



**HAL**  
open science

## Priors for the Bayesian star paradox

Mikael Falconnet

► **To cite this version:**

| Mikael Falconnet. Priors for the Bayesian star paradox. 2009. hal-00429744v2

**HAL Id: hal-00429744**

**<https://hal.science/hal-00429744v2>**

Preprint submitted on 17 Nov 2009 (v2), last revised 19 Nov 2010 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# PRIORS FOR THE BAYESIAN STAR PARADOX

MIKAEL FALCONNET

ABSTRACT. We show that the Bayesian star paradox, first proved mathematically by Steel and Matsen for a specific class of prior distribution, occurs in a wider context.

## INTRODUCTION

In phylogenetics, a particular resolved tree can be highly supported even when the data is generated by an unresolved star tree. This unfortunate aspect of the Bayesian approach to phylogeny reconstruction is called the *star paradox*. Recent studies highlight that the paradox can occur in the simplest setting, namely, for an unresolved rooted tree on three taxa and two states, see Yang and Rannala [7] and Lewis et al. [1] for example. Kolaczkowski and Thornton presented in [2] some simulations and suggested that artifactual high posteriors for a particular resolved tree might disappear for very long sequences. Previous simulations in Yang and Rannala's paper were plagued by numerical problems, which left unknown the nature of the limiting distribution on posterior probabilities. For an introduction to the Bayesian approach to phylogeny reconstruction see chapter 5 of Yang [5].

The statistical question which supports the star paradox is whether the Bayesian posterior distribution of the resolutions of a star tree becomes uniform when the length of the sequence tends to infinity, that is, in the case of three taxa, whether the posterior distribution of each resolution converges to  $1/3$ . In a recent paper, Steel and Matsen [3] disprove this, thus ruining Kolaczkowski and Thornton's hope, for a specific class of branch length priors which they call *tame*. More precisely, Steel and Matsen show that, for every tame prior and every fixed  $\varepsilon > 0$ , the posterior probability of any of the three possible trees stays above  $1 - \varepsilon$  with non vanishing probability when the length of the sequence goes to infinity. This result had been taken account by Yang in [6] and reinforced by theoretical results on the posterior probabilities by Susko in [4].

Our main result is that Steel and Matsen's conclusion holds for a wider class of priors, possibly not continuous, which we call *tempered*. Recall that Steel and Matsen consider smooth priors, whose densities satisfy some regularity conditions.

The paper is organized as follows. In section 1, we describe the Bayesian framework of the star paradox. In section 2, we define the class of tempered priors for branch lengths and we state our main result. In section 3, we prove that every tame prior, in Steel and Matsen's sense, is tempered, in the sense of this paper. In section 4, we provide examples of tempered, but not tame, prior distributions. Sections 5 and 6 are devoted to the proofs of intermediate results. Finally, in section 7, we state and prove an extension of a technical lemma due to Steel and Matsen, which allows us to extend their result.

---

*Date:* November 17, 2009.

*2000 Mathematics Subject Classification.* Primary: 60J28; 92D15. Secondary: 62C10.

*Key words and phrases.* Phylogenetic trees, Bayesian statistics, star trees.

## 1. BAYESIAN FRAMEWORK FOR ROOTED TREES ON THREE TAXA

Consider three taxa, encoded by the set  $\tau = \{1, 2, 3\}$ , with two possible states. Phylogenies on  $\tau$  are supported by one of the four following trees: the star tree  $R_0$  on three taxa and, for every taxon  $i \in \tau$ , the tree  $R_i$  such that  $i$  is the outlier, hence

$$R_1 = (1, (2, 3)), \quad R_2 = (2, (1, 3)), \quad R_3 = (3, (1, 2)).$$

The phylogeny based on  $R_0$  is specified by the common length of its three branches, denoted by  $t$ . For each  $i \in \tau$ , the phylogeny based on  $R_i$  is specified by a couple of branch lengths  $(t_e, t_i)$ , where  $t_e$  denotes the external branch length and  $t_i$  the internal branch length, see figure 1.

For instance, in the phylogeny based on  $R_1$ , the divergence of taxa 2 and 3 occurred  $t_e$  units of time ago and the divergence of taxon 1 from taxa 2 and 3 occurred  $t_i + t_e$  units of time ago.

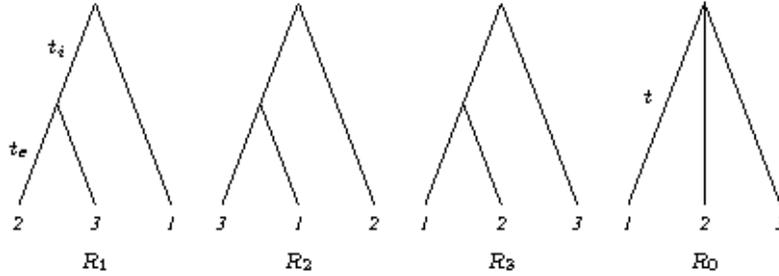


FIGURE 1. The four rooted trees for three species.

Four site patterns can occur on  $\tau$ :  $s_0$  denotes the pattern such that a given site coincides in the three taxa and, for every  $i \in \tau$ ,  $s_i$  denotes the pattern such that a given site coincide in the two other taxa and is different in taxon  $i$ . In other words, if one writes the site patterns in taxa 1, 2 and 3 in this order and  $x$  and  $y$  for any two different characters,

$$s_0 = xxx, \quad s_1 = yxx, \quad s_2 = xyx, \quad \text{and} \quad s_3 = xxy.$$

Let  $\pi$  denote the set of site patterns. As explained above, in the specific case of three taxa and two states evolving in a Jukes-Cantor model, one can choose  $\pi = \tau \cup \{0\}$ , in effect using  $i \in \pi$  as a shorthand for the collection of site patterns encoded by  $s_i$ . Assume that the counting of site pattern  $i$  is  $n_i$ . Then  $n = n_0 + n_1 + n_2 + n_3$  is the total length of the sequences and, in the independent Jukes-Cantor model considered in this paper, the quadruple  $(n_0, n_1, n_2, n_3)$  is a sufficient statistics of the sequence data. We use  $n_\sigma$  to denote any quadruple  $(n_0, n_1, n_2, n_3)$  of nonnegative integers such that  $|n_\sigma| = n_0 + n_1 + n_2 + n_3 = n \geq 1$ .

We assume that the sequences evolve according to a continuous-time Markov process with equal substitution rates 1 between the two characters.

For every  $i \in \sigma$  and every couple of branch lengths  $(t_e, t_i)$ , let  $p_i(t_e, t_i)$  denote the probability that site pattern  $s_i$  occurs on tree  $R_1$  with branch lengths  $(t_e, t_i)$ . Standard computations

provided by Yang and Rannala show that

$$\begin{aligned} 4p_0(t_e, t_i) &= 1 + e^{-4t_e} + 2e^{-4(t_i+t_e)}, \\ 4p_1(t_e, t_i) &= 1 + e^{-4t_e} - 2e^{-4(t_i+t_e)}, \\ 4p_2(t_e, t_i) &= 4p_3(t_e, t_i) = 1 - e^{-4t_e}. \end{aligned}$$

Let  $\mathfrak{T} = (T_e, T_i)$  denote a couple of positive random variables representing the branch lengths  $(t_e, t_i)$ . Let  $N = (N_0, N_1, N_2, N_3)$  denote a random variable representing the counts of sites patterns  $n_\sigma = (n_0, n_1, n_2, n_3)$ .

## 2. THE STAR TREE PARADOX

Assuming that every taxon in  $\tau$  evolved from a common ancestor, the aim of phylogeny reconstruction is to compute the most likely tree  $R_i$ . To do so, in the Bayesian approach, one places prior distributions on the trees  $R_i$  and on their branch lengths  $\mathfrak{T} = (T_e, T_i)$ .

Let  $\mathbb{P}(N = n_\sigma | R_i, \mathfrak{T})$  denote the probability that  $N = n_\sigma$  assuming that the data is generated along the tree  $R_i$  conditionally on the branch lengths  $\mathfrak{T} = (T_e, T_i)$ . One may consider  $R_1$  only since, for every  $n_\sigma$ , the symmetries of the setting yield the relations

$$\mathbb{P}(N = n_\sigma | R_2, \mathfrak{T}) = \mathbb{P}(N = (n_0, n_2, n_3, n_1) | R_1, \mathfrak{T}),$$

and

$$\mathbb{P}(N = n_\sigma | R_3, \mathfrak{T}) = \mathbb{P}(N = (n_0, n_3, n_1, n_2) | R_1, \mathfrak{T}).$$

**Notation 2.1.** For every  $i \in \tau$ , let  $\tau_i = \tau \setminus \{i\}$ . For every  $i \in \sigma$ , let  $P_i$  denote the random variable

$$P_i = p_i(\mathfrak{T}) = p_i(T_e, T_i).$$

For every  $i \in \tau$  and every  $n_\sigma$ , let  $\Pi_i(n_\sigma)$  denote the random variable

$$\Pi_i(n_\sigma) = P_0^{n_0} P_1^{n_1} P_2^{n_2+n_3}, \quad \text{with } \{i, j, k\} = \tau.$$

We recall that  $P_2 = P_3$  and we note that, if  $|n_\sigma| = n_0 + n_1 + n_2 + n_3 = n$  with  $n \geq 1$ , for every  $i \in \tau$ ,

$$\Pi_i(n_\sigma) = P_0^{n_0} P_1^{n_1} P_2^{n-n_0-n_1}.$$

Fix  $n_\sigma$  and assume that  $|n_\sigma| = n_0 + n_1 + n_2 + n_3 = n$  with  $n \geq 1$ . For every  $i \in \tau$ , the posterior probability of  $R_i$  conditionally on  $N = n_\sigma$  is

$$\mathbb{P}(R_i | N = n_\sigma) = \frac{n!}{n_0!n_1!n_2!n_3!} \frac{1}{\mathbb{P}(N = n_\sigma)} \mathbb{E}(\Pi_i(n_\sigma)).$$

Thus, for every  $i$  and  $j \in \tau$ ,

$$\frac{\mathbb{P}(R_i | N = n_\sigma)}{\mathbb{P}(R_j | N = n_\sigma)} = \frac{\mathbb{E}(\Pi_i(n_\sigma))}{\mathbb{E}(\Pi_j(n_\sigma))}.$$

For every  $\varepsilon > 0$  and every  $i \in \tau$ , let  $\mathcal{N}_i^\varepsilon$  denote the set of  $n_\sigma$  such that, for both indices  $j \in \tau$  such that  $j \neq i$ ,

$$\mathbb{E}(\Pi_i(n_\sigma)) \geq (2/\varepsilon) \mathbb{E}(\Pi_j(n_\sigma)).$$

For every  $i \in \tau$  and  $n_\sigma \in \mathcal{N}_i^\varepsilon$ ,

$$\mathbb{P}(R_i | N = n_\sigma) \geq 1 - \varepsilon,$$

which means that the posterior probability of tree  $R_i$  among the three possible trees is highly supported.

Recall that, under hypothesis  $R_0$  and for a tame prior distribution on  $\mathfrak{T} = (T_e, T_i)$ , Steel and Matsen prove that, for every  $i \in \tau$ ,  $\mathbb{P}(N \in \mathcal{N}_i^\varepsilon)$  does not go to 0 when the sequence length  $n$  goes to infinity, and consequently that the posterior probability  $\mathbb{P}(R_i|N)$  can be close to 1 even when the sequence length  $n$  is large.

We prove the same result for tempered prior distributions of  $\mathfrak{T} = (T_e, T_i)$ , which we now define.

**Notation 2.2.** (1) For every  $s \in [0, 1]$  and  $z \in [0, 3]$ , let

$$G_s(z) = \mathbb{P}(e^{-4T_e}(1 - e^{-4T_i}) \leq s | e^{-4T_e}(1 + 2e^{-4T_i}) = z).$$

(2) For every positive  $t$  and every  $i \in \sigma$ , let  $q_i$  denote the probability that site pattern  $s_i$  occurs on tree  $R_0$ , hence

$$4q_0 = 4p_0(0, t) = 1 + 3e^{-4t}, \quad 4q_1 = 4q_2 = 4q_3 = 1 - e^{-4t}.$$

(3) Let  $\ell_t$  denote a positive real number such that  $1 < 4q_0 - \ell_t$  and  $4q_0 + \ell_t < 4$ , for instance  $\ell_t = 3e^{-4t}(1 - e^{-4t})$ . Let  $I$  and  $I_t$  denote the intervals

$$I = [0, 3], \quad I_t = [4q_0 - 1 - \ell_t, 4q_0 - 1 + \ell_t] \subset ]0, 3[.$$

(4) For every positive  $t$  and integer  $n$ , let

$$Q_n(t) = \mathbb{P}(T_i \leq 1/n, t \leq T_e \leq t + 1/n).$$

**Definition 2.3** (Tempered priors). *The distribution of  $\mathfrak{T} = (T_e, T_i)$  is tempered if the following two conditions hold.*

(1) For every  $t$ , there exists  $s_0 \in ]0, 1]$ , an interval  $I_t$  around  $4q_0 - 1$ , bounded functions  $(F_i)_{i=0}^{k-1}$ , positive numbers  $\alpha$  and  $\kappa$ , and real numbers  $(\varepsilon_i)_{i=0}^k$  such that

$$0 = \varepsilon_0 < \varepsilon_1 < \dots < \varepsilon_{k-1} \leq 2 < \varepsilon_k,$$

and such that for every  $s \in [0, s_0]$  and every  $z \in I_t$ ,

$$\left| G_s(z) - \sum_{i=0}^{k-1} F_i(z) s^{\alpha + \varepsilon_i} \right| \leq \kappa s^{\alpha + \varepsilon_k}.$$

(2) For every positive  $t$ ,  $n^{-1} \log Q_n(t) \rightarrow 0$  when  $n \rightarrow \infty$ .

We now state our main result, which is the extension of Steel and Matsen's result to our more general setting.

**Theorem 2.4.** *Consider sequences of length  $n$  generated by a star tree  $R_0$  on 3 taxa with strictly positive edge length  $t$ . Let  $N$  be the resulting data, summarized by site pattern counts. Consider any prior on the three resolved trees  $(R_1, R_2, R_3)$  and a tempered prior distribution on their branch lengths  $\mathfrak{T} = (T_e, T_i)$ .*

*Then, for every  $i \in \tau$ , for every positive  $\varepsilon$ , there exists a positive  $\delta$  such that, when  $n$  is large enough,*

$$\mathbb{P}((\mathbb{P}(R_i|N) \geq 1 - \varepsilon) \geq \delta).$$

The rest of this section is devoted to the sketch of the proof of theorem 2.4. We use the definitions below. Note that the set  $F_c^{(n)}$  is not the set introduced by Steel and Matsen. For a technical reason in the proof of proposition 2.6 stated below, we had to modify their definition. Note however that propositions 2.6 and 2.7 below are adaptations of ideas in Steel and Matsen's paper.

**Notation 2.5.** For every  $i$  in  $\sigma$ , let  $\Delta_i$  denote the function defined as follows. For every nonnegative integers  $n_\sigma = (n_0, n_1, n_2, n_3)$  such that  $|n_\sigma| = n_0 + n_1 + n_2 + n_3 = n$  with  $n \geq 1$ ,

$$\Delta_0(n_\sigma) = \frac{n_0 - q_0 n}{\sqrt{n}},$$

and, for every  $i \in \tau$ ,

$$\Delta_i(n_\sigma) = \frac{n_i - 1/3(n - n_0)}{\sqrt{n}}.$$

For every  $c > 1$ , introduce

$$F_c^{(n)} = \{n_\sigma; |n_\sigma| = n, -2c \leq \Delta_2(n_\sigma) \leq -c, -2c \leq \Delta_3(n_\sigma) \leq -c, -c \leq \Delta_0(n_\sigma) \leq 0\}.$$

For every  $i \in \tau$  and every positive  $\eta$ , let  $A_\eta^i$  denote the event

$$A_\eta^i = \{\forall j \in \tau_i, \mathbb{E}(\Pi_i(N) | N) \geq \eta \mathbb{E}(\Pi_j(N) | N)\}.$$

Since  $\Delta_1 + \Delta_2 + \Delta_3 = 0$ , every  $n_\sigma$  in  $F_c^{(n)}$  is such that  $2c \leq \Delta_1(n_\sigma) \leq 4c$ . We note that  $F_c^{(n)}$  is not symmetric about  $\tau$  and gives a preference to 1. That is why we only deal with  $A_\eta^1$  in the following proof. To deal with  $A_\eta^i$ , it suffices to change  $F_c^{(n)}$ .

*Proof of theorem 2.4.* From the reasoning in section 2, it suffices to prove that for every positive  $\eta$ , there exists a positive  $\delta$  such that, when  $n$  is large enough,

$$\mathbb{P}(A_\eta^1) \geq \delta.$$

Suppose that one generates  $n \geq 1$  sites on the star tree  $R_0$  with given branch length  $t$  and let  $N$  be the counts of site patterns defined in section 1, hence  $N_0 + N_1 + N_2 + N_3 = n$ .

When  $n$  is large enough, central limit estimates show that the probability of the event  $\{N \in F_c^{(n)}\}$  is uniformly bounded from below, say by  $\delta > 0$ . Hence,

$$\mathbb{P}(A_\eta^1) \geq \delta \mathbb{P}(A_\eta^1 | N \in F_c^{(n)})$$

We wish to prove that there exists a positive  $\alpha$  independent of  $c$  such that for  $n$  large enough and for every  $n_\sigma \in F_c^{(n)}$ , and for every  $j \in \tau_1$ ,

$$\mathbb{E}(\Pi_1(n_\sigma)) \geq c^2 \alpha \mathbb{E}(\Pi_j(n_\sigma)).$$

This follows from the two propositions below, adapted from Steel and Matsen's paper.

**Proposition 2.6.** Fix  $t$  and assume that  $n_\sigma \in F_c^{(n)}$ . Then, when  $n$  is large enough, for every  $j \in \tau_1$ ,

$$\mathbb{E}(\Pi_j(n_\sigma) | 4P_0 - 1 \in I_t) \geq \mathbb{E}(\Pi_j(n_\sigma) | 4P_0 - 1 \notin I_t).$$

**Proposition 2.7.** Fix  $t$  and assume that  $n_\sigma \in F_c^{(n)}$ . Then, there exists a positive  $\alpha$ , independent of  $c$ , such that for every  $z \in I_t$ , and for every  $j \in \tau_1$ ,

$$\mathbb{E}(\Pi_1(n_\sigma) | 4P_0 - 1 = z) \geq c^2 \alpha \mathbb{E}(\Pi_j(n_\sigma) | 4P_0 - 1 = z).$$

From these two propositions, for every  $j \in \tau_1$ ,

$$\mathbb{E}(\Pi_1(n_\sigma)) \geq c^2 \alpha \mathbb{P}(4P_0 - 1 \in I_t) \mathbb{E}(\Pi_j(n_\sigma)).$$

Assume that  $c$  is so large that  $c^2 \alpha \mathbb{P}(4P_0 - 1 \in I_t) \geq \eta$ . Then, for every  $n_\sigma \in F_c^{(n)}$ , for every  $j \in \tau_1$ ,

$$\mathbb{E}(\Pi_1(n_\sigma)) \geq \eta \mathbb{E}(\Pi_j(n_\sigma)).$$

This implies that

$$\mathbb{P}\left(A_\eta^1 \mid N \in F_c^{(n)}\right) = 1,$$

which yields the theorem.  $\square$

### 3. THE TAME CASE

We show in this section that tame priors are tempered.

**Proposition 3.1.** *Assume that  $\mathfrak{T} = (T_e, T_i)$  has a smooth joint probability density  $\omega$  which is bounded and everywhere non zero. Then the distribution of  $\mathfrak{T} = (T_e, T_i)$  is tempered.*

In particular, every tame prior fulfills the hypothesis of proposition 3.1, hence every tame prior is tempered, as claimed in the introduction.

**Notation 3.2.** *Introduce the random variables*

$$(S_e, S_i) = \zeta(T_e, T_i), \quad \zeta(t_e, t_i) = (e^{-4t_e}, 1 + 2e^{-4t_i}),$$

that is,

$$S_e = e^{-4T_e}, \quad S_i = 1 + 2e^{-4T_i}.$$

Hence,  $G_s(z)$  is also

$$G_s(z) = \mathbb{P}(3S_e \leq 2s + z \mid S_e S_i = z),$$

and the distribution of  $(S_e, S_i)$  has a smooth joint probability density  $\bar{\omega}$ , defined on  $0 < x \leq 1 < y \leq 3$  by

$$\bar{\omega}(x, y) = \frac{\omega \circ \zeta^{-1}(x, y)}{16x(y-1)}.$$

*Proof of proposition 3.1.* For tame priors, the probability  $Q_n(t)$  introduced in condition (2) of definition 2.3 is of order  $1/n^2$ . Thus condition (2) of definition 2.3 holds.

The definition of  $G_s(z)$  as a conditional expectation can be rewritten as

$$G_s(z) = \mathbb{P}(3S_e \leq 2s + S_e S_i \mid S_e S_i = z).$$

Hence, for every measurable bounded function  $H$ ,

$$\mathbb{E}(H(S_e S_i); 3S_e \leq 2s + S_e S_i) = \mathbb{E}(H(S_e S_i) G_s(S_e S_i)),$$

that is,

$$\iint H(xy) \mathbf{1}\{3x \leq 2s + xy\} \bar{\omega}(x, y) dx dy = \iint H(xy) G_s(xy) \bar{\omega}(x, y) dx dy.$$

The change of variable  $z = xy$  yields

$$\iint H(z) \mathbf{1}\{3x \leq 2s + z\} \bar{\omega}(x, z/x) dz dx/x = \iint H(z) G_s(z) \bar{\omega}(x, z/x) dz dx/x.$$

This should hold for every measurable bounded function  $H$ , hence one can choose

$$G_s(z) = H_z(s)/H_z(\infty),$$

with

$$H_z(s) = \int \mathbf{1}\{3x \leq 2s + z\} \bar{\omega}(x, z/x) dx/x.$$

Since  $0 \leq S_e \leq 1 \leq S_i \leq 3$  almost surely, the integral defining  $H_z(s)$  may be further restricted to the range  $0 \leq x \leq 1$  and  $z/3 \leq x \leq z$ . Finally, for every  $s \geq 0$  and  $z \in [0, 3]$ ,

$$G_s(z) = H_z(s)/H_z(1),$$

where

$$H_z(s) = \int_{m(0,z)}^{m(s,z)} \varpi(x, z/x) dx/x, \quad \text{with } m(s, z) = \min\{1, z, (2s+z)/3\}.$$

Hence,  $m(0, z) = z/3$  and, for small positive values of  $s$ ,  $m(s, z) = m(0, z) + 2s/3$ . When  $0 \leq z \leq 1$ ,  $m(s, z) \rightarrow m(\infty, z) = z$  when  $s \rightarrow \infty$  and this limit is reached for  $s = z$ . When  $1 \leq z \leq 3$ ,  $m(s, z) \rightarrow m(\infty, z) = 1$  when  $s \rightarrow \infty$  and this limit is reached for  $s = (3-z)/2$ . In both cases,  $m(\infty, z) = m(1, z)$  hence  $H_z(\infty) = H_z(1)$ .

Because  $\omega$  and  $\zeta^{-1}$  are smooth, Taylor-Lagrange formula shows that, for every  $s \geq 0$  and every fixed  $z$ ,

$$H_z(s) = H_z(0) + H_z'(0)s + \frac{1}{2}H_z''(0)s^2 + \frac{1}{6}H_z^{(3)}(0)s^3 + \frac{1}{24} \int_0^s (x-s)^3 H_z^{(4)}(s) dx.$$

Simple computations yield  $H_z(0) = 0$  and the values of  $H_z'(0)$ ,  $H_z''(0)$  and  $H_z^{(3)}(0)$  as combinations of  $\omega$  and of partial derivatives of  $\omega$ , evaluated at the point  $(\vartheta, 0)$ , where  $3e^{-4\vartheta} = z$ .

Furthermore, the hypothesis on  $\omega$  ensures that  $H_z^{(4)}$  is bounded, in the following sense: there exists positive numbers  $s_0$  and  $\kappa_0$  such that for every  $s \in [0, s_0]$  and every  $z \in I_t$ ,

$$\left| H_z^{(4)}(s) \right| \leq 24\kappa_0.$$

Hence,  $\mathfrak{T} = (T_e, T_i)$  fulfills the first condition to be tempered, with

$$k = 3, \quad \alpha = 1, \quad \varepsilon_1 = 1, \quad \varepsilon_2 = 2, \quad \varepsilon_3 = 3, \quad \kappa = \kappa_0,$$

and, for every  $0 \leq i \leq 2$ ,

$$F_i(z) = H_z^{(i+1)}(0)/H_z(1).$$

Finally, since  $\omega$  is smooth, the functions  $F_i$  are bounded on  $I_t$ . □

#### 4. A DISCRETE EXAMPLE

We provide an example such that the condition of Steel and Matsen fails but our result applies.

**Definition 4.1.** Let  $a > 0$  and  $b > 0$ . Let  $(t_n)$ ,  $(y_n)$  and  $(r_n)$  denote sequences of positive numbers, indexed by  $n \geq 1$ , and  $r$  a positive number, defined by the formulas

$$t_n = n^{-a}, \quad y_n = 1 + 2e^{-4t_n} \quad r_n = y_n [n^{-b} - (n+1)^{-b}], \quad r = \sum_{n \geq 1} r_n.$$

**Proposition 4.2.** Choose positive parameters  $a$  and  $b$  such that  $3a < b$  and  $3a < 1$ . Assume that  $T_i$  is a discrete random variable such that, for every  $n \geq 1$ ,

$$\mathbb{P}(T_i = t_n) = r_n/r.$$

Assume that  $T_e$  is a continuous random variable, independent of  $T_i$ , with exponential law of parameter 4, that is, with density  $4e^{-4t}$  on  $t \geq 0$  with respect to the Lebesgue measure.

Then, the distribution of  $\mathfrak{T} = (T_e, T_i)$  is not tame but tempered, for the parameters

$$k = 3, \quad \alpha = b/a, \quad \varepsilon_1 = 1, \quad \varepsilon_2 = 2, \quad \varepsilon_3 = 3,$$

and some explicit functions  $F_0$ ,  $F_1$  and  $F_2$ .

*Proof of proposition 4.2.* Recall that, using the random variables  $S_e = e^{-4T_e}$  and  $S_i = 1 + 2e^{-4T_i}$ , the function  $G_s$  is characterized by the fact that, for every measurable bounded function  $H$ ,

$$\mathbb{E}(H(S_e S_i) : S_e(3 - S_i) \leq 2s) = \mathbb{E}(H(S_e S_i) G_s(S_e S_i)).$$

Here,  $S_e$  and  $S_i$  are independent, the distribution of  $S_e$  is uniform on  $[0, 1]$  and the distribution of  $S_i$  is discrete with

$$\mathbb{P}(S_i = y_n) = r_n/r.$$

Thus,

$$\sum_n r_n \int_0^1 H(xy_n) \mathbf{1}\{x(3 - y_n) \leq 2s\} dx = \sum_n r_n \int_0^1 H(xy_n) G_s(xy_n) dx.$$

The changes of variable  $z = y_n x$  in each integral yield

$$\sum_n (r_n/y_n) \int H(z) \mathbf{1}\{z \leq y_n\} \mathbf{1}\{3z \leq (2s + z)y_n\} dz = \sum_n (r_n/y_n) \int H(z) \mathbf{1}\{z \leq y_n\} G_s(z) dz.$$

This should hold for every measurable bounded function  $H$ , hence

$$G_s(z) = H_z(s)/H_z(\infty), \quad H_z(s) = \sum_n (r_n/y_n) \mathbf{1}\{z \leq y_n\} \mathbf{1}\{3z \leq (2s + z)y_n\}.$$

Since  $r_n/y_n = n^{-b} - (n+1)^{-b}$  for  $n \geq 1$ ,  $H_z(s) = n(z, s)^{-b}$  where

$$n(z, s) = \inf\{n \geq 1 \mid z \leq y_n, 3z \leq (2s + z)y_n\}.$$

Since  $y_n \rightarrow 3$  when  $n \rightarrow \infty$ ,  $n(z, s)$  is finite for every  $z < 3$  and  $s > 0$ .

For every  $z > 0$ , when  $s$  is large enough, namely  $s \geq (3 - z)/2$ , the condition  $3z \leq (2s + z)y_n$  becomes useless and

$$n(z, s) = \inf\{n \geq 1 \mid z \leq y_n\},$$

hence  $n(z, s)$  and  $H_z(s)$  are independent of  $s$ . If  $z \geq 1$ , this implies that  $n(z, s)$  and  $H_z(s)$  are independent of  $s \geq 1$ . If  $z < 1$  and  $s \geq 1$ , the conditions  $z \leq y_n$  and  $3z \leq (2s + z)y_n$  both hold for every  $n \geq 1$  hence  $n(z, s) = 1$  and  $H_z(s) = 1$ . In both cases,  $H_z(\infty) = H_z(1)$ .

We are interested in small positive values of  $s$ . For every  $z < 3$ , when  $s$  is small enough, namely  $s \leq (3 - z)/2$ , the condition  $z \leq y_n$  becomes useless and

$$n(z, s) = \inf\{n \geq 1 \mid 3z \leq (2s + z)y_n\},$$

When furthermore  $s < z$ ,  $n \geq n(z, s)$  is equivalent to the condition

$$n^{-a} \leq h(s/z), \quad \text{with} \quad h(u) = -\frac{1}{4} \ln \left( 1 - \frac{3u}{1 + 2u} \right), \quad 0 \leq u < 1.$$

Finally, for every  $s < \min\{z, (3 - z)/2\}$ ,  $n(z, s)$  is the unique integer such that

$$n(z, s) - 1 < h(s/z)^{-1/a} \leq n(z, s).$$

This reads as

$$h(u)^{b/a} [1 + h(u)^{1/a}]^{-b} < H_z(1) G_s(z) \leq h(u)^{b/a}, \quad u = s/z.$$

One sees that the function  $h$  is analytic and that  $h(u) = (3u/4) + o(u)$  when  $u \rightarrow 0$ , hence,

$$h(u)^{b/a} = (3u/4)^{b/a} (1 + a_1 u + a_2 u^2 + a_3 u^3 + o(u^3)),$$

when  $u \rightarrow 0$ , for given coefficients  $a_1$ ,  $a_2$  and  $a_3$ . Likewise, since  $1/a > 3$ ,  $h(u)^{1/a} = o(u^3)$  when  $u \rightarrow 0$ . This implies that

$$[1 + h(u)^{1/a}]^{-b} = 1 + o(u^3),$$

hence

$$H_z(1)G_s(z) = (3u/4)^{b/a}(1 + a_1u + a_2u^2 + a_3u^3 + o(u^3)).$$

This yields the first part of definition 2.3, with

$$k = 3, \quad \alpha = b/a, \quad (\varepsilon_1, \varepsilon_2, \varepsilon_3) = (1, 2, 3),$$

and

$$F_0(z) = (3/4z)^{b/a}/H_z(1), \quad F_1(z) = a_1F_0(z)/z, \quad F_2(z) = a_2F_0(z)/z^2.$$

The remaining step is to get rid of the dependencies over  $z$  of our upper bounds. For instance, the reasoning above provides as an error term a multiple of

$$u^{\alpha+3}/H_z(1) = s^{\alpha+3}/(z^{\alpha+3}H_z(1)),$$

instead of a constant multiple of  $s^{\alpha+3}$ . But  $\inf I_t > 0$ , hence the  $1/z^{\alpha+3}$  contribution is uniformly bounded.

As regards  $H_z(1)$ , we first note that  $H_z(1) = 1$  if  $z \leq 1$ . If  $z \geq 1$ , elementary computations show that  $H_z(1) \geq c$  if and only if  $n(z, 1) \leq c^{-1/b}$  if and only if  $\exp(-c^{a/b}) \geq (z-1)/2$ , which is implied by the fact that  $1 - c^{a/b} \geq (z-1)/2$ , which is equivalent to the upper bound  $c^{a/b} \leq (3-z)/2$ . Since  $\sup I_t < 3$ , this can be achieved uniformly over  $z \in I_t$  and  $1/H_z(1)$  is uniformly bounded as well.

Finally, we asked for an expansion valid on  $s \leq s_0$ , for a fixed  $s_0$ , and we proved an expansion valid over  $s/z \leq u_0$ , for a fixed  $u_0$ . But one can choose  $s_0 = u_0 \inf I_t$ . This concludes the proof that the conditions in the first part of definition 2.3 hold.

We now prove that the second part of definition 2.3 holds. Since  $T_i$  and  $T_e$  are independent, for every positive integer  $n$ ,

$$Q_n(t) = \mathbb{P}(T_i \leq 1/n) \mathbb{P}(t \leq T_e \leq t + 1/n).$$

One has

$$n\mathbb{P}(t \leq T_e \leq t + 1/n) \rightarrow 4e^{-4t} \quad \text{when } n \rightarrow +\infty,$$

and

$$\frac{1}{r(n^{1/a} + 1)^b} \leq \mathbb{P}(T_i \leq 1/n) \leq \frac{3}{rn^{b/a}}.$$

Since  $Q_n(t)$  is bounded from below by a multiple of  $1/n^{1+b/a}$ , the second point of definition 2.3 holds.  $\square$

## 5. PROOF OF PROPOSITION 2.6

The proof is decomposed into two intermediate results, stated as lemmata below and using estimates on auxiliary random variables introduced below.

**Notation 5.1.** For every  $n \geq 1$  and  $t > 0$ , let  $\Gamma_t(n) = [0, 1/n] \times [t, t + 1/n]$ .

For every  $t > 0$ , let  $\mu_t = q_0^{q_0} q_1^{q_1} q_2^{q_2} q_3^{q_3} = q_0^{q_0} q_1^{3q_1}$  and  $U_t$  denote the random variable

$$U_t = \prod_{i \in \sigma} (P_i/q_i)^{q_i}.$$

For every  $n_\sigma$ , for every  $j \in \tau_1$ , let  $W_j(n_\sigma)$  denote the random variable

$$W_j(n_\sigma) = P_0^{\Delta_0(n_\sigma)} P_1^{\Delta_j - \Delta_0/3(n_\sigma)} P_2^{\Delta_1 + \Delta_k - 2\Delta_0/3(n_\sigma)}, \quad \text{with } \{j, k\} = \tau_1.$$

One sees that

$$U_t = P_0^{q_0} P_1^{q_1} P_2^{2q_1} / \mu_t, \quad Q_n(t) = \mathbb{P}(\mathfrak{T} \in \Gamma_t(n)),$$

and

$$W_j = (P_0/P_2)^{\Delta_0} (P_1/P_2)^{\Delta_j - \Delta_0/3}.$$

**Lemma 5.2.** (1) For every  $n_\sigma \in F_c^{(n)}$ , for every  $j \in \tau_1$ ,  $W_j(n_\sigma) \leq 1$ .

(2) For every  $n_\sigma \in F_c^{(n)}$ , for every  $j \in \tau_1$ ,  $W_j(n_\sigma) \geq (q_1)^c$  on the event  $\{\mathfrak{T} \in \Gamma_t(n)\}$ .

(3) There exists a finite constant  $\kappa$  such that  $U_t^n \geq e^{-\kappa}$  uniformly on  $n \geq 1$  and  $\{\mathfrak{T} \in \Gamma_t(n)\}$ .

*Proof of lemma 5.2.* (1) For every  $\mathfrak{T}$ ,  $P_0 \geq P_1 \geq P_2$ . On  $F_c^{(n)}$ ,  $\Delta_0 \leq 0$  and for every  $j \in \tau_1$ ,  $\Delta_j - \Delta_0/3 \leq 0$  hence

$$(P_0/P_1)^{\Delta_0} \leq 1, \quad (P_0/P_2)^{\Delta_j - \Delta_0/3} \leq 1.$$

This proves the claim.

(2) One has  $P_0 \leq 1$  everywhere and  $P_1 \geq q_1$  and  $P_2 \geq q_1$  on the event  $\{\mathfrak{T} \in \Gamma_t(n)\}$ . On  $F_c^{(n)}$ ,  $\Delta_0 \leq 0$  and for every  $j \in \tau_1$ ,  $\Delta_j - \Delta_0/3 \leq 0$  hence  $W_j \geq q_2^{-\Delta_j - 2\Delta_0/3}$ . Finally, on  $F_c^{(n)}$ ,  $\Delta_j + 2\Delta_0/3 \leq -c$ . This proves the claim.

(3) For every  $\mathfrak{T} \in \Gamma_t(n)$ ,  $T_i \geq 0$  and  $T_e \geq t$ , hence  $P_1 \geq q_1$  and  $P_2 \geq q_2 = q_1$ . Likewise,  $T_i \leq 1/n$  and  $T_e \leq t + 1/n$  hence  $P_0 \geq p_0(1/n, t + 1/n) \geq q_0 - 5e^{-4t}(1 - e^{-4/n})/4$ . This yields that, for every  $n \geq 1$  and  $\mathfrak{T} \in \Gamma_t(n)$ ,

$$U_t^n \geq (1 - 5e^{-4t}/(q_0 n))^n \rightarrow \exp(-5e^{-4t}/q_0) > 0,$$

which implies the desired lower bound.  $\square$

**Lemma 5.3.** For every  $n_\sigma \in F_c^{(n)}$ , for every  $j \in \tau_1$ ,

$$\mathbb{E}(\Pi_j(n_\sigma) | 4P_0 - 1 \in I_t) \geq \mu_t^n Q_n(t) e^{-O(\sqrt{n})},$$

and

$$\mathbb{E}(\Pi_j(n_\sigma) | 4P_0 - 1 \notin I_t) \leq \mu_t^n e^{-(\ell_t^2/32)n}.$$

*Proof of lemma 5.3.* Since  $P_0 = p_0(\mathfrak{T})$ , for every  $\mathfrak{T} \in \Gamma_t(n)$ , when  $n$  is large,  $4P_0 - 1 \in I_t$ . Consequently,

$$\mathbb{E}(\Pi_j(n_\sigma) | 4P_1 - 1 \in I_t) \geq Q_n(t) \mathbb{E}(\Pi_j(n_\sigma) | \mathfrak{T} \in \Gamma_t(n)).$$

On the event  $\{\mathfrak{T} \in \Gamma_t(n)\}$ ,

$$\Pi_j(n_\sigma) = \mu_t^n U_t^n W_j(n_\sigma)^{\sqrt{n}} \geq \mu_t^n e^{-\kappa} (q_1)^{c\sqrt{n}},$$

from parts (2) and (3) of lemma 5.2, which proves the first part of the lemma.

Turning to the second part, let  $d_{KL}$  denote the Kullback-Leibler distance between probability measures. When  $4P_0 - 1 \notin I_t$ ,

$$d_{KL}(q_{0:3}, P_{0:3}) \geq (1/2) \|q_{0:3} - P_{0:3}\|_1^2 \geq (1/2)(q_0 - P_0)^2 \geq \ell_t^2/(32).$$

Note that

$$\Pi_j(n_\sigma) = \mu_t^n W_j(n_\sigma)^{\sqrt{n}} e^{-nd_{KL}(q_{0:3}, P_{0:3})},$$

hence the estimate on  $d_{KL}(q_{0:3}, P_{0:3})$ , and part (1) of lemma 5.2, imply the second part of the lemma.  $\square$

Turning finally to the proof of proposition 2.6, we note that  $Q_n(t) = e^{o(n)}$  because the distribution of  $\mathfrak{T}$  is tempered. Furthermore, lemma 5.3 shows that, when  $n$  is large enough,

$$\mathbb{E}(\Pi_j(n_\sigma) | 4P_0 - 1 \in I_t) \geq \mathbb{E}(\Pi_j(n_\sigma) | 4P_0 - 1 \notin I_t).$$

This concludes the proof of proposition 2.6.

## 6. PROOF OF PROPOSITION 2.7

Our proof of proposition 2.7 is based on lemma 6.2 and proposition 6.3 below.

**Notation 6.1.** For every  $u$  in  $[0, 1]$ , let  $\zeta(u) = (1 + 2u)(1 - u)^2$ . Let  $U$  and  $V$  denote the random variables defined as

$$U = (P_1 - P_2)/(1 - P_0), \quad V = \zeta(U).$$

**Lemma 6.2.** For every  $n_\sigma \in F_c^{(n)}$ , for every  $j \in \tau_1$ ,

$$\frac{\mathbb{E}(\Pi_1(n_\sigma) | P_0)}{\mathbb{E}(\Pi_j(n_\sigma) | P_0)} \geq 4c^2 n \frac{\mathbb{E}(V^s(1 - V) | P_0)}{\mathbb{E}(V^s | P_0)}, \quad s = (n - n_0)/3.$$

**Proposition 6.3.** Assume that the distribution of  $\mathfrak{T}$  is tempered. There exists  $\theta$  and  $\alpha$ , both positive and independent of  $c$ , such that for every  $s \geq \theta$ , on the event  $\{4P_0 - 1 \in I_t\}$ ,

$$4s \mathbb{E}(V^s(1 - V) | P_0) \geq \alpha \mathbb{E}(V^s | P_0).$$

Assuming this, the proof of proposition 2.7 is as follows. Let  $s$ ,  $\theta$  and  $\alpha$  as in lemma 6.2 and proposition 6.3. Since  $n - n_0 = (1 - q_0)n - \Delta_0 \sqrt{n} \geq (1 - q_0)n$  for every  $n_\sigma \in F_c^{(n)}$ ,  $s = (n - n_0)/3 \geq \theta$  when  $n$  is large enough. Furthermore,  $s \leq n/3$ . Finally, for every  $n_\sigma \in F_c^{(n)}$  with  $n$  large enough, on  $\{4P_0 - 1 \in I_t\}$ , for every  $j \in \tau_1$ ,

$$\mathbb{E}(\Pi_1(n_\sigma) | P_0) \geq 3c^2 \alpha \mathbb{E}(\Pi_j(n_\sigma) | P_0).$$

This concludes the proof of proposition 2.7.

*Proof of lemma 6.2.* Recall that, for every  $c > 1$ ,  $F_c^{(n)}$  is

$$F_c^{(n)} = \{n_\sigma : |n_\sigma| = n, -2c \leq \Delta_2(n_\sigma) \leq c, -2c \leq \Delta_3(n_\sigma) \leq c, -c \leq \Delta_0(n_\sigma) \leq 0\}.$$

Using the  $\Delta$  variables, one can rewrite  $\Pi_1$ ,  $\Pi_2$  and  $\Pi_3$  as

$$\Pi_i(n_\sigma) = P_0^{n_0} (P_1 P_2^2)^s (P_1/P_2)^{\Delta_i(n_\sigma) \sqrt{n}}, \quad i = 1, 2, 3, \quad s = (n - n_0)/3.$$

Assume that  $n_\sigma \in F_c^{(n)}$ . Then,  $\Delta_1(n_\sigma) \geq 2c$ , for every  $j \in \tau_1$ ,  $\Delta_j(n_\sigma) \leq 0$  and  $P_1 \geq P_2$ , hence

$$\Pi_1(n_\sigma) \geq P_0^{n_0} (P_1 P_2^2)^s (P_1/P_2)^{2c \sqrt{n}}, \quad \Pi_j(n_\sigma) \leq P_0^{n_0} (P_1 P_2^2)^s.$$

Furthermore,

$$P_1 P_2^2 = (1/27)V(1 - P_0)^3, \quad P_1/P_2 = (1 + 2U)/(1 - U),$$

hence for every  $j \in \tau_1$ ,

$$\frac{\mathbb{E}(\Pi_1(n_\sigma) | P_0)}{\mathbb{E}(\Pi_j(n_\sigma) | P_0)} \geq \frac{\mathbb{E}\left(V^s \left(\frac{1 + 2U}{1 - U}\right)^{2c \sqrt{n}} | P_0\right)}{\mathbb{E}(V^s | P_0)}.$$

Direct computations (or lemma 3.2 in Steel and Matsen [3]) show that, for every  $u$  in  $[0, 1]$  and every  $m \geq 3$ ,

$$\left(\frac{1 + 2u}{1 - u}\right)^m \geq m^2(1 - \zeta(u)),$$

hence

$$((1 + 2U)/(1 - U))^{2c\sqrt{n}} \geq 4c^2n(1 - V).$$

The conclusion of lemma 6.2 follows.  $\square$

*Proof of proposition 6.3.* We recall that  $U$  and  $V$  denote the random variables defined as

$$U = (P_1 - P_2)/(1 - P_0), \quad V = \zeta(U), \quad \zeta(u) = (1 + 2u)(1 - u)^2.$$

To use proposition 7.2, one must compute a Taylor expansion at  $v = 1^-$  or, equivalently, at  $u = 0^+$ , of the conditional probability

$$\mathbb{P}(V \geq v | P_0) = \mathbb{P}(U \leq u | P_0),$$

where  $u = \zeta^{-1}(v)$ . Besides, for  $v$  close to 1,

$$u = \zeta^{-1}(v) = w/\sqrt{3} + w^2/9 + 5w^3/54\sqrt{3} + O(w^4), \quad \text{with } w = \sqrt{1 - v}.$$

Since  $U = (P_1 - P_2)/(1 - P_0)$ ,

$$\mathbb{P}(U \leq u | 4P_0 - 1 = z) = \mathbb{P}(S_e(3 - S_i) \leq 2s | S_e S_i = z),$$

where we recall that

$$S_e = e^{-4T_e}, \quad S_i = 1 + 2e^{-4T_i}, \quad 2s = u(3 - z).$$

Keeping the notation given in definition 2.3, one has

$$G_s(z) = \mathbb{P}(S_e(3 - S_i) \leq 2s | S_e S_i = z).$$

Since the distribution of  $\mathfrak{T}$  is tempered, there exists  $n$  bounded functions  $(F_i)_{i=0}^{n-1}$  on  $I_t$ , a positive number  $\alpha$ ,  $n + 1$  real numbers

$$0 = \varepsilon_0 < \varepsilon_1 < \dots < \varepsilon_{n-1} \leq 2 < \varepsilon_n,$$

and two positive numbers  $\kappa$  and  $s_0$  such that for every  $0 \leq s \leq s_0$  and every  $z \in I_t$ ,

$$\left| G_s(z) - \sum_{i=0}^{n-1} F_i(z) s^{\alpha + \varepsilon_i} \right| \leq \kappa s^{\alpha + \varepsilon_n}.$$

Combining this with the relation  $2s = u(3 - z)$  and the expansion of  $u = \zeta^{-1}(v)$  along the powers of  $w$ , one sees that there exists bounded functions  $(f_i)_{i=0}^{n-1}$  on  $I_t$ , a positive number  $\kappa'$  and  $0 \leq v_0 < 1$  such that for every  $v_0 \leq v \leq 1$  and every  $z \in I_t$ ,

$$\left| \mathbb{P}(V \geq v | 4P_0 - 1 = z) - \sum_{i=0}^{n-1} f_i(z) (1 - v)^{\alpha/2 + \varepsilon_i/2} \right| \leq \kappa' (1 - v)^{\alpha/2 + \varepsilon_n/2}.$$

Since the functions  $f_i$  are bounded and positive on  $I_t$ , proposition 7.2 in section 7 implies that there exists a positive number  $\theta$  such that for every  $z \in I_t$  and every  $s \geq \theta$ , the conclusion of proposition 6.3 holds.  $\square$

## 7. EXTENSION OF STEEL AND MATSEN'S LEMMA

The Bayesian star paradox due to Steel and Matsen relies on a technical result which we slightly rephrase as follows. For every nonnegative real  $t$  and every  $[0, 1]$  valued random variable  $V$ , introduce

$$M_t = \mathbb{E}(V^t), \quad R_t = 1 - \frac{M_{t+1}}{M_t} = \frac{\mathbb{E}(V^t(1-V))}{\mathbb{E}(V^t)}.$$

**Proposition 7.1** (Steel and Matsen's lemma). *Let  $0 \leq \eta < 1$  and  $B > 0$ . There exists a finite  $K$ , which depends on  $\eta$  and  $B$  only, such that the following holds. For every  $[0, 1]$  valued random variable  $V$  with a smooth probability density function  $f$  such that  $f(1) > 0$  and  $|f'(v)| \leq Bf(1)$  for every  $\eta \leq v \leq 1$ , and for every integer  $k \geq K$ ,*

$$2kR_k \geq 1.$$

Indeed the asymptotics of  $R_k$  when  $k$  is large depends on the distribution of  $V$  around 1. We prove in the following proposition that the conclusion of Steel and Matsen's lemma above holds for a wider class of random variables.

**Proposition 7.2.** *Let  $V$  a random variable on  $[0, 1]$ . Suppose that there exists real numbers  $0 \leq v_0 < 1$ ,  $\alpha > 0$ ,  $(\varepsilon_i)_{i=0}^n$  and  $(\gamma_i)_{i=0}^n$ , such that*

$$0 = \varepsilon_0 < \varepsilon_1 < \dots < \varepsilon_{n-1} \leq 1 < \varepsilon_n,$$

and, for every  $v_0 \leq v \leq 1$ ,

$$\left| \mathbb{P}(V \geq v) - \sum_{i=0}^{n-1} \gamma_i (1-v)^{\alpha+\varepsilon_i} \right| \leq \gamma_n (1-v)^{\alpha+\varepsilon_n}.$$

Then there exists a finite  $\tau(\gamma_{0:n})$ , which depends continuously on  $\gamma_{0:n}$ , such that for every  $t \geq \tau(\gamma_{0:n})$ ,

$$2tR_t \geq \alpha.$$

**Remark 1.** *We insist on the fact that  $\tau(\gamma_{0:n})$  depends continuously on  $\gamma_{0:n}$ . To wit, in the proof of proposition 6.3, we apply proposition 7.2 with bounded functions of  $z$ . This means that for every  $z \in I_t$ , one gets a number  $\tau$  which depends on  $z$  through the bounded functions such the control on the distribution of  $V$  holds. The continuity of  $\tau$  ensures that there exists a number independent of  $z$  such that proposition 6.3 holds.*

**Remark 2.** *If one computes a Taylor expansion of the function  $v \mapsto \mathbb{P}(V \geq v)$  at  $v = 1^-$  under the conditions of Steel and Matsen's lemma, one can see that conditions of proposition 7.2 hold. It follows that proposition 7.2 is an extension of Steel and Matsen's lemma.*

*Proof of proposition 7.2.* For fixed values of  $\alpha$ ,  $(\gamma_i)_{i=0}^n$  and  $(\varepsilon_i)_{i=0}^n$ , introduce, for every  $t > 0$ ,

$$M_t^\pm = \int_0^1 t v^{t-1} F_\pm(v) dv, \quad \text{where } F_\pm(v) = \sum_{i=0}^{n-1} \gamma_i (1-v)^{\alpha+\varepsilon_i} \pm \gamma_n (1-v)^{\alpha+\varepsilon_n}.$$

Hence,

$$M_t = \int_0^1 t v^{t-1} \mathbb{P}(V \geq v) dv = M_t^\pm + \int_0^1 t v^{t-1} [\mathbb{P}(V \geq v) - F_\pm(v)] dv,$$

and

$$M_t^\pm = tB(t, \alpha + 1) \left( \sum_{i=1}^{n-1} \gamma_i \Lambda(\varepsilon_i, t) P(\varepsilon_i, t) \pm \gamma_n \Lambda(\varepsilon_n, t) P(\varepsilon_n, t) \right),$$

where

$$\Lambda(\varepsilon, t) = \frac{\Gamma(\{t\} + \alpha + 1)}{\Gamma(\{t\} + \alpha + \varepsilon + 1)}, \quad P(\varepsilon, t) = \prod_{\ell=1}^{\lfloor t \rfloor + 1} \left( 1 - \frac{\varepsilon}{\alpha + \varepsilon + \{t\} + \ell} \right),$$

and  $B$  denotes the beta function

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

From the control of the distribution of  $V$ ,

$$M_t^- - \gamma v_0^t \leq M_t \leq M_t^+ + v_0^t \quad \text{where } \gamma = \sum_{i=0}^n |\gamma_i|.$$

Combining this with the general expression of  $M_t^\pm$  given above, one gets

$$\frac{M_{t+1}}{M_t} \leq \frac{(t+1)B(t+1, \alpha+1)\chi_+(t+1) + v_0^{t+1}}{tB(t, \alpha+1)\chi_-(t) - \gamma v_0^t},$$

where

$$\chi_\pm(t) = \sum_{i=0}^{n-1} \gamma_i \Lambda(\varepsilon_i, t) P(\varepsilon_i, t) \pm \gamma_n \Lambda(\varepsilon_n, t) P(\varepsilon_n, t).$$

Using the fact that

$$\frac{(t+1)B(t+1, \alpha+1)}{tB(t, \alpha+1)} = \frac{t+1}{t+\alpha+1},$$

and that

$$tB(t, \alpha+1)Q_\alpha(t) \geq 1, \quad \text{where } Q_\alpha(t) = \frac{(t+\alpha)(t+\alpha-1)\dots(t+\{\alpha\})}{\Gamma(\alpha+1)},$$

one sees that

$$\frac{M_{t+1}}{M_t} \leq \frac{t+1}{t+\alpha+1} \frac{\gamma_0 + \chi_+(t+1) + Q_\alpha(t+1)v_0^{t+1}}{\gamma_0 + \chi_-(t) - \gamma Q_\alpha(t)v_0^t}.$$

Furthermore,

$$\frac{\gamma_0 + \chi_+(t+1) + Q_\alpha(t+1)v_0^{t+1}}{\gamma_0 + \chi_-(t) - Q_\alpha(t)v_0^t} = 1 + \frac{\chi_+(t+1) - \chi_-(t) + \kappa(t)v_0^t}{\gamma_0 + \chi_-(t) - \gamma Q_\alpha(t)v_0^t}.$$

where  $\kappa(t) = v_0 Q_\alpha(t+1) + \gamma Q_\alpha(t)$  is a polynomial function in  $t$ .

From lemma 7.3 below, there exists a positive number  $C$  which depend on  $\alpha$  and  $\varepsilon_{0,n}$  only such that

$$\chi_+(t+1) - \chi_-(t) \leq [2\gamma_n + \varepsilon_n \gamma] C t^{-\beta}, \quad \chi_-(t) \geq -C \gamma t^{-\varepsilon_1}.$$

where

$$\beta = \min\{\varepsilon_n, 1 + \varepsilon_1\}, \quad 1 < \beta \leq 2.$$

Combining these estimates on  $\chi_+(t+1)$  and  $\chi_-(t)$ , one sees that there exists finite continuous functions  $\tau_1$  and  $A$  of  $(\gamma_{0:n}, \alpha, \varepsilon_{0:n})$ , such that, for every  $t \geq \tau_1$ ,

$$R_t \geq \alpha/t - A/t^\beta.$$

Choosing  $\tau = \max(\tau_1, \tau_2)$  yields proposition 7.2 as soon as, for every  $t \geq \tau$  (recall that  $\beta > 1$ ),

$$2At \leq \alpha t^\beta. \quad \square$$

**Lemma 7.3.** *Let  $\beta = \min\{\varepsilon_n, 1 + \varepsilon_1\}$ . There exists a positive number  $C$ , which depends on  $\alpha$  and  $\varepsilon_{0:n}$  only, such that*

$$\chi_+(t+1) - \chi_-(t) \leq [2\gamma_n + \varepsilon_n\gamma]Ct^{-\beta}, \quad \chi_-(t) \geq -C\gamma t^{-\varepsilon_1}.$$

*Proof of lemma 7.3.* For every real number  $t \geq 1$  and every  $1 \leq i \leq n$ ,

$$e^{-S(\varepsilon_i, t) - T(\varepsilon_i, t)} \leq P(\varepsilon_i, t) \leq e^{-S(\varepsilon_i, t)},$$

where

$$S(\varepsilon, t) = \sum_{\ell=1}^{[t]+1} \frac{\varepsilon}{\alpha + \varepsilon + \{t\} + \ell} \quad \text{and} \quad T(\varepsilon, t) = \sum_{\ell=1}^{[t]+1} \frac{\varepsilon^2}{(\alpha + \varepsilon + \{t\} + \ell)^2}.$$

Thus, there exists two positive real numbers  $C_i^-$  and  $C_i^+$  such that for every real number  $t \geq 1$ ,  $C_i^- \leq t^{\varepsilon_i} P(\varepsilon_i, t) \leq C_i^+$ , and one can choose  $C_i^+ = (\alpha + \varepsilon_i + 3)^{\varepsilon_i}$ .

Let  $C = \max\{C_i^+; 1 \leq i \leq n\}$ . Using the two relations

$$P(\varepsilon_i, t) - P(\varepsilon_i, t+1) = P(\varepsilon_i, t) \frac{\varepsilon_i}{\alpha + \varepsilon_i + t + 2},$$

and

$$P(\varepsilon_n, t) + P(\varepsilon_n, t+1) = P(\varepsilon_n, t) \left( 2 - \frac{\varepsilon_n}{\alpha + \varepsilon_n + t + 2} \right),$$

one sees that

$$\chi_+(t+1) - \chi_-(t) = 2\gamma_n \Lambda(\varepsilon_n, t) P(\varepsilon_n, t) - \sum_{i=1}^n \gamma_i \Lambda(\varepsilon_i, t) P(\varepsilon_i, t) \frac{\varepsilon_i}{\alpha + \varepsilon_i + t + 2}.$$

For every  $1 \leq i \leq n$ , the function  $\Lambda(\varepsilon_i, \cdot)$  is positive and bounded by 1. Hence,

$$\begin{aligned} \chi_+(t+1) - \chi_-(t) &\leq 2\gamma_n P(\varepsilon_n, t) + \sum_{i=1}^n |\gamma_i| P(\varepsilon_i, t) \frac{\varepsilon_i}{\alpha + \varepsilon_i + t + 2} \\ &\leq C \left( 2\gamma_n t^{-\varepsilon_n} + \gamma \varepsilon_n t^{-(1+\varepsilon_1)} \right), \end{aligned}$$

and the first inequality in the statement of the lemma holds. The same kind of estimates yields

$$\chi_-(t) \geq - \sum_{i=0}^{n-1} |\gamma_i| \Lambda(\varepsilon_i, t) P(\varepsilon_i, t) - \gamma_n \Lambda(\varepsilon_n, t) P(\varepsilon_n, t),$$

hence the second inequality holds. This concludes the proof of lemma 7.3.  $\square$

#### ACKNOWLEDGMENTS

I would like to thank Mike Steel for some helpful comments.

#### REFERENCES

- [1] M.T. Holder, P. Lewis, and K.E. Holsinger. Polytomies and bayesian phylogenetic inference. *Systematic Biology*, 54(2):241–253, 2005.
- [2] B. Kolaczowski and J.W. Thornton. Is there a star tree paradox? *Mol. Biol. Evol.*, 23:1819–1823, 2006.
- [3] M. Steel and F. A. Matsen. The Bayesian ‘star paradox’ persists for long finite sequences. *Molecular Biology and Evolution*, 24:1075–1079, 2007.
- [4] E. Susko. On the distributions of bootstrap support and posterior distributions for a star tree. *Systematic Biology*, 57(4):602–612, 2008.
- [5] Z. Yang. *Computational Molecular Evolution*. Oxford Series in Ecology and Evolution, 2006.

- [6] Z. Yang. Fair-Balance paradox, star-tree paradox, and bayesian phylogenetics. *Molecular Biology and Evolution*, 24:1639–1655, 2007.
- [7] Z. Yang and B. Rannala. Branch-length prior influences Bayesian posterior probability of phylogeny. *Syst. Biol.*, 54(3):455–470, 2005.

UNIVERSITÉ JOSEPH FOURIER GRENOBLE 1, INSTITUT FOURIER UMR 5582 UJF-CNRS, 100 RUE DES MATHS, BP 74, 38402 SAINT MARTIN D'HÈRES, FRANCE