortunately, general principles constrain the selection of prior distributions. For instance, priors of arbitrary width yield Bayes factors that favour the null model irrespective of the observed data (e.g., the Jeffreys-Lindley-Bartlett paradox Bartlett, 1957; Jeffreys, 1961; Lindley, 1957). Furthermore, the prior on the test-relevant parameter must be proper, as improper priors contain suppressed normalisation constants and may lead to Bayes factors of arbitrary value. Consequently, in Bayes factor hypothesis testing we cannot use popular "non-informative" priors selected by formal rules (Kass and Wasserman, 1996) such as the right-Haar prior (e.g., Berger et al., 1998; Ghosh, 2011) and Jeffreys's parameterisation-invariant prior (e.g., Jeffreys, 1946; Ly et al., 2017c) which are both improper. We also require that a reasonable Bayes factor does not depend on the units of measurement that the researcher chooses to represent the data. Naturally, we also desire that the Bayes factors are calculable in the sense that both integrals in the numerator and denominator are solvable for any data set d.

Other desiderata unfortunately hold only for continuous random variables. For instance, the desideratum of *predictive matching* states that the Bayes factor ought to be perfectly indifferent, i.e., $\text{BF}_{10}(d) = 1$, in case the data are completely uninformative; the desideratum of *information consistency* states that the Bayes factor ought to provide infinite support for the alternative hypothesis in case the data are overwhelmingly informative (for a review see Bayarri et al., 2012; see also Bayarri and Berger, 2013). In case of discrete data it is not clear what constitutes completely uninformative and overwhelmingly informative.

Here we propose a new and relatively general desideratum that further constrains the selection of prior distributions: *limit-consistency.* This desideratum holds regardless of whether the data are discrete or continuous, and applies whenever the test at hand features a comparison between two or more processes or groups. Consider again a comparison between two Poisson processes and assume that the measurement of the first process is terminated early, whereas the measurement of the second process continuous indefinitely. In the limit, knowledge about the second process will reach perfection, but knowledge about the interrupted process will remain incomplete. Consequently, there exists a bound on the level of evidence that can be obtained in a test that compares the two processes. As measurement for the second process continues, the Bayes factors ought to approach a finite limit.

The practical value of limit-consistency as a constraint on the selection of prior distributions will now be demonstrated through Harold Jeffreys's test for the equality of two Poisson rates.

6.2 Jeffreys's Bayes factor for the comparison of two Poisson rates

Jeffreys's derivation starts with a rewrite of the joint distribution of the two processes $Y_1(t_1)$ and $Y_2(t_2)$, i.e.,

$$f(d \,|\, \lambda_1, \lambda_2, \mathcal{M}_1) = \frac{(\lambda_1 t_1)^{y_1}}{y_1!} e^{-\lambda_1 t_1} \frac{(\lambda_2 t_2)^{y_2}}{y_2!} e^{-\lambda_2 t_2} \tag{6.2.1}$$

in terms of the relative timed rate $\theta = \frac{\lambda_1 t_1}{\lambda_1 t_1 + \lambda_2 t_2}$ and the total timed rate $\zeta = \lambda_1 t_1 + \lambda_2 t_2$, that is,

$$f(d \,|\, \theta, \zeta, \mathcal{M}_1) = \underbrace{\binom{y_.}{y_1} \theta^{y_1} (1-\theta)^{y_2}}_{f(y_1 \,|\, y_., \theta)} \underbrace{\frac{\zeta^{y_.}}{y_.!} e^{-\zeta}}_{f(y_. \,|\, \zeta)} \tag{6.2.2}$$

where $y_. = y_1 + y_2$ denotes the total number of observations across the two processes. Setting $\lambda_2 = \lambda_1$ shows that $\mathcal{M}_0$ can be perceived as a restriction of $\mathcal{M}_1$ with the relative timed rate θ known and fixed at $\theta_0 = \frac{t_1}{t_.}$, where $t_. = t_1 + t_2$, thus,

$$f(d \,|\, \zeta, \mathcal{M}_0) = f(y_1 \,|\, y_., \tfrac{t_1}{t_.}) f(y_. \,|\, \zeta). \tag{6.2.3}$$

The factorisation of the two likelihood functions allowed Jeffreys to conceptualise the two-sample Poisson problem as a conditional binomial test, if the same prior on the common parameter ζ is chosen.[2] This can be achieved by assigning independent gamma priors $\lambda_i \sim \text{Gam}(\alpha_i, \beta t_i)$ in $\mathcal{M}_1$ and $\lambda_1 \sim \text{Gam}(\alpha_., \beta t_.)$ in $\mathcal{M}_0$, where $\alpha_. = \alpha_1 + \alpha_2$; the induced prior on the total timed rate is then $\zeta \sim \text{Gam}(\alpha_., \beta)$ under both models.

Under the alternative hypothesis, the test relevant parameter θ receives a beta prior, that is,

$$\pi_\eta(\theta, \zeta \,|\, \mathcal{M}_1) = \underbrace{\tfrac{1}{\mathcal{B}(\alpha_1, \alpha_2)} \theta^{\alpha_1 - 1} (1-\theta)^{\alpha_2 - 1}}_{\text{Beta}(\theta \,;\, \alpha_1, \alpha_2)} \underbrace{\tfrac{\beta^{\alpha_.}}{\Gamma(\alpha_.)} \zeta^{\alpha_. - 1} e^{-\beta \zeta}}_{\text{Gam}(\zeta \,;\, \alpha_., \beta)} \tag{6.2.4}$$

where $\eta = (\alpha_1, \alpha_2, \beta)$ denotes the hyperparameters and $\mathcal{B}$ denotes the beta function. Hence, the prior factorises and, consequently, so do the marginal likelihoods:

$$p_\eta(d \,|\, \mathcal{M}_1) = p_{\alpha_1, \alpha_2}(y_1 \,|\, y_.) p_{\alpha_., \beta}(y_.), \tag{6.2.5}$$

$$p_\eta(d \,|\, \mathcal{M}_0) = f(y_1 \,|\, y_., \tfrac{t_1}{t_.}) p_{\alpha_., \beta}(y_.), \tag{6.2.6}$$

where $f(y_1 \,|\, y_., \frac{t_1}{t_.})$ is given in Eq. (6.2.2), and where

$$p_{\alpha_1, \alpha_2}(y_1 \,|\, y_.) = \binom{y_.}{y_1} \frac{\mathcal{B}(y_1 + \alpha_1, y_2 + \alpha_2)}{\mathcal{B}(\alpha_1, \alpha_2)}, \tag{6.2.7}$$

$$p_{\alpha_., \beta}(y_.) = \frac{\Gamma(\alpha_. + y_.)}{\Gamma(\alpha_.) y_.!} \Big(\frac{\beta}{1+\beta} \Big)^{\alpha_.} \Big(\frac{1}{1+\beta} \Big)^{y_.}. \tag{6.2.8}$$

[2] With $t_1 = t_2$ we have $\theta_0 = \frac{1}{2}$ and note the resemblance to the frequentist conditional binomial test proposed by Przyborowski and Wilenski (1940), which was published one year after the first edition of Jeffreys's book.

Dividing Eq. (6.2.5) by Eq. (6.2.6) shows that

$$\mathrm{BF}^{J}_{10\,;\,\alpha_1,\alpha_2}(d) = \frac{\mathcal{B}(y_1+\alpha_1, y_2+\alpha_2)}{\mathcal{B}(\alpha_1,\alpha_2)(\frac{t_1}{t_.})^{y_1}(\frac{t_2}{t_.})^{y_2}}, \tag{6.2.9}$$

where $d = (y_1, t_1, y_2, t_2)$. Jeffreys proposed to set $\alpha_1 = \alpha_2 = a$, i.e, $\mathrm{BF}^{J}_{10\,;\,a}(d) = \mathrm{BF}^{J}_{10\,;\,a,a}(d)$ with $a = 1$, that is,

$$\mathrm{BF}^{J}_{10\,;\,1}(d) = \frac{\mathcal{B}(y_1+1, y_2+1)}{(\frac{t_1}{t_.})^{y_1}(\frac{t_2}{t_.})^{y_2}}, \tag{6.2.10}$$

to compare the model $\mathcal{M}_1$ with differing Poisson rates against $\mathcal{M}_0$ in which the two rates are the same.

6.2.1 Properties of Jeffreys's Bayes factor $\mathrm{BF}^{J}_{10\,;\,a}(d)$

6.2.1.1 Invariances

Observe that by setting $\lambda_i \sim \mathrm{Gam}(\alpha_i, \beta t_i)$ and by specifying the test relevant parameter to be the unitless quantity θ, we have effectively assigned a beta prior $\mathrm{Beta}(\alpha_1, \alpha_2)$ to θ. Within this framework, the setting $\alpha_1 = \alpha_2 = a$ leads to a Bayes factor $\mathrm{BF}^{J}_{10\,;\,a}(d)$ that does not depend on how the processes are labeled, as the same output is obtained for $\tilde{d} = (y_2, t_2, y_1, t_1)$ and for d.

Furthermore, the measurement scale for the times t_1 and t_2 does not affect the outcome, as the Bayes factor $\mathrm{BF}^{J}_{10\,;\,a}(d)$ only depends on the ratio $\frac{t_1}{t_.}$.

6.2.1.2 Uninformativeness for balanced outcomes

Jeffreys's choice for $a = 1$ was inspired by the Bayes factor he developed for the binomial problem $X \sim \mathrm{Bin}(\theta, n)$. Jeffreys (1961, p. 257) noted that when testing $\mathcal{H}_0 : \theta = 1/2$ against $\mathcal{H}_1 : \theta \in (0,1)$ the Bayes factor $\mathrm{BF}^{J}_{10\,;\,a}(x, n) = 1$ after a single observation $n = 1$ independently of whether we observe a success $x = 1$ or a failure $x = 0$. Jeffreys mentions that this behaviour also extends to the case when the number of observed successes x equals the number of failures $n - x$, (i.e., $n = 2x$), as an additional observation will once again not change the Bayes factor, meaning $\mathrm{BF}^{J}_{10\,;\,a}(x, 2x) = \mathrm{BF}^{J}_{10\,;\,a}(x, 2x+1) = \mathrm{BF}^{J}_{10\,;\,a}(x+1, 2x+1)$. Note that this property holds for any other $a > 0$, but only if $\mathcal{H}_0 : \theta = 1/2$. When applied to the Poisson case, this means that when the two exposure times are the same (i.e., $t_1 = t_2$) and the observed counts are the same (i.e., $y_1 = y_2$), Jeffreys's Bayes factor does not change when a single additional count is added to one of the processes.

6.2.1.3 Limit-inconsistency

Unfortunately, Jeffreys's Bayes factor $\mathrm{BF}^{J}_{10\,;\,a}(d)$ with $a > 0$ is limit-inconsistent, that is, $\mathrm{BF}^{J}_{10\,;\,a}(d)$ does not stabilise once data collection of the first process is interrupted at t_1 and data acquisition of the second process continues indefinitely. The following property shows that Jeffreys's Bayes factor $\mathrm{BF}^{J}_{10\,;\,a}(d)$ will eventually favour the simpler model $\mathcal{M}_0$, regardless of the data.

Property 6.2.1 ($\mathrm{BF}^{J}_{10\,;\,a}(d)$ is limit-inconsistent). *Let $y_1, t_1, \lambda_2 > 0$ be fixed. Then for every $a > 0$ (i.e., a symmetric beta distribution on θ), Jeffreys's Bayes factor $\mathrm{BF}^{J}_{10\,;\,a}(d)$ tends to zero as t_2 grows.* ◇

Proof. We proof by contradiction and assume that for all t_2 the logarithm of the Bayes factor $\mathrm{BF}^{J}_{10\,;\,a}(d)$ is bounded from below, thus, the existence of constant M such that

$$\log \mathrm{BF}^{J}_{10\,;\,a}(d) = \log p_a(y_1 \mid y_.) - \log f(y_1 \mid y_., \tfrac{t_1}{t_.}) \geq M \tag{6.2.11}$$

for all t_2. To simplify matters we use $y_2 = E(Y_2(t_2)) = \lambda_2 t_2$. The asymptotic behaviour of $-\log f(y_1 \mid y_., \frac{t_1}{t_1+t_2})$ for t_2 large can be described by the following two series

$$-y_1 \log(\tfrac{t_1}{t_1+t_2}) = y_1[\log(\tfrac{t_2}{t_1}) + \tfrac{t_1}{t_2}] + \mathcal{O}(\tfrac{1}{t_2^2}), \tag{6.2.12}$$

$$-\lambda_2 t_2 \log(\tfrac{t_2}{t_1+t_2}) = \lambda_2 t_1[1 - \tfrac{t_1}{2t_2}] + \mathcal{O}(\tfrac{1}{t_2^2}). \tag{6.2.13}$$

In addition, Stirling's approximation of the beta function implies that the logarithm of the Bayes factor behaves as

$$\log \mathrm{BF}^{J}_{10\,;\,a}(d) \sim \log \Gamma(y_1 + a) - (y_1 + a) \log(\lambda_2 t_2 + a) + y_1 \log\left(\frac{t_2}{t_1}\right) + \lambda_2 t_1, \tag{6.2.14}$$

when t_2 is large. The assumption that $\log \mathrm{BF}^{J}_{10\,;\,a}(d)$ is bounded from below leads to a contradiction as a rewrite now shows that

$$\log \Gamma(y_1 + a) + \lambda_2 t_1 - M \geq y_1 \log(\lambda_2 t_1 + \tfrac{a t_1}{t_2}) + a \log(\lambda_2 t_2 + a), \tag{6.2.15}$$

from which we can incorrectly conclude that the logarithm is a bounded function. Thus, $\mathrm{BF}^{J}_{10\,;\,a}(d) \to 0$ as $t_2 \to \infty$. □

Two concrete examples are given in Fig. 6.1. In both panels, the dashed line represents the logarithm of $\mathrm{BF}^{J}_{10\,;\,a}(d)$ with $a = 1$. In the left panel $y_1 = t_1 = 1$ and $y_2 = \lambda_2 t_2$ with $\lambda_2 = 5$. The dashed line decreases, meaning that Jeffreys's Bayes factor eventually indicates evidence for the null hypothesis $\lambda_1 = \lambda_2$ as $\lambda_2 = 5$ is estimated more precisely but the estimate of λ_1 remains highly uncertain.

In the right panel $y_1 = t_1 = 1$ and $\lambda_2 = 1$. For small values of t_2, Jeffreys's Bayes factor $\mathrm{BF}^{J}_{10\,;\,a}(d)$ with $a = 1$ now conveys evidence in favour of the null, which is what is expected in this situation. However, the dashed line again decreases, showing that the evidence for the null grows without bound even though the information about the first process is limited and uncertain. Choosing another $a > 0$ does not solve the problem as the bias for the null model is driven by the term $a \log(\lambda_2 t_2 + a)$ in the asymptotic expansion in Eq. (6.2.15).

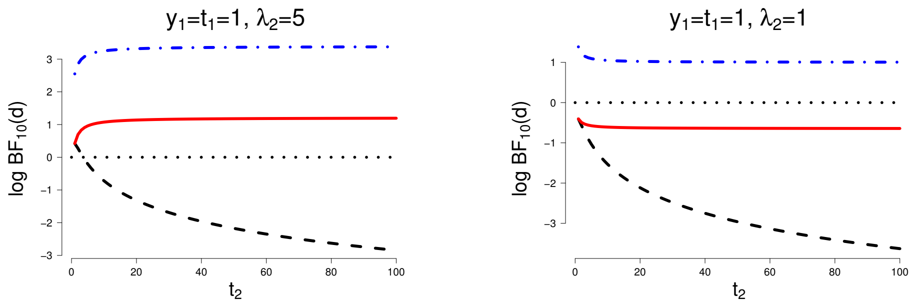

Figure 6.1: Jeffreys's Bayes factor $\mathrm{BF}^{J}_{10\,;\,1}(d)$ (dashed lines in both panels) is limit-inconsistent and increasingly favours the null hypothesis as the exposure time of the second process lengthens. This bias can be eliminated by setting $a = 0$ in the prior distribution, but the resulting Bayes factor $\mathrm{BF}^{J}_{10\,;\,0}(d)$ (dot-dashed lines) unduly favours the alternative model. The localised Bayes factor $\mathrm{BF}_{10\,;\,a}(d)$ (solid line) is limit-consistent. Left panel: The log of the Bayes factor based on $y_1 = t_1 = 1$ and $\lambda_2 = 5$, which should yield some evidence in favour of the alternative hypothesis as t_2 grows. Right panel: The log of the Bayes factors based on $y_1 = t_1 = 1$ and $\lambda_2 = 1$, which should yield some evidence in favour of the null hypothesis as t_2 grows.

6.3 A limit-consistent Bayes factor for the comparison of two Poisson rates

One way to obtain a limit-consistent Bayes factor is by setting $a = 0$, as the bias term in the asymptotic expansion then cancels. The dot-dashed line in Fig. 6.1 confirms that $\mathrm{BF}^{J}_{10\,;\,0}(d)$ stabilises as t_2 increases. The problem with $a = 0$, however, is that we then effectively use an improper prior with an unspecified normalisation constant on the test relevant parameter θ and this choice introduces new problems. Fig. 6.1 shows the undesirable consequence: in both panels, $\mathrm{BF}^{J}_{10\,;\,0}(d)$ overvalues the support in favour of the alternative hypothesis; this is particularly poignant for the example shown in the right panel, where $y_1 = t_1 = 1$ and $\lambda_2 = 1$, which ought to result in evidence for the null hypothesis. Moreover, $a = 0$ yields infinite support for the alternative when $y_1 = 0$, $y_2 = 1$ and $t_1 = t_2$.

In order to obtain a Bayes factor that is limit-consistent we now consider the localised beta distribution that expands the beta distribution with an additional parameter θ_0 that allows the model $\mathcal{M}_1$ to be centred on the simpler model $\mathcal{M}_0$. This centring occurs on the logit scale. Recall that the standard beta distribution $\mathrm{Beta}(\alpha_1, \alpha_2)$ reparameterised as $\phi = \log(\frac{\theta}{1-\theta})$ is given by[3]

$$\int_{-\infty}^{\infty} \pi(\phi\,;\, \alpha_1, \alpha_2) \mathrm{d}\phi = \frac{1}{\mathcal{B}(\alpha_1, \alpha_2)} \int_{-\infty}^{\infty} \frac{e^{\phi \alpha_1}}{(1 + e^{\phi})^{\alpha_1 + \alpha_2}} \mathrm{d}\phi. \tag{6.3.1}$$

[3]Thus, $\int \mathrm{d}\phi = \int \theta^{-1}(1-\theta)^{-1}\mathrm{d}\theta$ and equivalently, $\theta = \frac{e^{\phi}}{1+e^{\phi}}$ and therefore, $\int \mathrm{d}\theta = \int \frac{e^{\phi}}{(1+e^{\phi})^2}\mathrm{d}\phi$.

As ϕ ranges over the real line, its location can be shifted by ϕ_0 resulting in

$$\pi_{\alpha_1,\alpha_2}(\phi\,;\,\phi_0) = \frac{1}{\mathcal{B}(\alpha_1,\alpha_2)} \frac{e^{\alpha_1(\phi-\phi_0)}}{(1+e^{\phi-\phi_0})^{\alpha_1+\alpha_2}}. \tag{6.3.2}$$

Back transforming this distribution with $\phi_0 = \log(\frac{\theta_0}{1-\theta_0})$ yields the localised beta distribution

$$\pi_{\alpha_1,\alpha_2}(\theta\,;\,\theta_0) = \underbrace{\frac{1}{\mathcal{B}(\alpha_1,\alpha_2)}\theta^{\alpha_1-1}(1-\theta)^{\alpha_2-1}}_{\text{Beta}(\theta\,;\,\alpha_1,\alpha_2)}(\tfrac{1-\theta_0}{\theta_0})^{\alpha_1}(1-[2-\tfrac{1}{\theta_0}]\theta)^{-(\alpha_1+\alpha_2)}, \tag{6.3.3}$$

where $\text{Beta}(\theta\,;\,\alpha_1,\alpha_2)$ refers to the beta density. Note that with $\theta_0 = 1/2$ (i.e., $t_1 = t_2$), $\alpha_1 = \alpha_2 = a$ and $a = 1$, we retrieve Jeffreys's choice on θ in the two-sample Poisson problem. Hence, a Bayes factor constructed from this prior retains the desirable behaviour of $\text{BF}^J_{10\,;\,a}(d)$ for $a > 0$ at $t_1 = t_2$. To derive this localised Bayes factor we have to calculate the marginal likelihood with this new prior.

Property 6.3.1 (The marginal likelihood of a binomially distributed random variable with a localised beta prior)**.** *Let $f(y_1 \,|\, y_., \theta)$ be the binomial pmf and θ distributed according to a localised beta prior. The marginal likelihood is then*

$$p_{\alpha_1,\alpha_2,\theta_0}(y_1 \,|\, y_.) = \binom{y_.}{y_1}\frac{\mathcal{B}(\alpha_1+y_1,\alpha_2+y_2)}{\mathcal{B}(\alpha_1,\alpha_2)} \times {}_2F_1(\alpha_1+\alpha_2,\alpha_1+y_1\,;\,\alpha_1+\alpha_2+y_.\,;\,2-\tfrac{1}{\theta_0})(\tfrac{1-\theta_0}{\theta_0})^{\alpha_1} \tag{6.3.4}$$

where ${}_2F_1(u,v\,;\,w\,;\,z)$ denotes Gauss' hypergeometric function. ◇

Proof. With $C = \binom{y_.}{y_1}(\frac{1-\theta_0}{\theta_0})^{\alpha_1}/\mathcal{B}(\alpha_1,\alpha_2)$, $u_1 = y + \alpha_1$, $u_2 = y_2 + \alpha_2$, $v = \alpha_1 + \alpha_2$ and by definition of the prior predictive, we have

$$p_{\alpha_1,\alpha_2,\theta_0}(y_1 \,|\, y_.) = C\int_0^1 \theta^{u_1-1}(1-\theta)^{u_2-1}(1-\theta[2-\tfrac{1}{\theta_0}])^{-v}\text{d}\theta, \tag{6.3.5}$$

$$= C\mathcal{B}(u_1,u_2){}_2F_1(v,u_1\,;\,u_1+u_2\,;\,2-\tfrac{1}{\theta_0}). \tag{6.3.6}$$

The last equality follows from Euler's integral representation of the hypergeometric function (Abramowitz and Stegun, 1964, p. 558). □

Hence, with the gamma prior $\zeta \sim \text{Gam}(\alpha_., \beta)$ on the total timed rate as before, and a beta prior localised at $\theta_0 = (\frac{t_1}{t_.})$ on the relative timed rate θ, we have the following Bayes factor for the two-sample Poisson problem:

$$\text{BF}_{10\,;\,\alpha_1,\alpha_2}(d) = \frac{\mathcal{B}(\alpha_1+y_1,\alpha_2+y_2)}{\mathcal{B}(\alpha_1,\alpha_2)(\frac{t_1}{t_.})^{y_1}(\frac{t_2}{t_.})^{y_2}} \times {}_2F_1(\alpha_1+\alpha_2,\alpha_1+y_1\,;\,\alpha_1+\alpha_2+y_1+y_2\,;\,\tfrac{t_1-t_2}{t_1})(\tfrac{t_2}{t_1})^{\alpha_1}. \tag{6.3.7}$$

This is essentially Jeffreys's Bayes factor $\text{BF}^J_{10\,;\,\alpha_1,\alpha_2}(d)$, but with a correction factor for the localisation at $\frac{t_1}{t_.}$.

6.3.1 Properties of the new Bayes factor $\mathrm{BF}_{10\,;\,a}(d)$

Property 6.3.2 (Invariance). *With $\alpha_1 = \alpha_2 = a$ we have*

$$\mathrm{BF}_{10\,;\,a}(d) = \mathrm{BF}^J_{10\,;\,a}(d)(\tfrac{t_2}{t_1})^a \; {}_2F_1(2a, a + y_1\,;\, 2a + y_1 + y_2\,;\, 1 - \tfrac{t_2}{t_1}) \tag{6.3.8}$$

$$= \mathrm{BF}^J_{10\,;\,a}(d)(\tfrac{t_1}{t_2})^a \; {}_2F_1(2a, a + y_2\,;\, 2a + y_1 + y_2\,;\, 1 - \tfrac{t_1}{t_2}) \tag{6.3.9}$$

which implies that this Bayes factor is invariant under relabeling and independent of the units for the times t_1 and t_2. ◇

Proof. Using Pfaff's transform (Gradshteyn and Ryzhik, 2007), we find

$${}_2F_1(2a, a + y_1\,;\, 2a + y_.\,;\, 1 - \tfrac{t_2}{t_1}) = (\tfrac{t_2}{t_1})^{-2a} {}_2F_1(2a, a + y_2\,;\, 2a + n\,;\, 1 - \tfrac{t_1}{t_2}).$$

Multiplying both sides by $\mathrm{BF}^J_{10\,;\,a}(d)(\tfrac{t_2}{t_1})^a$ yields the assertion. □

Property 6.3.3 (Limit-consistency). *Suppose that the data collection of the first process is halted at t_1 resulting in y_1 observations. Furthermore, let $y_2 = \lambda_2 t_2$ for some $\lambda_2 > 0$, then the Bayes factor $\mathrm{BF}_{10\,;\,a}(d)$ converges to a limit as t_2 grows indefinitely. Thus,*

$$g_a(y_1, t_1, \lambda_2) = \lim_{t_2 \to \infty} \mathrm{BF}_{10\,;\,a}(y_1, t_1, \lambda_2 t_2, t_2) \tag{6.3.10}$$

exists. For $a = 1$, the solution g_a can be well approximated by

$$\begin{aligned} \tilde{g}(y_1, t_1, \lambda_2) = {} & \log \Gamma(y_1 + a) + \lambda_2 t_1 - y_1 \log(\lambda_2 t_1) + a \log\left(\frac{\lambda_2}{t_1}\right) \\ & + 2a\left[\frac{y_1 - a}{y_1 + \lambda_2 t_1 - a} + \log\left(\frac{t_1}{y_1 + \lambda_2 t_1 - a}\right)\right] \\ & - (y_1 + a)\log\left(\frac{y_1 + \lambda_2 t_1 + a}{y_1 + \lambda_2 t_1 - a}\right) \\ & - \frac{1}{2}\log\left(1 - \frac{8a(a + y_1)}{3a + \lambda_2 t_1 + y_1 + |y_1 + \lambda_2 t_1 - a|}\right) \end{aligned} \tag{6.3.11}$$

which we verified numerically. ◇

Proof. To study the asymptotic behaviour of $\mathrm{BF}_{10\,;\,a}(d)$ we consider Eq. (6.3.9) as the hypergeometric function with arguments smaller than one, thus, $1 - \tfrac{t_1}{t_2}$, are easier to handle as $t_1 \ll t_2$. We first provide some intuition.

Recall that the asymptotic behaviour of $\mathrm{BF}^J_{10\,;\,a}(d)$ is given by

$$\log \mathrm{BF}^J_{10\,;\,a}(d) \sim \log \Gamma(y_1 + a) - y_1 \log(\lambda_2 t_1) - a \log(\lambda_2 t_2 + a) + \lambda_2 t_1. \tag{6.3.12}$$

Hence, to show that $\log \mathrm{BF}_{10\,;\,a}(d)$ stabilises, we have to show that the logarithms of the additional factors of Eq. (6.3.9), that is, $(\tfrac{t_1}{t_2})^a$ and ${}_2F_1(2a, a + \lambda_2 t_2\,;\, 2a + n\,;\, 1 - \tfrac{t_1}{t_2})$ behave as $a \log(t_2)$ because this cancels out the bias-driving term $a \log(\lambda_2 t_2 + a)$. To see that this is possible, we consider the asymptotic behaviour of the hypergeometric function for the argument and the parameters separately.

Suppose that the argument $z = 1 - \frac{t_1}{t_2}$ is fixed, then the parameters $v = a + \lambda_2 t_2$ and $w = 2a + y_1 + \lambda_2 t_2$ of the hypergeometric function ${}_2F_1(u, v\,;\, w\,;\, z)$ will be of the same order whenever t_2 is large (Temme, 2003). As a result, we obtain

$$ {}_2F_1(2a, a + \lambda_2 t_2\,;\, 2a + \lambda_2 t_2 + y_1\,;\, 1 - \tfrac{t_1}{t_2}) \approx (1 - z)^{-2a} = (\tfrac{t_1}{t_2})^{-2a}. \qquad (6.3.13) $$

Multiplying both sides of Eq. (6.3.13) by $\mathrm{BF}^{J}_{10\,;\,a}(d)(\frac{t_1}{t_2})^a$, taking the logarithm, and considering the asymptotic expansion with respect to t_2 shows that the bias-driving factor is adjusted to $-a \log(\lambda_2 t_1 + a\frac{t_1}{t_2})$, which suggests that $\log \mathrm{BF}_{10\,;\,a}(d)$ indeed stabilises.

Similarly, suppose that the parameters $v = a + \lambda_2 t_2$ and $w = 2a + y_1 + \lambda_2 t_2$ are fixed, then for t_2 large, the argument of the hypergeometric function ${}_2F_1(u, v\,;\, w\,;\, z)$ can be taken to be one at the expense of a small approximation error. This roughly implies that for $y_1 > a$ we have

$$ {}_2F_1(2a, a + \lambda_2 t_2\,;\, 2a + y_1 + \lambda_2 t_2\,;\, 1) = \frac{\Gamma(2a + y_1 + \lambda_2 t_2)\Gamma(y_1 - a)}{\Gamma(y_1 + \lambda_2 t_2)\Gamma(y_1 + a)}, \qquad (6.3.14) $$

whenever t_2 is large enough. Again, multiplying both sides by $\mathrm{BF}^{J}_{10\,;\,a}(d)(\frac{t_1}{t_2})^a$ and writing out the beta function in $\mathrm{BF}^{J}_{10\,;\,a}(d)$ then shows that

$$ \mathrm{BF}_{10\,;\,a}(d) \approx \mathcal{B}(\lambda_2 t_2 + a, y_1 - a)(\tfrac{t_1}{t_2})^a \qquad (6.3.15) $$

whenever t_2 is large enough. An asymptotic expansion for t_2 large as in Prop. 6.2.1 then shows that

$$ \log \mathrm{BF}_{10\,;\,a}(d) \approx \log \Gamma(y_1 - a) + (a - y_1) \log(\lambda_2 t_1 + \tfrac{a t_1}{t_2}) + \lambda_2 t_1 + \mathcal{O}(\tfrac{1}{t_2^2}), \qquad (6.3.16) $$

which suggests that $\log \mathrm{BF}_{10\,;\,a}(d)$ converges to a finite number as t_2 grows indefinitely.

For a rigorous proof of the result and the derivation of $g_a(y_1, t_1, \lambda_2)$, we used a Laplace approximation to the hypergeometric function ${}_2F_1(2a, a + \lambda_2 t_2\,;\, 2a + \lambda_2 t_2 + y_1\,;\, 1 - \frac{t_1}{t_2})$ as described by Butler and Wood (2002) at each fixed t_2. We then used `Mathematica` to derive the limit which led to a function that spanned over four pages and therefore is not presented here.

By serendipity[4] we were able to approximate the four-page equation by $\tilde{g}(y_1, t_1, \lambda_2)$ given above. Numerical experiments confirm that $\tilde{g}$ approximates the true g_a well, see Table 6.1, and that g_a is in neighborhood of $\mathrm{BF}_{10\,;\,a}(d)$ with t_2 large. The error that stands out occurs with $y_1 = 2$ and $t_1 = 5$, which leads to the correct limit $g_1(2, 5, 1) = 0.047$ and the approximated limit $\tilde{g}(2, 5, 1) = 0.019$. □

The logarithm of the Bayes factor $\mathrm{BF}_{10\,;\,a}(d)$ as a function of t_2 is depicted as the solid line in Fig. 6.1. In the left panel, an exact calculation shows that

[4]The replacement of $4ax(c - b)$ by $4a(c - a)$ in the definition of $\hat{y}$ in Butler and Wood (2002, p. 1164).

Table 6.1: With $\lambda_2 = 1$ and relative error $\frac{g_1(y_1,t_1,\lambda_2)-\tilde{g}(y_1,t_1,\lambda_2)}{g_0(y_1,t_1,\lambda_2)}$ in percentage.

	$t_1 = 2$	$t_1 = 5$	$t_1 = 10$	$t_1 = 25$	$t_1 = 100$	$t_1 = 250$
$y_1 = 2$	-3.1	58.3	0.4	3e-2	4e-4	3e-5
$y_1 = 5$	-5.6	-0.1	4.4	2e-2	4e-4	3e-5
$y_1 = 10$	-5e-2	-0.8	-1e-2	2e-2	4e-4	3e-5
$y_1 = 25$	-1e-3	-3e-3	-7e-3	-7e-4	3e-4	2e-5
$y_1 = 100$	-3e-6	-8e-6	-2e-5	-5e-5	-8e-6	2e-5
$y_1 = 250$	-5e-8	-2e-7	-4e-7	-1e-6	-5e-6	-5e-7

Table 6.2: With $\lambda_2 = 5$ and relative error $\frac{g_1(y_1,t_1,\lambda_2)-\tilde{g}(y_1,t_1,\lambda_2)}{g_0(y_1,t_1,\lambda_2)}$ in percentage.

	$t_1 = 2$	$t_1 = 5$	$t_1 = 10$	$t_1 = 25$	$t_1 = 100$	$t_1 = 250$
$y_1 = 2$	-0.4	3e-2	3e-3	2e-4	2e-4	3e-6
$y_1 = 5$	4.4	2e-2	3e-3	2e-4	3e-6	2e-7
$y_1 = 10$	-1e-2	2e-2	3e-3	2e-4	3e-6	2e-7
$y_1 = 25$	-7e-3	-7e-4	3e-3	2e-4	3e-6	2e-7
$y_1 = 100$	-2e-5	-5e-5	-1e-4	-5e-3	2e-6	2e-7
$y_1 = 250$	-4e-7	-1e-6	-2e-6	-7e-7	2e-6	2e-7

Table 6.3: With $\lambda_2 = 0.001$ and relative error $\frac{g_1(y_1,t_1,\lambda_2)-\tilde{g}(y_1,t_1,\lambda_2)}{g_0(y_1,t_1,\lambda_2)}$ in percentage.

	$t_1 = 2$	$t_1 = 5$	$t_1 = 10$	$t_1 = 25$	$t_1 = 100$	$t_1 = 250$
$y_1 = 2$	-4e-2	-0.12	-0.27	-0.7	-3.2	-7.6
$y_1 = 5$	-3e-4	-9e-4	-2e-3	-6e-3	-3-2	0.1
$y_1 = 10$	-1e-5	-4e-5	-9e-5	-3e-4	-1e-3	-4e-3
$y_1 = 25$	-3e-7	-7e-7	-2e-6	-5e-6	-2e-5	-7e-5
$y_1 = 100$	-8e-10	-2e-9	-5e-9	-1e-8	-7e-8	-2e-7
$y_1 = 250$	-2e-11	-5e-11	-1e-10	-3e-10	-1e-9	-2e-9

$g_1(1,1,5) = 1.21$, whereas the approximation yields $\tilde{g}(1,1,5) = 1.15$; this means that when the first process yields $y_1 = t_1 = 1$ and stops, the evidence in favour of the alternative hypothesis is then bounded by $\text{BF}_{10\,;\,1}(d) \leq e^{1.21} \approx 3.35$. For the case depicted in the right panel, an exact calculation shows that $g_1(1,1,1) = -0.63$, whereas the approximation yields $\tilde{g}(1,1,1) = -0.90$; this means that when the first process yields $y_1 = t_1 = 1$ and stops, the evidence in favour of the null hypothesis is then bounded by $\text{BF}_{01\,;\,1}(d) \leq e^{0.63} \approx 2.46$.

6.4 Discussion

We proposed the new desideratum of *limit-consistency* that can help guide the specification of prior distributions for tests that involve a comparison of two or more processes or groups. As a concrete illustration of the added value of the desideratum we rederived Jeffreys's (1939) Bayes factor $\mathrm{BF}^{J}_{10\,;\,a}(d)$ for the two-sample Poisson problem. We proved that this Bayes factor is not limit-consistent: when t_1 is fixed and t_2 grows indefinitely, the Bayes factor increasingly supports the null, regardless of the data. This implies that researchers who use Jeffreys's Bayes factor for the comparison of two Poisson rates can bias the evidence in favour of the null by selectively investing resources in data collection for one of the two processes. We then proposed a generalisation of Jeffreys's test that eliminates the bias-driving term; consequently, this localised Bayes factor is limit-consistent while retaining the positive features of Jeffreys's original test.

The proof of limit-consistency can perhaps be sharpened – a four-page long definition of the limit function $g_a(y_1, t_1, \lambda_2)$ based on `Mathematica` output is not intuitive, and the serendipitous approximation $\tilde{g}(y_1, t_1, \lambda_2)$ may warrant more research. Insights might be gained from the fact that the unnormalised posterior for θ with the localised beta prior as a function of t_2 is either log-concave or log-convex, depending on y_1, t_2 and λ_2. Additional insight might be acquired from studying the differential equation corresponding to the hypergeometric function in the Bayes factor, or other saddle points methods as hinted at by Cvitković et al. (2017), López and Pagola (2011), and Temme (2003).

In our derivation of the localised Bayes factor $\mathrm{BF}_{10\,;\,a}(d)$ we extended Jeffreys's proposal of using a beta distribution on the test-relevant parameter θ by adding a location parameter on the logit scale. By representing the problem this way we could use the methods and intuitions that Jeffreys developed for his Bayesian t-test (Jeffreys, 1948, pp. 242–248; Ly et al., 2016a, 2016b).

We believe that the localised Bayes factor $\mathrm{BF}_{10\,;\,a}(d)$ for comparing two Poisson rates is consistent with Jeffreys's general philosophy of testing – more so, perhaps, than Jeffreys's own proposal from 1939. The desideratum of limit-consistency appears logical and compelling, and we hope that it can be helpful in a broad range of discrete data problems.