Bernoulli **22**(3), 2016, 1709–1728
DOI: 10.3150/15-BEJ708

Approximation of improper priors

CHRISTELE BIOCHE* and PIERRE DRUILHET**

[1]*Laboratoire de Mathématiques, UMR CNRS 6620, Clermont Université, Université Blaise Pascal, 63177 Aubière Cedex, France.*
E-mail: *Christele.Bioche@math.univ-bpclermont.fr, **Pierre.Druilhet@univ-bpclermont.fr

We propose a convergence mode for positive Radon measures which allows a sequence of probability measures to have an improper limiting measure. We define a sequence of vague priors as a sequence of probability measures that converges to an improper prior. We consider some cases where vague priors have necessarily large variances and other cases where they have not. We study the consequences of the convergence of prior distributions on the posterior analysis. Then we give some constructions of vague priors that approximate the Haar measures or the Jeffreys priors. We also revisit the Jeffreys–Lindley paradox.

Keywords: approximation of improper priors; conjugate priors; convergence of prior; Jeffreys–Lindley paradox; non-informative priors; the Jeffreys prior; vague priors

1. Introduction

Improper priors such as flat priors (Laplace [21]), Jeffreys priors (Jeffreys [17]), reference priors (Berger *et al.* [4]) or the Haar measures (Eaton [13]) are often used in Bayesian analysis when no prior information is available. The posterior distribution is obtained by applying the formal Bayes rule. There are several approaches to justify the use of improper priors in statistics. Taraldsen and Lindqvist [26] explain how the theory of conditional probability spaces developed by Rényi [23] is related to a theory for statistics that includes improper priors. Their article is based on a generalization of Kolmogorov's theory to the σ-finite measures. They show in particular by examples that this theory is different from the alternative theory of improper priors provided by Hartigan [14]. For many authors, the inference based on an improper prior Π is legitimated as limit of inferences based on proper priors Π_n. However, there are several ways to define this limit. For example, Jeffreys [18], Stone [25], Bernardo and Smith ([6], Proposition 5.11), Jaynes [16] consider the convergence, for any given observation x, of the posterior distributions $\Pi_n(\cdot|x)$ to $\Pi(\cdot|x)$ for some convergence mode such as total variation. Stone [24] consider a convergence mode involving both the posterior distribution and the marginal distribution.

All these convergence modes are related to the statistical model through the likelihood. Moreover, there is no standard convergence mode such that a sequence Π_n of proper priors may converge to an improper prior Π independently on the statistical model. Consider, for example, a sequence of normal distributions $\mathcal{N}(0,n)$ with zero mean and variance equal to n; it is often admitted that this sequence converges to the Laplace prior since for many statistical models the Bayes estimate related to $\mathcal{N}(0,n)$ converges to the Bayes estimate for the Laplace prior. A question then arises: does the limiting behaviour of a sequence of proper priors depend on the statistical model? Is there any intrinsic convergence mode?

1350-7265

The aim of this paper is to define a convergence mode on the set of prior distributions without reference to any statistical model. In Section 2, we define this convergence mode. We show that a sequence of vague priors is related to at most one improper prior. We also show that any improper distribution can be approximated by proper distributions and reciprocally. In Section 3, we give some conditions on the likelihood to derive convergence of posterior distributions and Bayesian estimators from the convergence of prior distributions. In Section 4, we give some examples of construction of sequences of probability measures which converge to improper priors such as the Haar measure or the Jeffreys prior. In Section 6, we give a special interest in the convergence of Beta distributions. In Section 7, we revisit the Jeffreys–Lindley paradox in the light of our convergence mode.

2. Definition, properties and examples of q-vague convergence

Let X be a random variable and assume that $X|\theta \sim P_\theta, \theta \in \Theta$. We assume that Θ is in $\mathbb{R}$, $\mathbb{R}^p$ with $p > 1$, or a countable set. In the Bayesian paradigm, a prior distribution Π is given on Θ. In this article, we always assume that a prior Π is a positive Radon measure, that is, a positive measure which is finite on compact sets. So, a prior may be proper or improper. We denote by π the density function with respect to the Lebesgue measure in the continuous case and the counting measure in the discrete case, or more generally to some σ-finite measure. If Π is a probability measure, we can use the Bayes formula to write the posterior density:

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_\Theta f(x|\theta)\pi(\theta)\,\mathrm{d}\theta}, \tag{1}$$

where $f(x|\theta)$ is the likelihood function.

If Π is an improper measure but $\int_\Theta f(x|\theta)\pi(\theta)\,\mathrm{d}\theta < +\infty$, we can formally apply the Bayes formula to get a posterior distribution which will be proper. Now, if we replace Π by $\alpha\Pi$, for $\alpha > 0$, we obtain the same posterior distribution. So, in this case, the posterior distribution is proper and independent of changes in the scaling of the prior. If Π is an improper measure with $\int_\Theta f(x|\theta)\pi(\theta)\,\mathrm{d}\theta = +\infty$, we cannot apply the Bayes formula. But in this article, we allow posterior distribution to be improper, and in this case we will define it by $\pi(\theta|x) = f(x|\theta)\pi(\theta)$ up to within a scalar factor.

We denote by $\mathcal{C}_K(\Theta)$ the space of real-valued continuous functions on Θ with compact support and by $\mathcal{C}_K^+(\Theta)$ the positive functions in $\mathcal{C}_K(\Theta)$. When there is no ambiguity on the space, they will be simply denoted by $\mathcal{C}_K$ or $\mathcal{C}_K^+$. We also introduce the notation $\mathcal{C}_b(\Theta)$ for the space of bounded continuous functions on Θ, and $\mathcal{C}_0(\Theta)$ for the space of continuous functions g such that for all $\varepsilon > 0$, there exists a compact $K \subset \Theta$ such that for all $\theta \in K^c$, $g(\theta) < \varepsilon$. We use the notation $\Pi(h) = \int_\Theta h\,\mathrm{d}\Pi$ where h is a measurable real-valued function, and $|\Pi| = \Pi(1) = \int_\Theta \mathrm{d}\Pi$, the total mass of Π.

We recall the two classic kinds of convergence of measures (Bauer [2]). A sequence of probability measures $\{\Pi_n\}_n$ converges narrowly (also said weakly) to a probability measure Π if, for every function ϕ in $\mathcal{C}_b(\Theta)$, $\{\Pi_n(\phi)\}_n$ converges to $\Pi(\phi)$. A sequence of positive Radon measures $\{\Pi_n\}_n$ converges vaguely to a positive Radon measure Π if, for every function ϕ in

$\mathcal{C}_K(\Theta)$, $\{\Pi_n(\phi)\}_n$ converges to $\Pi(\phi)$. We also recall a characterization of vague convergence for a sequence of probability measures which will be useful later in the article.

Lemma 2.1 (Billingsley [8], page 393). *If $\{\Pi_n\}_n$ is a sequence of probability measures and Π is a probability measure, then $\{\Pi_n\}_n$ converges vaguely to Π iff for all $g \in \mathcal{C}_0(\Theta)$, $\{\Pi_n(g)\}_n$ converges to $\Pi(g)$.*

2.1. Convergence of prior distribution sequences

In this section, we define a new convergence mode for sequences of positive Radon measures. The aim is to propose a formalization of an usual practice which consists of approximate an improper prior with a sequence of proper priors.

Definition 2.2. *A sequence of positive Radon measures $\{\Pi_n\}_n$ is said to converge q-vaguely to a positive Radon measure Π if there exists a sequence of positive real numbers $\{a_n\}_n$ such that $\{a_n\Pi_n\}_n$ converges vaguely to Π.*

Let us justify this definition. In equation (1), if we replace Π by $\alpha\Pi$, for $\alpha > 0$, we obtain the same posterior distribution, which means that the prior distribution is defined up to within a scalar factor. So, it is natural to define the equivalence relation $\sim$ on the space of positive Radon measures by

$$\Pi \sim \Pi' \quad \Longleftrightarrow \quad \exists\alpha > 0 \qquad \text{such that } \Pi = \alpha\Pi'. \tag{2}$$

Then it is natural to define the quotient space of positive Radon measures by the equivalence relation $\sim$. We denote by $\overline{\Pi}$ the equivalence class of Π, that is, $\overline{\Pi} = \{\widetilde{\Pi}/\exists\alpha > 0, \widetilde{\Pi} = \alpha\Pi\}$. The q-vague convergence corresponds to the standard quotient topology on this quotient space.

Remark 2.3. One referee pointed out that similar quotient spaces for σ-finite measures were considered by Taraldsen and Lindqvist [27] to define conditional measures.

Proposition 2.4. *Let $\{\Pi_n\}_n$ and Π be positive Radon measures. The sequence $\{\Pi_n\}_n$ converges q-vaguely to Π iff $\{\overline{\Pi}_n\}_n$ converges to $\overline{\Pi}$ for the quotient topology.*

Proof. • Direct part: Assume that $\lim_{n\to\infty}\overline{\Pi}_n = \overline{\Pi}$. The space of positive Radon measures is a metrisable space so it admits a countable neighbourhood base. Thus, there exists a decreasing sequence of open sets $\{O_i\}_{i\in\mathbb{N}}$ in the space of positive Radon measures such that for all $i \in \mathbb{N}$, $\Pi \in O_i$ and $\bigcap_{i\in\mathbb{N}} O_i = \{\Pi\}$. So, for all $i \in \mathbb{N}$, $\overline{\Pi} \in \overline{O}_i$. For any O_i, there exists N_i such that for all $n > N_i$, $\overline{\Pi}_n \in \overline{O}_i$. Without lost of generality, we can choose N_i such that $N_i > N_{i-1}$. For all n such that $N_i \leq n < N_{i+1}$, $\Pi_n \in \mathcal{C}(O_i)$ where $\mathcal{C}(O_i) = \{\lambda x \text{ with } \lambda > 0 \text{ and } x \in O_i\}$, that is, for all n such that $N_i \leq n < N_{i+1}$, there exists $a_n > 0$ such that $a_n\Pi_n \in O_i$. Moreover, since $\bigcap_{i\in\mathbb{N}} O_i = \{\Pi\}$, $\lim_{n\to\infty} a_n\Pi_n = \Pi$.

• Converse part: Assume that $\{a_n \Pi_n\}_n$ converges to Π. Since the canonical mapping ϕ defined by

$$\phi : \mathcal{R} \to \mathcal{R}/\sim, \qquad \Pi \mapsto \overline{\Pi}, \tag{3}$$

where $\mathcal{R}$ is the space of positive Radon measures, is continuous, $\{\phi(a_n\Pi_n)\} = \{\overline{\Pi}_n\}$ converges to $\phi(\Pi) = \overline{\Pi}$. □

The following proposition shows that a sequence of prior measures cannot converge q-vaguely to more than one limit up to within a scalar factor.

Theorem 2.5. *Let $\{\Pi_n\}_n$ be a sequence of priors such that $\{\Pi_n\}_n$ converges q-vaguely to both Π_a and Π_b, then necessarily there exists $\alpha > 0$ such that $\Pi_a = \alpha\Pi_b$.*

Proof. This is a direct consequence of Proposition A.1 that states that $\overline{\mathcal{R}}$ is a Hausdorff space. However, we give here a direct proof that does not involve abstract topological concept.

Assume that $\{\Pi_n\}_n$ converges q-vaguely to both Π_a and Π_b. From Definition 2.2, there exist two sequences of positive scalars $\{a_n\}_n$ and $\{b_n\}_n$ such that $\{a_n\Pi_n\}_n$, respectively, $\{b_n\Pi_n\}_n$, converges vaguely to Π_a, respectively, Π_b. We have to prove that $\Pi_b = \alpha\Pi_a$ for some positive scalar α. Since $\Pi_a \neq 0$ and $\Pi_b \neq 0$, there exist h_a and h_b in $\mathcal{C}_K^+$ such that $\Pi_a(h_a) > 0$ and $\Pi_b(h_b) > 0$. Put $h_0 = h_a + h_b$; we have $\Pi_a(h_0) > 0$ and $\Pi_b(h_0) > 0$. Moreover, $\lim_{n\to\infty} a_n\Pi_n(h_0) = \Pi_a(h_0)$ and $\lim_{n\to\infty} b_n\Pi_n(h_0) = \Pi_b(h_0)$. So, there exists N such that for $n \geq N$, $a_n\Pi_n(h_0) > 0$ and $b_n\Pi_n(h_0) > 0$. For any h in $\mathcal{C}_K$ and $n > N$, $\lim_{n\to\infty} \frac{\Pi_n(h)}{\Pi_n(h_0)} = \lim_{n\to\infty} \frac{a_n\Pi_n(h)}{a_n\Pi_n(h_0)} = \frac{\Pi_a(h)}{\Pi_a(h_0)}$ and $\lim_{n\to\infty} \frac{\Pi_n(h)}{\Pi_n(h_0)} = \lim_{n\to\infty} \frac{b_n\Pi_n(h)}{b_n\Pi_n(h_0)} = \frac{\Pi_b(h)}{\Pi_b(h_0)}$. By uniqueness of the limit in $\mathbb{R}$, $\frac{\Pi_a(h)}{\Pi_a(h_0)} = \frac{\Pi_b(h)}{\Pi_b(h_0)}$. Therefore, $\Pi_a = \frac{\Pi_a(h_0)}{\Pi_b(h_0)}\Pi_b$. The result follows. □

Theorem 2.6 motivates to include the improper priors in the theory since it shows these are obtained naturally from limits of proper priors. This can be compared with a completion of a metric space.

Theorem 2.6. *Any improper measure may be approximated by a sequence of probability measures and conversely, any proper measure may be approximated by a sequence of improper measures.*

Proof. • Consider an improper measure Π and $\{K_n\}_n$ an increasing sequence of compacts such that $\Theta = \bigcup_n K_n$. Then $\Pi_n = \Pi\mathbb{1}_{K_n}$ is a proper measure so, $\frac{1}{|\Pi_n|}\Pi_n$ is a probability measure. Moreover, $\{\Pi_n\}_n$ converges vaguely to Π, so $\{\frac{1}{|\Pi_n|}\Pi_n\}$ converges q-vaguely to Π.

• Let Π be a probability measure. Consider the sequence $\Pi_n = \Pi + \alpha_n\Pi'$ where Π' is an improper measure and $\{\alpha_n\}_n$ is a decreasing sequence which converges to 0. Then, for all $n \in \mathbb{N}$, Π_n is an improper measure and $\{\Pi_n\}_n$ converges q-vaguely to Π. □

In many statistical models, there are several parameterizations of interest. We show that the q-vague convergence is invariant by change of parameterization. Consider a new parameterization $\eta = h(\theta)$ where h is a homeomorphism. We denote by $\widetilde{\Pi}_n = \Pi_n \circ h^{-1}$ and $\widetilde{\Pi} = \Pi \circ h^{-1}$ the prior distribution on η derived from the prior distribution on θ. The following proposition establishes a link between q-vague convergence of $\{\Pi_n\}_n$ and $\{\widetilde{\Pi}_n\}_n$.

Proposition 2.7. *Let $\{\Pi_n\}_n$ be a sequence of priors which converges q-vaguely to Π. Let h be a homeomorphism and consider the parameterization $\eta = h(\theta)$. Then $\{\widetilde{\Pi}_n\}_n$ converges q-vaguely to $\widetilde{\Pi}$.*

Proof. From the change of variables formula, $\int g(h(\theta))\,\mathrm{d}\Pi_n(\theta) = \int g(\eta)\,\mathrm{d}\widetilde{\Pi}_n(\eta)$ and $\int g(h(\theta))\,\mathrm{d}\Pi(\theta) = \int g(\eta)\,\mathrm{d}\widetilde{\Pi}(\eta)$. Moreover, if $\{\Pi_n\}_n$ converges q-vaguely to Π, from Definition 2.2 there exists $\{a_n\}_n$ such that $\{a_n\Pi_n\}_n$ converges vaguely to Π. Note that for all $g \in \mathcal{C}_K$, $g \circ h \in \mathcal{C}_K$. So, for all $g \in \mathcal{C}_K$, $\lim_{n\to\infty} a_n \int g(h(\theta))\,\mathrm{d}\Pi_n(\theta) = \int g(h(\theta))\,\mathrm{d}\Pi(\theta)$, that is, $\lim_{n\to\infty} a_n \int g(\eta)\,\mathrm{d}\widetilde{\Pi}_n(\eta) = \int g(\eta)\,\mathrm{d}\widetilde{\Pi}(\eta)$. Thus, $\{\widetilde{\Pi}_n\}_n$ converges q-vaguely to $\widetilde{\Pi}$. □

2.2. Convergence when approximants are probabilities

In this section, the sequence of approximants $\{\Pi_n\}_n$ is assumed to be a sequence of probability measures. Then we can establish some links between q-vague and narrow convergence.

Indeed, if $\{\Pi_n\}_n$ is a sequence of probabilities and Θ is a compact set, q-vague convergence is equivalent to narrow convergence.

More generally, we give a necessary and sufficient condition for the narrow convergence of a sequence of probabilities which converges q-vaguely to a probability. We recall that a sequence of bounded measures $\{\Pi_n\}_n$ is said to be tight if, for each $\varepsilon > 0$, there exists a compact set K such that, for all n, $\Pi_n(K^c) < \varepsilon$.

Proposition 2.8. *Let $\{\Pi_n\}_n$ and Π be probability measures such that $\{\Pi_n\}_n$ converges q-vaguely to Π. Then $\{\Pi_n\}_n$ converges narrowly to Π iff $\{\Pi_n\}_n$ is tight.*

Proof. Direct part: $\{\Pi_n\}_n$ converges narrowly to Π a probability measure so $\{\Pi_n\}_n$ is tight.

Converse part: Let us show that if $\{\Pi_{n_k}\}_k$ is any subsequence of $\{\Pi_n\}_n$ which converges narrowly then $\{\Pi_{n_k}\}_k$ converges to Π. From Billingsley ([8], Theorem 25.10), there exists a subsequence $\{\Pi_{n_k}\}_k$ of $\{\Pi_n\}_n$ which converges narrowly to some probability measure, say $\widetilde{\Pi}$. Since $\{\Pi_{n_k}\}_k$ is a sequence of probabilities which converges narrowly to $\widetilde{\Pi}$, from Definition 2.2, $\{\Pi_{n_k}\}_k$ converges q-vaguely to $\widetilde{\Pi}$. So, from Theorem 2.5, there exists $\alpha > 0$ such that $\Pi = \alpha\widetilde{\Pi}$, but Π and $\widetilde{\Pi}$ are probabilities. So $\Pi = \widetilde{\Pi}$. The result follows from Billingsley ([8], Corollary of Theorem 25.10, page 346). □

Now, we also assume that the limiting measure Π is an improper measure. Then we can give a result about the sequence $\{a_n\}_n$ which will be useful thereafter.

Lemma 2.9. *Let* $\{\Pi_n\}_n$ *be a sequence of probability measures and* $\{a_n\}_n$ *a sequence of positive scalars such that* $\{a_n\Pi_n\}_n$ *converges vaguely to* Π. *If* Π *is improper, then necessarily* $\lim_{n\to\infty} a_n = +\infty$.

Proof. We assume that $\{a_n\Pi_n\}_n$ converges vaguely to Π so, we have $\Pi(\Theta) \leq \liminf_n a_n\Pi_n(\Theta)$ (see Bauer [2], Theorem 30.3). But for all $n \in \mathbb{N}$, $\Pi_n(\Theta) = 1$ so $\Pi(\Theta) \leq \liminf a_n$. Moreover, $\Pi(\Theta) = +\infty$ so $\liminf_n a_n = +\infty$. The result follows. □

Lemma 2.10 (Lang [20], page 38). *Let* E *be* $\mathbb{R}$, $\mathbb{R}^p$ *with* $p > 1$ *or a countable set, for all compact* $K_0 \subset (\bigcup_{n>0} \mathring{K}_n) = E$, *there exists a function* $h \in \mathcal{C}_K(E)$ *such that* $\mathbb{1}_{K_0} \leq h \leq 1$.

When a sequence of proper priors is used to approximate an improper prior, the mass tends to concentrate outside any compact set.

Proposition 2.11. *Let* $\{\Pi_n\}_n$ *be a sequence of probability measures which converges* q*-vaguely to an improper prior* Π. *Then, for any compact* K *in* Θ, $\lim_{n\to\infty} \Pi_n(K) = 0$, *and consequently,* $\lim_{n\to\infty} \Pi_n(K^c) = 1$.

Proof. From Definition 2.2, there exists $\{a_n\}_n$ such that $\lim_{n\to\infty} a_n\Pi_n(h) = \Pi(h)$ for any h in $\mathcal{C}_K$. From Lemma 2.9, $\lim_{n\to\infty} a_n = +\infty$ whereas $\Pi(h) < +\infty$, so $\lim_{n\to\infty} \Pi_n(h) = 0$. Let K_0 be a compact set in Θ. From Lemma 2.10, there exists a function $h \in \mathcal{C}_K$ such that $\mathbb{1}_{K_0} \leq h$. So $\Pi_n(K_0) \leq \Pi_n(h)$ and $\lim_{n\to\infty} \Pi_n(K_0) = 0$. Since $\Pi_n(K_0) + \Pi_n(K_0^c) = 1$ for all $n \in \mathbb{N}$, thus $\lim_{n\to\infty} \Pi_n(K_0^c) = 1$. □

Many authors consider that few knowledge on the parameter is represented by priors with large variance. Here, we establish some links between the q-vague convergence of priors and the convergence of the sequence of corresponding variances.

Proposition 2.12. *Let* $\{\Pi_n\}_n$ *be a sequence of probabilities on* $\mathbb{R}$ *such that* $\mathbb{E}_{\Pi_n}(\theta)$ *is a constant. If* $\{\Pi_n\}_n$ *converges* q*-vaguely to an improper prior* Π *whose support is* $\mathbb{R}$, *then* $\lim_{n\to\infty} \mathrm{Var}_{\Pi_n}(\theta) = +\infty$.

Proof. Since $\mathbb{E}_{\Pi_n}(\theta)$ is constant, $\lim_{n\to\infty} \mathrm{Var}_{\Pi_n}(\theta) = +\infty$ iff $\lim_{n\to\infty} \mathbb{E}_{\Pi_n}(\theta^2) = +\infty$. For any $r > 0$, we have $\mathbb{E}_{\Pi_n}(\theta^2) \geq \int_{[-r,r]^c} \theta^2 \, d\Pi_n(\theta)$ so $\mathbb{E}_{\Pi_n}(\theta^2) \geq r^2\Pi_n([-r, r]^c)$. From Proposition 2.11, $\lim_{n\to\infty} \Pi_n([-r, r]^c) = 1$ and then $\lim_{n\to\infty} \mathbb{E}_{\Pi_n}(\theta^2) \geq r^2$. Since this holds for any $r > 0$, $\lim_{n\to\infty} \mathbb{E}_{\Pi_n}(\theta^2) = +\infty$. □

Corollary 2.13. *Let* $\{\Pi_n\}_n$ *be a sequence of probabilities with constant mean which approximate the Lebesgue measure* $\lambda_{\mathbb{R}}$. *Then, necessarily,* $\lim_{n\to\infty} \mathrm{Var}_{\Pi_n}(\theta) = +\infty$.

However, we will see in the examples in Section 5.4.1 that when we do not assume the expectation to be constant; the variance does not necessarily diverge.

2.3. Characterization of q-vague convergence

In this section, we establish several sufficient conditions for the q-vague convergence of $\{\Pi_n\}_n$ to Π through their probability density function (p.d.f.). When Θ is continuous, then π_n and π are the standard p.d.f. with respect to the Lebesgue measure. When Θ is discrete, then $\pi(\theta_0) = \Pi(\theta = \theta_0)$, that is, π is the p.d.f. with respect to the counting measure.

When $\Theta = \{\theta_i\}_{i\in I}$ is a discrete set with $I \subset \mathbb{N}$, we give an easy-to-check characterization of the q-vague convergence.

Proposition 2.14. *Let $\{\Pi_n\}_n$ and Π be priors on $\Theta = \{\theta_i\}_{i\in I}$, $I \subset \mathbb{N}$. The sequence $\{\Pi_n\}_n$ converges q-vaguely to Π iff there exists a sequence of positive real numbers $\{a_n\}_n$ such that for all $i \in I$, $\lim_{n\to\infty} a_n\pi_n(\theta_i) = \pi(\theta_i)$.*

Proof. It is a direct consequence of Definition 2.2 applied to the discreet case. □

Now, we consider the continuous case.

Proposition 2.15. *Let $\{\Pi_n\}_n$ and Π be continuous priors on Θ in $\mathbb{R}$ or $\mathbb{R}^p$ with $p > 1$. Assume that*:

(1) *there exists a sequence of positive real numbers $\{a_n\}_n$ such that the sequence $\{a_n\pi_n\}_n$ converges pointwise to π*,

(2) *there exists a continuous function $g : \Theta \to \mathbb{R}^+$ and $N \in \mathbb{N}$ such that for all $n > N$ and $\theta \in \Theta$, $a_n\pi_n(\theta) < g(\theta)$.*

Then $\{\Pi_n\}_n$ converges q-vaguely to Π.

Proof. Let h be in $\mathcal{C}_K(\Theta)$. Then, $a_n h(\theta)\pi_n(\theta) \le \|h\| g \mathbb{1}_K(\theta)$ where $\|h\| = \max_{\theta\in\Theta} h(\theta)$. Since $\|h\| g \mathbb{1}_K(\theta)$ is Lebesgue integrable, by dominated convergence theorem, $\lim_{n\to\infty} \int a_n\pi_n(\theta) \times h(\theta)\,\mathrm{d}\theta = \int \pi(\theta)h(\theta)\,\mathrm{d}\theta$. □

The following result will be useful to establish a result in Section 4.2.

Proposition 2.16. *Let $\{\Pi_n\}_n$ and Π be priors. Assume that*:

(1) *there exists a sequence of positive real numbers $\{a_n\}_n$ such that the sequence $\{a_n\pi_n\}_n$ converges pointwise to π*,

(2′) *for any compact set K, there exists a scalar M and some $N \in \mathbb{N}$ such that for $n > N$, $\sup_{\theta\in K} a_n\pi_n(\theta) < M$.*

Then $\{\Pi_n\}_n$ converges q-vaguely to Π.

Proof. The proof is similar to the proof of Proposition 2.15 with $a_n\pi_n(\theta)h(\theta) \le M \sup_{\theta\in K} |h(\theta)| \mathbb{1}_K(\theta)$. □

Remark 2.17. Proposition 2.15 and Proposition 2.16 hold if $\pi(\theta)$ is the p.d.f. with respect to any positive Radon measure.

3. Convergence of posterior distributions and estimators

Consider the model $X|\theta \sim P_\theta$, $\theta \in \Theta$. We denote by $f(x|\theta)$ the likelihood. The priors Π_n on Θ represent our prior knowledge. We always assume that $\int_\Theta f(x|\theta)\,\mathrm{d}\Pi(\theta) > 0$.

For a measure Π and a measurable function g, we define the measure $g\Pi$ by $g\Pi(f) = \Pi(gf) = \int f(\theta)g(\theta)\,\mathrm{d}\Pi(\theta)$ for any f whenever the integrals are defined; $g\Pi$ is also denoted $g\,\mathrm{d}\Pi$ or $\Pi \circ g^{-1}$ by some authors.

In this paper, we define the posterior on θ, $\Pi(\cdot|x)$, by $\pi(\theta|x) \propto f(x|\theta)\pi(\theta)$. Thus, the posterior $\Pi(\cdot|x)$ may be proper or improper. There are three possible cases. First, if we use a proper prior, by applying the Bayes formula, we obtain a posterior which is a probability measure. If the prior is an improper measure such that $\int_\Theta f(x|\theta)\pi(\theta)\,\mathrm{d}\theta < +\infty$, we can formally apply the Bayes rule, which provides a posterior probability measure by renormalization. At last, if the prior is an improper measure such that $\int_\Theta f(x|\theta)\pi(\theta)\,\mathrm{d}\theta = +\infty$, the posterior is an improper measure defined by $\pi(\theta|x) = f(x|\theta)\pi(\theta)$ up to within a scalar factor.

In this section, we study the consequences of the q-vague convergence of $\{\Pi_n\}_n$ on the posterior analysis. In the general case where the posteriors may be proper or improper, we give a result about the q-vague convergence of posteriors $\{\Pi_n(\cdot|x)\}_n$ to $\Pi(\cdot|x)$. When posteriors are probability measures, we can establish results about the narrow convergence instead of the q-vague convergence.

Proposition 3.1. *Let $\{\Pi_n\}_n$ be a sequence of priors which converges q-vaguely to Π. Assume that, $\theta \longmapsto f(x|\theta)$ is a non-zero continuous function on Θ. Then $\{\Pi_n(\cdot|x)\}_n$ converges q-vaguely to $\Pi(\cdot|x)$.*

Moreover, if $\{\Pi_n(\cdot|x)\}_n$ is a tight sequence of probabilities and $\Pi(\cdot|x)$ is a probability, then $\{\Pi_n(\cdot|x)\}_n$ converges narrowly to $\Pi(\cdot|x)$.

Proof. Assume that $\{\Pi_n\}_n$ converges q-vaguely to Π. From Definition 2.2, there exists a sequence of positive scalars $\{a_n\}_n$ such that $\{a_n\Pi_n\}_n$ converges vaguely to Π. So, for any $h \in \mathcal{C}_K$, $\lim_{n\to\infty} a_n\Pi_n(h) = \Pi(h)$. Since $f(x|\cdot)$ is a continuous function, $f(x|\cdot)h \in \mathcal{C}_K$ and $\lim_{n\to\infty} a_n\Pi_n(f(x|\cdot)h) = \Pi(f(x|\cdot)h)$. But $\Pi_n(f(x|\cdot)h) = f(x|\cdot)\Pi_n(h)$ and $\Pi(f(x|\cdot)h) = f(x|\cdot)\Pi(h)$. So, $\{a_n f(x|\cdot)\Pi_n\}$ converges vaguely to $f(x|\cdot)\Pi$, or equivalently $\{f(x|\cdot)\Pi_n\}_n$ converges q-vaguely to $f(x|\cdot)\Pi$.

If $\{\Pi_n(\cdot|x)\}_n$ is a tight sequence of probabilities and $\Pi(\cdot|x)$ is a probability, the second result follows from Proposition 2.8. □

Remark 3.2. If Θ is discrete, then $f(x|\theta)$ is necessary continuous for the discrete topology.

The following results are based on Proposition 3.1 with easier-to-check assumptions.

Corollary 3.3. *Let $\{\Pi_n\}_n$ and Π be priors. Assume that*:

(1) *there exists a sequence of positive real numbers $\{a_n\}_n$ such that the sequence $\{a_n\pi_n\}_n$ converges pointwise to π*,

(2) *$\{a_n\pi_n(\theta)\}_n$ is non-decreasing for all $\theta \in \Theta$*,

(3) $\theta \longmapsto f(x|\theta)$ *is continuous and positive*,
(4) *all the posteriors* $\Pi_n(\cdot|x)$ *and* $\Pi(\cdot|x)$ *are proper.*

Then, $\{\Pi_n(\cdot|x)\}_n$ *converges narrowly to* $\Pi(\cdot|x)$.

Proof. The sequence $\{a_n f\pi_n\}_n$ is a non-decreasing sequence of non-negative functions. By monotone convergence theorem, $\lim_{n\to\infty}\int a_n f(x|\theta)\pi_n(\theta)\,\mathrm{d}\theta = \int \lim_{n\to\infty} a_n f(x|\theta) \times \pi_n(\theta)\,\mathrm{d}\theta = \int f(x|\theta)\pi(\theta)\,\mathrm{d}\theta$. So, $\{a_n\Pi_n(f)\}_n$ converges to $\Pi(f) > 0$. So there exists N such that for all $n > N$, $a_n\Pi_n(f) \geq \frac{1}{2}\Pi(f)$. Consider $\{K_m\}_m$ an increasing sequence of compact sets such that $\bigcup K_m = \Theta$. The sequence $\{K_m^c\}_m$ decreases to $\varnothing$ so $\lim_{m\to\infty}\Pi(f\mathbb{1}_{K_m^c}) = 0$. Thus, for all $\varepsilon > 0$, there exists M such that, for all $m \geq M$, $\Pi(f\mathbb{1}_{K_m^c}) \leq \varepsilon$. So, for all $n > N$, $\frac{f\Pi_n(K_M^c)}{\Pi_n(f)} = \frac{fa_n\Pi_n(K_M^c)}{a_n\Pi_n(f)} \leq \frac{2a_n\Pi_n(f\mathbb{1}_{K_M^c})}{\Pi(f)} \leq \frac{2\Pi(f\mathbb{1}_{K_M^c})}{\Pi(f)} \leq \frac{2\varepsilon}{\Pi(f)}$. The second inequality comes from assumption (3). Thus, $\{\frac{f\Pi_n}{\Pi_n(f)}\}_n$ is tight. The result follows from Proposition 2.8. □

Corollary 3.4. *Let* $\{\Pi_n\}_n$ *and* Π *be priors. Assume that*:

(1) *there exists a sequence of positive real numbers* $\{a_n\}_n$ *such that the sequence* $\{a_n\pi_n\}_n$ *converges pointwise to* π,

(2) *there exists a continuous function* $g : \Theta \to \mathbb{R}^+$ *such that* fg *is Lebesgue integrable and for all* $n \in \mathbb{N}$ *and* $\theta \in \Theta$, $a_n\pi_n(\theta) < g(\theta)$,

(3) $\theta \longmapsto f(x|\theta)$ *is continuous and positive*,

(4) *all the posteriors* $\Pi_n(\cdot|x)$ *and* $\Pi(\cdot|x)$ *are proper.*

Then $\{\Pi_n(\cdot|x)\}_n$ *converges narrowly to* $\Pi(\cdot|x)$.

Proof. From Proposition 2.15, assumptions (1) and (2) imply that $\{\Pi_n\}_n$ converges q-vaguely to Π. From assumption (2), for all n, $a_n f(x|\theta)\pi_n(\theta) \leq f(x|\theta)g(\theta)$. Since fg is Lebesgue integrable, by dominated convergence theorem, $\lim_{n\to\infty}\int a_n f(x|\theta)\pi_n(\theta)\,\mathrm{d}\theta = \int \lim_{n\to\infty} a_n f(x|\theta)\pi_n(\theta)\,\mathrm{d}\theta = \int f(x|\theta)\pi(\theta)\,\mathrm{d}\theta$. Thus, $\{a_n\Pi_n(f)\}_n$ converges to $\Pi(f) > 0$ so there exists N such that for all $n > N$, $a_n\Pi_n(f) \geq \frac{1}{2}\Pi(f)$. Consider $\{K_m\}_{m\in\mathbb{N}}$ an increasing sequence of compact sets such that $\bigcup K_m = \Theta$. The sequence $\{K_m^c\}_{m\in\mathbb{N}}$ decreases to $\varnothing$ so $\lim_{m\to\infty}\lambda(fg\mathbb{1}_{K_m^c}) = 0$. Thus, for all $\varepsilon > 0$, there exists M such that for all $m \geq M$, $\lambda(fg\mathbb{1}_{K_m^c}) \leq \varepsilon$. So, for all $n > N$, $\frac{fa_n\Pi_n(K_M^c)}{a_n\Pi_n(f)} \leq \frac{2a_n\Pi_n(f\mathbb{1}_{K_M^c})}{\Pi(f)} \leq \frac{2\lambda(fg\mathbb{1}_{K_M^c})}{\Pi(f)} \leq \frac{2\varepsilon}{\Pi(f)}$. Thus, $\{\Pi_n(\cdot|x)\}_n$ is a tight sequence of probabilities. The result follows from Proposition 3.1. □

The following result will be useful to explain the Jeffreys–Lindley paradox (see Section 7).

Corollary 3.5. *Consider a sequence of probabilities* $\{\Pi_n\}_n$ *which converges vaguely to the proper measure* Π. *Assume that*:

(1) $\theta \longmapsto f(x|\theta)$ *is continuous and non-negative*,
(2) $f(x|\cdot) \in \mathcal{C}_0(\Theta)$.

Then $\{\Pi_n(\cdot|x)\}_n$ *converges narrowly to* $\Pi(\cdot|x)$.

Proof. Since the Π_n and Π are proper measures and $f(\cdot|\theta)$ is a p.d.f., $\Pi_n(\cdot|x)$ and $\Pi(\cdot|x)$ are probabilities. We assume that $\{\Pi_n\}_n$ converges vaguely, and so q-vaguely, to Π and that f satisfies (1). So, from Proposition 3.1, $\{\Pi_n(\cdot|x)\}_n$ converges q-vaguely to $\Pi(\cdot|x)$. From Lemma 2.1, $\{\Pi_n(f)\}_n$ converges to $\Pi(f)$. So, there exists N such that for $n > N$, $\Pi_n(f) > \frac{\Pi(f)}{2}$. Moreover, from assumption (2), for all $\varepsilon > 0$, there exists a compact K such that for all $\theta \in K^c$, $f(\theta|x) \leq \varepsilon$. Thus, for all $n > N$, $\frac{f\Pi_n(K^c)}{\Pi_n(f)} \leq \frac{2\Pi_n(f\mathbb{1}_{K^c})}{\Pi(f)} \leq \frac{2\varepsilon}{\Pi(f)}$. Thus, $\{\frac{f\Pi_n}{\Pi_n(f)}\}_n$ is tight. The result follows from Proposition 3.1. □

Now, we establish some links between the q-vague convergence of $\{\Pi_n\}_n$ and the convergence of the Bayes estimates $\mathbb{E}_{\Pi_n}(\theta|x)$.

Proposition 3.6. *Let $\{\Pi_n\}_n$ be a sequence of priors which converges q-vaguely to Π. Assume that*:

(1) $\theta \longmapsto f(x|\theta)$ *is a non-zero continuous function on* Θ,

(2) *the family* $\{\Pi_n(\cdot|x)\}_n$ *is a family of probabilities uniformly integrable* (*see Billingsley* [7], *page* 32).

Then $\lim_{n\to\infty} \mathbb{E}_{\Pi_n}(\theta|x) = \mathbb{E}_{\Pi}(\theta|x)$.

Proof. From Proposition 3.1, $\{\Pi_n(\theta|x)\}_n$ converges q-vaguely to $\Pi(\theta|x)$. For all n, $\Pi_n(\cdot|x)$ and $\Pi(\cdot|x)$ are probability measures and $\{\Pi_n(\cdot|x)\}_n$ uniformly integrable implies that $\{\Pi_n(\cdot|x)\}_n$ is tight. So, from Proposition 3.1, $\{\Pi_n(\theta|x)\}_n$ converges narrowly to $\Pi(\theta|x)$. The result follows from Billingsley ([7], Theorem 5.4). □

We give an other version of Proposition 3.6 with a more restrictive but easier-to-check condition than uniform integrability.

Corollary 3.7. *Let $\{\Pi_n\}_n$ be a sequence of priors which converges q-vaguely to Π. Assume that $\theta \longmapsto f(x|\theta)$ is a non-zero continuous function on Θ, and that $\{\Pi_n(\cdot|x)\}_n$ is a family of probabilities such that $\{\mathrm{Var}_{\Pi_n}(\theta|x)\}_n$ is bounded above. Then $\lim_{n\to\infty} \mathbb{E}_{\Pi_n}(\theta|x) = \mathbb{E}_{\Pi}(\theta|x)$.*

Proof. This is a consequence of Billingsley ([7], page 32) and Proposition 3.6. □

4. Some constructions of sequences of vague priors

In this section, we give some constructions of sequences of probability measures that approximate a given improper prior such as the Haar measures or the Jeffreys prior. We have shown in the proof of Proposition 2.6 that any improper prior may be approximated by truncation. Here, we give other constructions for the Haar measure or the Jeffreys prior.

4.1. Location and scale models

The parameter θ is said to be a location parameter if there exists a p.d.f. g such that $f(x|\theta) = g(x-\theta)$. For instance, it is the case when $X|\theta \sim \mathcal{N}(\theta, \sigma^2)$ with known σ^2. The underlying group is $(\mathbb{R}, +)$ and the Haar measure $\lambda_{\mathbb{R}}$ is improper.

Proposition 4.1. *Let Π be a continuous probability measure on $\mathbb{R}$. Assume that the p.d.f. $\pi(\theta)$ of Π with respect to the Lebesgue measure $\lambda_{\mathbb{R}}$ is bounded above by a continuous and increasing function and is continuous at $\theta = 0$ with $\pi(0) > 0$. We define Π_n by $\pi_n(\theta) = \frac{1}{n}\pi(\frac{\theta}{n})$. Then, $\{\Pi_n\}_{n>0}$ converges q-vaguely to $\lambda_{\mathbb{R}}$.*

Proof. Put $\pi_n(\theta) = \frac{1}{n}\pi(\frac{\theta}{n})$. Put $a_n = n$, then $\lim_{n\to\infty} a_n\pi_n(\theta) = \lim_{n\to\infty}\pi(\frac{\theta}{n}) = \pi(0) > 0$ since π is continuous at 0. Moreover, π is bounded above by a continuous and increasing function, so there exists g such that, for all $\theta \in \mathbb{R}$ and for all $n > 0$, $\pi(\frac{\theta}{n}) \le g(\frac{\theta}{n}) \le g(\theta)$. The result follows from Proposition 2.15. □

We note that Hartigan [15] used a dual approach. He reduced the influence of the prior by letting the conditional variance σ^2 reducing to 0. He arrived at similar conclusions. He assumed that Π is locally uniform at 0, but it is equivalent to assuming that Π is continuous and positive at 0. We replace his condition "π tail-bounded" by the condition "π bounded".

Remark 4.2. Proposition 4.1 holds with the assumption "π bounded" instead of "π bounded above by a continuous and increasing function".

We now study the scale model. The strictly positive parameter σ is said to be a scale parameter if $f(x|\sigma) = \frac{1}{\sigma}g(\frac{x}{\sigma})$ where g is a p.d.f. If σ is a scale parameter, $\log(\sigma)$ is a location parameter for $\log(X)$. Here, the concerned group is $(\mathbb{R}^+ \setminus \{0\}, \times)$ and the Haar measure $\frac{1}{\sigma}\lambda_{\mathbb{R}^+\setminus\{0\}}$ is improper. The following proposition is the equivalent of Proposition 4.1 for the Haar measure on $(\mathbb{R}^+ \setminus \{0\}, \times)$.

Corollary 4.3. *Let Π be a continuous probability measure on $\mathbb{R}^+ \setminus \{0\}$. Assume that the p.d.f. $\pi(\sigma)$ of Π with respect to the Lebesgue measure $\lambda_{\mathbb{R}^+\setminus\{0\}}$ is bounded above by a continuous and increasing function and is continuous at $\sigma = 1$ with $\pi(1) > 0$. We define Π_n by $\pi_n(\sigma) = \frac{1}{n}\sigma^{1/n-1}\pi(\sigma^{1/n})$. Then $\{\Pi_n\}_{n>0}$ converges q-vaguely to $\frac{1}{\sigma}\lambda_{\mathbb{R}^+\setminus\{0\}}$.*

Proof. Put $\theta = \log(\sigma)$. From Proposition 2.7, $\widetilde{\pi}(\theta) = \mathrm{e}^{\theta}\pi(\mathrm{e}^{\theta})$ which is bounded above by the continuous and increasing function $\mathrm{e}^{\theta}g(\mathrm{e}^{\theta})$. The result follows from Proposition 4.1. □

4.2. Jeffreys conjugate priors (JCPs)

The Jeffreys prior is one of the most popular prior when no information is available, but in many cases, is improper. Consider that the distribution $X|\theta$ belongs to an exponential family, i.e., $f(x|\theta) = \exp\{\theta \cdot t(x) - \phi(\theta)\}h(x)$, for some functions $t(x)$, $h(x)$ and $\phi(\theta)$, and $\theta \in \Theta$, where Θ is an open set in $\mathbb{R}^p$, $p \geq 1$, such that $f(x|\theta)$ is a well-defined p.d.f. We assume that $\phi(\theta)$ and $I_\theta(\theta)$ are continuous. These conditions are satisfied if $t(X)$ is not concentrated on an hyperplane almost surely (see Barndorff-Nielsen [1]). Druilhet and Pommeret [12] proposed a class of conjugate priors that aims to approximate the Jeffreys prior and that is invariant with respect to smooth reparameterization. The notion of approximation was defined only from an intuitive point of view. We can now give a more rigorous approach by using the q-vague convergence.

Denote by $\pi^J(\theta) = |I_\theta(\theta)|^{1/2}$ the p.d.f. of the Jeffreys prior with respect to the Lebesgue measure, where θ is the natural parameter of the exponential family and $I_\theta(\theta)$ is the determinant of the Fisher information matrix. The JCPs are defined through their p.d.f. with respect to the Lebesgue measure by

$$\pi^J_{\alpha,\beta}(\theta) \propto \exp\{\alpha.\theta - \beta\phi(\theta)\}\left|I_\theta(\theta)\right|^{1/2},$$

and for a smooth reparameterization $\theta \to \eta$ by

$$\pi^J_{\alpha,\beta}(\eta) \propto \exp\{\alpha.\theta(\eta) - \beta\phi(\theta(\eta))\}\left|I_\eta(\eta)\right|^{1/2}.$$

Proposition 4.4. *Let* $\{(\alpha_n, \beta_n)\}_n$ *be a sequence of real numbers that converges to* $(0, 0)$. *Then, for the natural parameter* θ *or for any smooth reparameterization* η, $\{\Pi^J_{\alpha_n,\beta_n}\}_n$ *converges* q*-vaguely to* Π^J.

Proof. Choose $\{a_n\}_n$ such that $a_n\pi^J_{\alpha_n,\beta_n}(\theta) = \exp\{\alpha_n\theta - \beta_n\phi(\theta)\}|I_\theta(\theta)|^{1/2}$, which converges pointwise to $|I_\theta(\theta)|^{1/2}$. Put $\gamma_n = (\alpha_n, \beta_n)$ and $\psi(\theta) = (\theta, -\phi(\theta))$. We have $\gamma_n \cdot \psi(\theta) = \alpha_n\theta - \beta_n\phi(\theta)$. By Cauchy–Schwarz inequality, $\gamma_n \cdot \psi(\theta) \leq \|\gamma_n\|\|\psi(\theta)\|$. Since γ_n converges to $(0, 0)$, there exists N such that, for $n > N$, $\|\gamma_n\| < 1$. Let K be a compact set in Θ, by continuity of $\psi(\theta)$, since $\phi(\theta)$ is continuous, and by continuity of $I_\theta(\theta)$, there exist M_1 and M_2 such that, for all $\theta \in K$, $\|\psi(\theta)\| < M_1$ and $|I_\theta(\theta)|^{1/2} < M_2$. Therefore, $a_n\pi^J_{\alpha_n,\beta_n}(\theta) \leq M_2 \exp\{M_1\}$. The result follows from Proposition 2.16. □

Even if we have the convergence to the Jeffreys prior, we have no guaranty that $\Pi^J_{\alpha_n,\beta_n}$ is a proper prior and there is no general result to characterize this property such as in Diaconis and Ylvisaker [11] for usual conjugate priors. For example, consider inverse Gaussian models with likelihood $f(x; \mu, \lambda) = (\frac{\lambda}{2\pi x^3})^{1/2} \exp(\frac{-\lambda(x-\mu)^2}{2\mu^2 x})\mathbb{1}_{\{x>0\}}$ where $\mu > 0$ denotes the mean parameter and $\lambda > 0$ stands for the shape parameter. Considering the parameterization $(\psi = \frac{1}{\mu}, \lambda)$, the JCPs are given by $\pi^J_{\alpha,\beta}(\psi, \lambda) \propto \mathrm{e}^{-(\lambda/2)(\alpha_1\psi^2 - 2\beta\psi + \alpha_2)}\psi^{-1/2}\lambda^{(\beta-1)/2}$. Druilhet and Pommeret [12] showed that $\pi^J_{\alpha,\beta}(\psi, \lambda)$ is proper iff $\alpha_1 > 0$, $\alpha_2 > 0$ and $-\frac{1}{2} \leq \beta < \sqrt{\alpha_1\alpha_2}$. So, we may consider the sequences $\alpha_{1,n} = \alpha_{2,n} = \frac{1}{n}$ and $\beta_n = \frac{1}{2n}$. By Proposition 4.4, $\Pi^J_{\alpha_n,\beta_n}(\psi, \lambda)$ is therefore a sequence of proper priors that converges q-vaguely to the Jeffreys prior Π^J.

Remark 4.5. For any continuous function g on Θ, we can define $\pi^g_{\alpha,\beta}(\theta) \propto \exp\{\alpha.\theta - \beta\phi(\theta)\}g(\theta)$ and $\pi^g(\theta) = g(\theta)$. Similarly to Proposition 4.4, it can be shown that $\{\Pi^g_{\alpha_n,\beta_n}\}$ converges q-vaguely to Π^g.

5. Some examples

In this section, we consider some usual distributions and we look at the q-vague limiting measure.

5.1. Approximation of flat prior from Uniform distributions

5.1.1. *The discrete case*

Consider $\Theta = \mathbb{N}$, and $\Pi_n = \mathcal{U}(\{0, 1, \ldots, n\})$ the Uniform distribution on $\{0, \ldots, n\}$. Then $\{\Pi_n\}_n$ converges q-vaguely to the counting measure.

Indeed, $\pi_n(\theta) = \frac{1}{n+1}\mathbb{1}_{\{0,1,\ldots,n\}}(\theta)$. Put $a_n = n + 1$, then, for $\theta \in \mathbb{N}$, $\lim_{n\to\infty} a_n\pi_n(\theta) = \lim_{n\to\infty} \mathbb{1}_{\{0,1,\ldots,n\}}(\theta) = 1$. The result follows from Proposition 2.14.

5.1.2. *The continuous case*

Let $\Theta = \mathbb{R}$, and $\Pi_n = \mathcal{U}([-n, n])$ the Uniform distribution on $[-n, n]$. Then $\{\Pi_n\}_n$ converges q-vaguely to the Lebesgue measure $\lambda_{\mathbb{R}}$.

It corresponds to a location model so the result follows from Proposition 4.1 with $\Pi = \mathcal{U}([-1, 1])$.

5.2. Poisson distribution

Here is an example where a family of proper priors does not converge q-vaguely. Let $\Theta = \mathbb{N}$ and Π_n be the Poisson distribution with $\pi_n(\theta) = \exp(-n)\frac{n^\theta}{\theta!}$. Assume that there exists Π such that $\{\Pi_n\}_n$ converges q-vaguely to Π. Then, from Proposition 2.14, there exists a sequence $\{a_n\}_n$ such that for all $\theta \in \Theta$, $\lim_{n\to\infty} a_n\pi_n(\theta) = \pi(\theta)$. Consider $\theta_0 \in \Theta$ such that $\pi(\theta_0) > 0$. There exists N such that, for all $n > N$, $\pi_n(\theta_0) > 0$. Consider $\theta > \theta_0$, for all $n > N$, $\frac{\pi_n(\theta)}{\pi_n(\theta_0)} = \frac{\theta_0!}{\theta!}n^{\theta-\theta_0}$ and $\lim_{n\to\infty} \frac{\pi_n(\theta)}{\pi_n(\theta_0)} = \frac{\pi(\theta)}{\pi(\theta_0)} < +\infty$. On the other side, $\lim_{n\to\infty} \frac{\theta_0!}{\theta!}n^{\theta-\theta_0} = +\infty$. This is a contradiction. So, there is no prior Π such that $\{\Pi_n\}_n$ converges q-vaguely to Π.

5.3. Normal distribution

Let $\Theta = \mathbb{R}$ and $\Pi_n = \mathcal{N}(0, n)$ the normal distribution with zero mean and variance equal to n. Then $\{\Pi_n\}_n$ converges q-vaguely to the Lebesgue measure on $\mathbb{R}$.

Indeed, $\pi_n(\theta) = \frac{1}{\sqrt{2\pi n}}\mathrm{e}^{-\theta^2/2n}$ and $\pi(\theta) = 1$. Put $a_n = \sqrt{2\pi n}$, $n > 0$. Then $\{a_n\pi_n\}_{n>0}$ converges pointwise to 1. Moreover, for all n and all θ, $a_n\pi_n(\theta) < 2$. The result follows from Proposition 2.15.

Remark 5.1. From Theorem 2.5, $\{\mathcal{N}(0, n)\}_{n>0}$ cannot converge to another limiting measure than the Lebesgue measure (up to within a scalar factor).

More generally, it can be shown that the limiting measure is the same for $\{\mathcal{N}(\mu_n, n)\}_n$ where $\{\mu_n\}_n$ is a constant or a bounded sequence. So, we consider now the case where $\lim_{n\to\infty} \mu_n = +\infty$ by taking $\mu_n = n$.

Proposition 5.2. *We have three cases for the convergence of $\mathcal{N}(n, \sigma_n^2)$:*

1. *If $\lim_{n\to\infty} \frac{n}{\sigma_n^2} = +\infty$, then $\{\mathcal{N}(n, \sigma_n^2)\}_n$ does not converge q-vaguely.*
2. *If $\lim_{n\to\infty} \frac{n}{\sigma_n^2} = c$ with $0 < c < \infty$, then $\{\mathcal{N}(n, \sigma_n^2)\}_n$ converges q-vaguely to $e^{c\theta}\, d\theta$.*
3. *If $\lim_{n\to\infty} \frac{n}{\sigma_n^2} = 0$, then $\{\mathcal{N}(n, \sigma_n^2)\}_n$ converges q-vaguely to $\lambda_{\mathbb{R}}$.*

Proof. For all $n > 0$, we denote by $\Pi_n = \mathcal{N}(n, \sigma_n^2)$, and by π_n the p.d.f. with respect to the Lebesgue measure, $\pi_n(\theta) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp(-\frac{(\theta-n)^2}{2\sigma_n^2})$.

1. Put $\widetilde{\pi}_n(\theta) = \exp(-\frac{\theta^2}{2\sigma_n^2} + \frac{\theta n}{\sigma_n^2})$ and $\widetilde{\pi}(\theta) = e^{n^2/(2\sigma_n^2)}\pi(\theta)$. So $\{\Pi_n\}_n$ converges q-vaguely iff $\{\widetilde{\Pi}_n\}_n$ converges q-vaguely. Assume that there exists $\widetilde{\Pi}$ such that $\{\widetilde{\Pi}_n\}_n$ converges q-vaguely to $\widetilde{\Pi}$. Then there exists a sequence $\{a_n\}_n$ such that $\{a_n\widetilde{\Pi}_n\}_n$ converges vaguely to $\widetilde{\Pi}$. Since $\widetilde{\Pi} \neq 0$, there exists an interval $[A_1, A_2]$ such that $-\infty < A_1 < A_2 < +\infty$ and $0 < \widetilde{\Pi}([A_1, A_2]) < +\infty$. Consider $[B_1, B_2]$ such that $A_2 < B_1 < B_2 < +\infty$. There exists N such that for $n > N$, $\theta \longmapsto -\frac{\theta^2}{2n} + \frac{\theta n}{\sigma_n^2}$ is non-decreasing. For a such n, $\widetilde{\Pi}_n([B_1, B_2]) \geq (B_2 - B_1)\exp(-\frac{B_1}{2\sigma_n^2} + \frac{B_1 n}{\sigma_n^2})$ and $\widetilde{\Pi}_n([A_1, A_2]) \leq (A_2 - A_1)\exp(-\frac{A_2}{2\sigma_n^2} + \frac{A_2 n}{\sigma_n^2})$. So $\frac{\widetilde{\Pi}_n([B_1,B_2])}{\widetilde{\Pi}_n([A_1,A_2])} \geq \frac{B_2-B_1}{A_2-A_1}\exp(C(n))$ with $C(n) = \frac{n(B_1-A_2)}{\sigma_n^2} - \frac{(B_1^2-A_2^2)}{2\sigma_n^2} \geq \frac{n(B_1-A_2)}{2\sigma_n^2}$. Thus, $\lim_{n\to\infty} \frac{\widetilde{\Pi}_n([B_1,B_2])}{\widetilde{\Pi}_n([A_1,A_2])} = +\infty$ but $\lim_{n\to\infty} \frac{\widetilde{\Pi}_n([B_1,B_2])}{\widetilde{\Pi}_n([A_1,A_2])} = \frac{\widetilde{\Pi}([B_1,B_2])}{\widetilde{\Pi}([A_1,A_2])} < +\infty$. So, $\{\Pi_n\}_n$ does not converge q-vaguely.

2. Put $a_n = \frac{1}{\sqrt{2\pi}\sigma_n}\exp(-\frac{n^2}{2\sigma_n^2})$. Then $\lim_{n\to\infty} a_n\pi_n(\theta) = \lim_{n\to\infty} \exp(-\frac{\theta^2}{2\sigma_n^2} + \frac{\theta n}{\sigma_n^2}) = e^{c\theta}$. Since $\lim_{n\to\infty} \frac{n}{\sigma_n^2} = c$, there exists N such that for all $n > N$, $\frac{n}{\sigma_n^2} \in [c - \varepsilon, c + \varepsilon]$. So, for all $n > N$, $\exp(-\frac{\theta^2}{2\sigma_n^2} + \frac{\theta n}{\sigma_n^2}) \leq \exp((c + \varepsilon)\theta)$ which is continuous. The result follows from Proposition 2.15.

3. This is the same reasoning as point 2. with $\lim_{n\to\infty} a_n\pi_n(\theta) = 1$ and $a_n\pi_n(\theta) \leq 1 + \varepsilon$ for all $n > N$ and N large enough. □

Example 5.3. Assume that $X|\theta \sim \mathcal{N}(\theta, \sigma^2)$, σ^2 known, and put the prior $\Pi_n = \mathcal{N}(0, n)$ on θ. Then $\Pi_n(\theta|x) = \mathcal{N}(\frac{nx}{\sigma^2+n}, \frac{\sigma^2 n}{\sigma^2+n})$. From Section 5.3, the two first hypotheses are satisfied and $\{\mathcal{N}(0, n)\}_n$ converges q-vaguely to the Lebesgue measure $\lambda_{\mathbb{R}}$ so here, $\Pi = \lambda_{\mathbb{R}}$. Moreover, $\theta \longmapsto f(x|\theta)$ is continuous and positive on Θ and $\Pi(\cdot|x) = \mathcal{N}(x, \sigma^2)$ is proper. So, from Theorem 3.3, $\{\mathcal{N}(\frac{nx}{\sigma^2+n}, \frac{\sigma^2 n}{\sigma^2+n})\}_n$ converges narrowly to $\mathcal{N}(x, \sigma^2)$.

Example 5.4. To continue Example 5.3, $\mathrm{Var}_{\Pi_n}(\theta|x) = \frac{\sigma^2 n}{\sigma^2+n}$ is bounded above by σ^2 and the other hypothesis of Proposition 3.7 have already been verified in Example 5.3. So, from Proposition 3.7, $\lim_{n\to\infty} \mathbb{E}_{\Pi_n}(\theta|x) = \mathbb{E}_{\Pi}(\theta)$. Indeed, $\lim_{n\to\infty} \mathbb{E}_{\Pi_n}(\theta) = \lim_{n\to\infty} \frac{nx}{\sigma^2+n} = x = \mathbb{E}_{\Pi}(\theta)$.

5.4. Gamma distribution

5.4.1. *Approximation of* $\Pi = \frac{1}{\theta}\mathbb{1}_{\theta>0}\,\mathrm{d}\theta$

Let $\Theta = \mathbb{R}_+$ and $\Pi_n = \gamma(\alpha_n, \beta_n)$ the Gamma distributions with $\lim_{n\to\infty}(\alpha_n, \beta_n) = (0, 0)$. We have $\pi_n(\theta) = \frac{\beta_n{}^{\alpha_n}}{\Gamma(\alpha_n)}\theta^{\alpha_n-1}\mathrm{e}^{-\beta_n\theta}$. Put $a_n = \frac{\Gamma(\alpha_n)}{\beta_n{}^{\alpha_n}}$. Then $a_n\pi_n(\theta) = \theta^{\alpha_n-1}\mathrm{e}^{-\beta_n\theta}$ and $\{a_n\pi_n(\theta)\}_n$ converges to $\frac{1}{\theta}$. Put $g(\theta) = \frac{1}{\theta}\mathbb{1}_{]0,1]}(\theta) + \mathbb{1}_{]1,+\infty[}(\theta)$. The sequence $\{\alpha_n\}_n$ goes to 0 so there exists N such that for all $n > N$, $\alpha_n < 1$. So, for $n > N$ and for $\theta > 0$, $a_n\pi_n(\theta) \le \theta^{\alpha_n-1} \le g(\theta)$. Since g is a continuous function on $\mathbb{R}_+^*$, from Proposition 2.15, $\{\Pi_n\}_n$ converges q-vaguely to $\frac{1}{\theta}\,\mathrm{d}\theta$.

Recall that for $\theta \sim \gamma(a, b)$, $\mathbb{E}(\theta) = \frac{a}{b}$ and $\mathrm{Var}(\theta) = \frac{a}{b^2}$. We can see below that the same convergence may be obtained with different convergences of the mean and variance.

- For $\Pi_n = \gamma(\frac{1}{n}, \frac{1}{n})$, $\mathbb{E}_{\Pi_n}(\theta) = 1$ for all n and $\lim_{n\to\infty}\mathrm{Var}_{\Pi_n}(\theta) = \lim_{n\to\infty} n = +\infty$.
- For $\Pi_n = \gamma(\frac{1}{n}, \frac{1}{\sqrt{n}})$, $\lim_{n\to\infty}\mathbb{E}_{\Pi_n}(\theta) = \lim_{n\to\infty}\frac{1}{\sqrt{n}} = 0$ and $\lim_{n\to\infty}\mathrm{Var}_{\Pi_n}(\theta) = 1$ for all n.
- For $\Pi_n = \gamma(\frac{1}{n}, \frac{1}{n^{1/3}})$, $\lim_{n\to\infty}\mathbb{E}_{\Pi_n}(\theta) = \lim_{n\to\infty} n^{-2/3} = 0$ and $\lim_{n\to\infty}\mathrm{Var}_{\Pi_n}(\theta) = \lim_{n\to\infty} n^{-1/3} = 0$.
- For $\Pi_n = \gamma(\frac{1}{n}, \frac{1}{n^2})$, $\lim_{n\to\infty}\mathbb{E}_{\Pi_n}(\theta) = \lim_{n\to\infty} n = +\infty$ and $\lim_{n\to\infty}\mathrm{Var}_{\Pi_n}(\theta) = \lim_{n\to\infty} n^3 = +\infty$.
- For $\Pi_n = \gamma(\frac{1}{n}, \frac{1}{n^{2/3}})$, $\lim_{n\to\infty}\mathbb{E}_{\Pi_n}(\theta) = n^{-1/2} = 0$ and $\lim_{n\to\infty}\mathrm{Var}_{\Pi_n}(\theta) = \lim_{n\to\infty} n^{1/3} = +\infty$.

More generally, if $\liminf_n \mathbb{E}_{\Pi_n}(\theta) > 0$ then $\lim_{n\to\infty}\mathrm{Var}_{\Pi_n}(\theta) = +\infty$, since $\mathrm{Var}_{\Pi_n}(\theta) = \frac{\mathbb{E}_{\Pi_n}(\theta)}{\beta_n}$ with $\lim_{n\to\infty}\beta_n = 0$.

5.4.2. *Approximation of* $\Pi = \frac{1}{\theta}\mathrm{e}^{-\theta}\mathbb{1}_{\theta>0}\,\mathrm{d}\theta$

Let us show that $\{\gamma(\alpha_n, 1)\}$ converges q-vaguely to $\frac{1}{\theta}\mathrm{e}^{-\theta}\mathbb{1}_{\theta>0}\,\mathrm{d}\theta$ when $\{\alpha_n\}$ goes to 0. Put $\Pi_n = \{\gamma(\alpha_n, 1)\}$. Then $\pi_n(\theta) = \frac{1}{\Gamma(\alpha_n)}\theta^{\alpha_n-1}\mathrm{e}^{-\theta}\mathbb{1}_{\theta>0}$ is the p.d.f. of Π_n. Put $a_n = \Gamma(\alpha_n)$, then $a_n\pi_n(\theta) = \theta^{\alpha_n-1}\mathrm{e}^{-\theta}\mathbb{1}_{\theta>0}$ converges to $\pi(\theta) = \frac{1}{\theta}\mathrm{e}^{-\theta}\mathbb{1}_{\theta>0}$. Moreover, since $\{\alpha_n\}_n$ goes to 0, there exists N such that for $n > N$, $\alpha_n < 1$. Put $g(\theta) = \frac{1}{\theta}\mathbb{1}_{]0,1]}(\theta) + \mathbb{1}_{]1,+\infty[}(\theta)$. So, for $n > N$ and $\theta > 0$, $a_n\pi_n(\theta) \le \theta^{\alpha_n-1} \le g(\theta)$. The function g is continuous so from Proposition 2.15, $\{\gamma(\alpha_n, 1)\}_n$ converges q-vaguely to $\frac{1}{\theta}\mathrm{e}^{-\theta}\mathbb{1}_{\theta>0}\,\mathrm{d}\theta$. Since $\lim_{n\to\infty}\alpha_n = 0$, we necessarily have $\lim_{n\to\infty}\mathbb{E}_{\Pi_n}(\theta) = 0$ and $\lim_{n\to\infty}\mathrm{Var}_{\Pi_n}(\theta) = 0$.

6. Convergence of Beta distributions

We now consider a more complex example which often appears in literature; see, for example, Tuyl *et al.* [28]. Let X represents the number of successes in N Bernoulli trials, and θ be the probability of a success in a single trial. Since the Beta distribution and the Binomial distribution form a conjugate pair, a common prior distribution on θ is $\beta(\alpha, \alpha)$ which have mean and median equal to $\frac{1}{2}$. Three "plausible" non-informative priors were listed by Berger ([3], page 89):

the Bayes–Laplace prior $\beta(1,1)$, the Jeffreys prior $\beta(\frac{1}{2},\frac{1}{2})$ and the improper Haldane prior, wrote down $\beta(0,0)$, whose density is $\pi_H(\theta)=\frac{1}{\theta(1-\theta)}$ with respect to the Lebesgue measure on $]0,1[$. If we want $\beta(\alpha,\alpha)$ with large variance, necessarily α must be close to 0. Thus, we choose $\beta(\frac{1}{n},\frac{1}{n})$. The density of $\Pi_n=\beta(\frac{1}{n},\frac{1}{n})$ with respect to the Lebesgue measure on $]0;1[$ is $\pi_n(\theta)=\frac{1}{B(1/n,1/n)}\theta^{1/n-1}(1-\theta)^{1/n-1}$. As mentioned, for example, by Bernardo [5] or Lane and Sudderth [19], there are two possible limiting distributions for $\beta(\frac{1}{n},\frac{1}{n})$ when n goes to $+\infty$. The first one is $\frac{1}{2}(\delta_0+\delta_1)$ which is the limiting measure given by the standard probability theory. The second one is the Haldane prior Π_H which is deduced from the posterior distributions and estimators (Lehmann and Casella [22]). We show that it depends on the space where θ lives. Choosing $]0,1[$ or $[0,1]$ does not matter for $\beta(\frac{1}{n},\frac{1}{n})$ but it matters for the limiting distributions. We may note that the Haldane prior is a Radon measure on $]0,1[$ but not on $[0,1]$ and that $\frac{1}{2}(\delta_0+\delta_1)$ is not defined on $]0,1[$.

6.1. Convergence on]0, 1[

In this section, we study the convergences on $]0,1[$ of $\{\beta(\frac{1}{n},\frac{1}{n})\}_{n>0}$, of the sequence of posteriors and of the sequence of estimators.

Put $a_n=B(\frac{1}{n},\frac{1}{n})$, then $a_n\pi_n(\theta)=\theta^{1/n-1}(1-\theta)^{1/n-1}$ converges to $\pi_H(\theta)=[\theta(1-\theta)]^{-1}$ and for any θ and n, $a_n\pi_n(\theta)<5$. Therefore, from Theorem 2.15, $\{\beta(\frac{1}{n},\frac{1}{n})\}_{n>0}$ converges q-vaguely to Π_H.

Consider the sequence of posteriors. The sequence of priors $\{\Pi_n\}_n$ converges q-vaguely to Π_H and $\theta\longmapsto f(x|\theta)$ is continuous on Θ. Then, from Lemma 3.1:

- if $x=0$, $\{\Pi_n(\theta|x)\}_n$ converges q-vaguely to the improper measures with p.d.f. $\pi(\theta)=(1-\theta)^{N-1}\theta^{-1}$,
- if $x=N$, $\{\Pi_n(\theta|x)\}_n$ converges q-vaguely to the improper measures with p.d.f. $\pi(\theta)=\theta^{N-1}(1-\theta)^{-1}$,
- if $0<x<N$, $\{\Pi_n(\theta|x)\}_n$ converges q-vaguely to $\Pi_H(\theta|x)=\beta(x,N-x)$.

For $0<x<N$, $\beta(x,N-x)$ is proper and $\theta\longmapsto f(x|\theta)$ is continuous and positive. So, from Theorem 3.3, $\{\Pi_n(\theta|x)\}_{n>0}$ converges narrowly to $\Pi_H(\theta|x)=\beta(x,N-x)$.

Consider now the Bayes estimators $\mathbb{E}_{\Pi_n}(\theta|x)=\frac{1+nx}{2+nN}$ which tend to $\frac{x}{N}$. So:

- If $x=0$, $\lim_{n\to\infty}\mathbb{E}_{\Pi_n}(\theta|x=0)=0$ whereas $\mathbb{E}_{\Pi_H}(\theta|x=0)=\frac{1}{N}$.
- If $x=N$, $\lim_{n\to\infty}\mathbb{E}_{\Pi_n}(\theta|x=N)=1$ whereas $\mathbb{E}_{\Pi_H}(\theta|x=N)=+\infty$.
- If $0<x<N$, $\lim_{n\to\infty}\mathbb{E}_{\Pi_n}(\theta|x)=\frac{x}{N}=\mathbb{E}_{\Pi_H}(\theta|x)$.

For $x=0$ and $x=N$, $\Pi_H(\cdot|x)$ is an improper measure. In this case, $\mathbb{E}_{\Pi_H}(\theta|x)=\int_\Theta\theta\,\mathrm{d}\Pi_H(\theta|x)$.

6.2. Convergence on [0, 1]

In this section, we study the convergences on $[0,1]$ of $\{\beta(\frac{1}{n},\frac{1}{n})\}_{n>0}$, of the sequence of posteriors and of the sequence of estimators.

For all n and for $0 < t < 1$, $\Pi_n([0, t[) + \Pi_n([t, 1 - t]) + \Pi_n(]1 - t, 1]) = 1$. But on $]0, 1[$, $\{\beta(\frac{1}{n}, \frac{1}{n})\}_{n>0}$ converges q-vaguely to the improper measure Π_H, so $\lim_{n\to\infty} \Pi_n([t, 1 - t]) = 0$. Moreover, for all n, $\Pi_n([0, t[) = \Pi_n(]1 - t, 1])$. Thus, for all $0 < t < 1$, $\lim_{n\to\infty} \Pi_n([0, t[) = \frac{1}{2}$. From Billingsley ([8], page 192), $\{\beta(\frac{1}{n}, \frac{1}{n})\}_{n>0}$ converges narrowly to $\frac{1}{2}(\delta_0 + \delta_1) = \Pi_{\{0,1\}}$. By Theorem 2.5, $\{\beta(\frac{1}{n}, \frac{1}{n})\}_{n>0}$ cannot converge to an other limit such as the Haldane measure, which is not a Radon measure on [0, 1].

The limit of the posterior distributions can be deduced from the limit of the prior distributions only for $x = 0$ and $x = N$.

- If $x = 0$, $\{\Pi_n(\theta|x = 0)\}$ converges narrowly to $\Pi_{\{0,1\}}(\theta|x = 0) = \delta_0$.
- If $x = N$, $\{\Pi_n(\theta|x = N)\}$ converges narrowly to $\Pi_{\{0,1\}}(\theta|x = N) = \delta_1$.
- If $0 < x < N$, $\{\Pi_n(\theta|x)\}$ converges narrowly to $\beta(x, N - x)$ whereas $\Pi_{\{0,1\}}(\theta|x)$ does not exist.

Similarly, the limit of the estimators can be deduced from the limit of the prior distributions only for $x = 0$ and $x = N$.

- If $x = 0$, $\lim_{n\to\infty} \mathbb{E}_{\Pi_n}(\theta|x = 0) = 0 = \mathbb{E}_{\Pi_{\{0,1\}}}(\theta|x = 0)$.
- If $x = N$, $\lim_{n\to\infty} \mathbb{E}_{\Pi_n}(\theta|x = N) = 1 = \mathbb{E}_{\Pi_{\{0,1\}}}(\theta|x = N)$.
- If $0 < x < N$, $\lim_{n\to\infty} \mathbb{E}_{\Pi_n}(\theta|x) = \frac{x}{N}$ whereas $\mathbb{E}_{\Pi_{\{0,1\}}}(\theta|x)$ does not exist.

7. The Jeffreys–Lindley paradox

Consider the standard Gaussian model $X|\theta \sim \mathcal{N}(\theta, 1)$ and the point null hypothesis H_0: $\theta = 0$ tested against H_1: $\theta \neq 0$. If we use the prior $\pi(\theta) = \frac{1}{2}\mathbb{1}_{\theta=0} + \frac{1}{2}\mathbb{1}_{\theta\neq 0}$ with respect to the measure $\delta_0 + \lambda_{\mathbb{R}}$, it corresponds to the mass $\frac{1}{2}$ on H_0 and the Laplace prior on H_1. The posterior probability of H_0 is $\Pi(\theta = 0|x) = [1 + \sqrt{2\pi}\mathrm{e}^{x^2/2}]^{-1}$ so $\Pi(\theta = 0|x) \le [1 + \sqrt{2\pi}]^{-1} \approx 0.285$ whatever the data are. An alternative is to use a sequence of proper priors $\{\Pi_n\}_n$ whose p.d.f. are $\pi_n(\theta) = \frac{1}{2}\mathbb{1}_{\theta=0} + \frac{1}{2}\mathbb{1}_{\theta\neq 0}\frac{1}{\sqrt{2\pi}n}\mathrm{e}^{-\theta^2/(2n^2)}$. With these priors, we have $\pi_n(\theta = 0|x) = [1 + \sqrt{\frac{1}{1+n^2}}\mathrm{e}^{(n^2x^2)/(2(1+n^2))}]^{-1}$ which converges to 1. This limit differs from the "non-informative" answer $[1 + \sqrt{2\pi}\mathrm{e}^{x^2/2}]^{-1}$ and is considered as a paradox. In the light of the concept of q-vague convergence, this result is not paradoxal since, as shown in Proposition 7.1, the sequence of priors $\{\frac{1}{2}\delta_0 + \frac{1}{2}\mathcal{N}(0, n^2)\}_n$ converges vaguely to $\frac{1}{2}\delta_0$, and, the limiting posterior distribution corresponds to the posterior of the limit of the prior distributions. The following proposition generalizes this example.

Proposition 7.1. *Consider a partition*: $\Theta = \Theta_0 \cup \Theta_1$ *where* $\Theta_0 = \{\theta_0\}$. *Let* $\{\widetilde{\Pi}_n\}_n$ *be a sequence of probabilities on* Θ *which converges q-vaguely to the improper measure* $\widetilde{\Pi}$ *and such that* $\widetilde{\Pi}_n(\theta_0) = \widetilde{\Pi}(\theta_0) = 0$. *Put* $\Pi_n = \rho\delta_{\theta_0} + (1 - \rho)\widetilde{\Pi}_n$ *where* $0 < \rho < 1$, *then* $\{\Pi_n\}_n$ *converges vaguely to* $\rho\delta_{\theta_0}$.

Moreover, assume that $\theta \longmapsto f(x|\theta)$ *is continuous and belongs to* $\mathcal{C}_0$. *Then* $\{\Pi_n(\cdot|x)\}$ *converges narrowly to* $\Pi(\cdot|x)$.

Proof. From Definition 2.2, there exists $\{a_n\}_n$ such that $\{a_n\widetilde{\Pi}_n\}_n$ converges vaguely to $\widetilde{\Pi}$. For $g \in \mathcal{C}_K$, $\Pi_n(g) = \rho g(\theta_0) + (1-\rho)\widetilde{\Pi}_n(g) = \rho g(\theta_0) + \frac{1-\rho}{a_n} a_n\widetilde{\Pi}_n(g)$. But $\lim_{n\to\infty} a_n\widetilde{\Pi}_n(g) = \widetilde{\Pi}(g) < \infty$. So, $\lim_{n\to\infty} \frac{1-\rho}{a_n} a_n\widetilde{\Pi}_n(g) = 0$ since, from Lemma 2.9, $\lim_{n\to\infty} a_n = +\infty$. Thus, $\lim_{n\to\infty}\Pi_n(g) = \rho g(\theta_0)$. The first result follows.

The second part is a direct consequence of Theorem 3.5. □

In the Proposition 7.1, it is assumed that $\theta \longmapsto f(x|\theta) \in \mathcal{C}_0(\Theta)$. Now, we consider the case where the limit of the likelihood $f(x|\theta)$ when θ is outside of any compact is not 0 but $f(x|\theta_0)$. In that case, the limit of the posterior probabilities is the same as the limit of the prior probabilities, as stated in the following proposition.

Proposition 7.2. *Consider the same notations and assumptions of Proposition 7.1. Moreover, assume that $\theta \longmapsto f(x|\theta)$ is continuous and such that for all $\varepsilon > 0$, there exists a compact K such that for all $\theta \in K^c$, $|f(x|\theta) - f(x|\theta_0)| \le \varepsilon$. Then $\lim_{n\to\infty}\Pi_n(\theta = \theta_0|x) = \Pi(\theta=\theta_0)$ and $\lim_{n\to\infty}\Pi_n(\theta \neq \theta_0|x) = \Pi(\theta \neq \theta_0)$.*

Proof. By Bayes formula: $\Pi_n(\theta=\theta_0|x) = \frac{\rho f(x|\theta_0)}{\rho f(x|\theta_0) + (1-\rho)\int_\Theta f(x|\theta)\,\mathrm{d}\widetilde{\Pi}_n(\theta)}$. But, for all $\varepsilon > 0$, there exists a compact K such that, for all $\theta \in K^c$, $|f(x|\theta) - f(x|\theta_0)| \le \varepsilon$. So $\int_\Theta f(x|\theta)\,\mathrm{d}\widetilde{\Pi}_n(\theta) = \int_K f(x|\theta)\,\mathrm{d}\widetilde{\Pi}_n(\theta) + \int_{K^c} f(x|\theta)\,\mathrm{d}\widetilde{\Pi}_n(\theta)$, where:

- $(f(x|\theta_0) - \varepsilon)\widetilde{\Pi}_n(K^c) \le \int_{K^c} f(x|\theta)\,\mathrm{d}\widetilde{\Pi}_n(\theta) \le (f(x|\theta_0) + \varepsilon)\widetilde{\Pi}_n(K^c)$. From Proposition 2.11, $\lim_{n\to\infty}\widetilde{\Pi}_n(K^c) = 1$. So, $\lim_{n\to\infty}\int_{K^c} f(x|\theta)\,\mathrm{d}\widetilde{\Pi}_n(\theta) = f(x|\theta_0)$.
- There exists $g \in \mathcal{C}_K(\Theta)$ such that $0 \le g \le 1$ and $g\mathbb{1}_K = 1$. For a such g,

$$\lim_{n\to\infty}\int_K f(x|\theta)\,\mathrm{d}\widetilde{\Pi}_n(\theta) \le \lim_{n\to\infty}\frac{1}{a_n}a_n\int_\Theta g(\theta)f(x|\theta)\,\mathrm{d}\widetilde{\Pi}_n(\theta) = 0$$

  since $\lim_{n\to\infty} a_n\int_\Theta g(\theta)f(x|\theta)\,\mathrm{d}\widetilde{\Pi}_n(\theta) = \int_\Theta g(\theta)f(x|\theta)\,\mathrm{d}\widetilde{\Pi}(\theta) < +\infty$ and $\lim_{n\to\infty} a_n = +\infty$ from Lemma 2.9.

Thus, $\lim_{n\to\infty}\Pi_n(\theta=\theta_0|x) = \frac{\rho f(x|\theta_0)}{\rho f(x|\theta_0) + (1-\rho)f(x|\theta_0)} = \rho = \Pi(\theta=\theta_0)$. □

To illustrate this result in a more general case, we consider an example proposed by Dauxois *et al.* [10]. They consider a model choice between $\mathcal{P}(m)$ the Poisson distribution, $\mathcal{B}(N,m)$ the Binomial distribution and $\mathcal{NB}(N,m)$ the Negative Binomial distribution. These models belong to the general framework of Natural Exponential Families (NEFs) and are determined by their variance function $V(m) = am^2 + m$ where m is the mean parameter. Thus, a null value for a relates to the Poisson NEF, a negative one to the Binomial NEF and a positive one to the Negative Binomial NEF. The prior distribution chosen on the parameter a is Π_K defined by

$$\Pi_K(a) = \begin{cases} \frac{1}{3}, & \text{if } a = 0, \\ \frac{1}{3K}, & \text{if } \frac{1}{a} \in \{1,\dots,K\}, \\ \frac{1}{3K}, & \text{if } -\frac{1}{a} \in \{n_0,\dots,n_0+K-1\}, \end{cases}$$

where K is an hyperparameter. Note that $\Pi_K(a=0) = \Pi_K(a>0) = \Pi_K(a<0) = \frac{1}{3}$.

Dauxois *et al.* [10] showed that the sequence of posterior distributions does not converge to δ_0 as in the previous case but $\Pi_K(a=0|X=x)$, $\Pi_K(a>0|X=x)$ and $\Pi_K(a<0|X=x)$ converge to the prior probabilities $\Pi_K(a=0)$, $\Pi_K(a>0)$ and $\Pi_K(a<0)$ whatever the data are when $K \to +\infty$.

Appendix: Properties of the quotient space

Proposition A.1. *$\overline{\mathcal{R}}$ is a Hausdorff space.*

Proof. This proof is based on two results of Bourbaki [9].

Step 1. $\mathcal{R}$ is a topological space and $\Gamma = \{\sigma_\alpha : \Pi \longmapsto \alpha\Pi, \alpha \in \mathbb{R}_+^*\}$ is a homeomorphism group of $\mathcal{R}$. We consider the equivalence relation: $\Pi \sim \Pi'$ iff there exists $\alpha > 0$ such that $\Pi = \alpha\Pi'$, that is, there exists $\sigma_\alpha \in \Gamma$ such that $\Pi = \sigma_\alpha(\Pi')$. So, from Bourbaki ([9], Section I.31), $\sim$ is open.

Step 2. Let us show that $G = \{(\Pi, \alpha\Pi), (\Pi, \alpha\Pi) \in \mathcal{R} \times \mathcal{R}\}$ which is the graph of $\sim$ is closed. Let $\{(\Pi_n, \alpha_n\Pi_n)\}_{n>0}$ be a sequence in G such that $\lim_{n\to\infty}(\Pi_n, \alpha_n\Pi_n) = (\Pi_0, \Pi_0')$. The aim is to show that $(\Pi_0, \Pi_0') \in G$, that is, (Π_0, Π_0') takes the form $(\Pi_0, \alpha_0\Pi_0)$ where $\alpha_0\Pi_0 \neq 0$. Since $\Pi_0 \neq 0$, there exists $f_0 \in \mathcal{C}_K$ such that $\Pi_0(f_0) > 0$. Moreover, $\lim_{n\to\infty}\Pi_n(f_0) = \Pi_0(f_0)$ so there exists N such that for all $n \geq N$, $\Pi_n(f_0) > 0$. For all $n \geq N$, $\lim_{n\to\infty}\alpha_n = \lim_{n\to\infty}\frac{\alpha_n\Pi_n(f_0)}{\Pi_n(f_0)} = \frac{\Pi_0'(f_0)}{\Pi_0(f_0)} = \alpha_0$. Thus, for all $f \in \mathcal{C}_K$, $\lim_{n\to\infty}\alpha_n\Pi_n(f) = \alpha_0\Pi_0(f)$ and $\lim_{n\to\infty}\alpha_n\Pi_n(f) = \Pi_0'(f)$. Since $\mathcal{R}$ is a Hausdorff space, $\alpha_0\Pi_0(f) = \Pi_0'(f)$. So, the graph of $\sim$, G, is closed. The result follows from Bourbaki ([9], Section I.55). □

Acknowledgment

The authors are grateful to Professors C.P. Robert, J. Rousseau and S. Dachian for helpful discussions. We acknowledge the comments from reviewers which resulted in an improved paper.

References

[1] Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*. Chichester: Wiley. MR0489333

[2] Bauer, H. (2001). *Measure and Integration Theory*. *De Gruyter Studies in Mathematics* **26**. Berlin: de Gruyter. MR1897176

[3] Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR0804611

[4] Berger, J.O., Bernardo, J.M. and Sun, D. (2009). The formal definition of reference priors. *Ann. Statist.* **37** 905–938. MR2502655

[5] Bernardo, J.-M. (1979). Reference posterior distributions for Bayesian inference. *J. Roy. Statist. Soc. Ser. B* **41** 113–147. MR0547240

[6] Bernardo, J.-M. and Smith, A. F. M. (1994). *Bayesian Theory*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Chichester: Wiley. MR1309682
[7] Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley. MR0233396
[8] Billingsley, P. (1986). *Probability and Measure*, 2nd ed. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. New York: Wiley. MR0830424
[9] Bourbaki, N. (1971). *Elements of Mathematics*. *General Topology*. *Part* 1. Paris: Hermann.
[10] Dauxois, J.-Y., Druilhet, P. and Pommeret, D. (2006). A Bayesian choice between Poisson, binomial and negative binomial models. *TEST* **15** 423–432. MR2273734
[11] Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families. *Ann*. *Statist*. **7** 269–281. MR0520238
[12] Druilhet, P. and Pommeret, D. (2012). Invariant conjugate analysis for exponential families. *Bayesian Anal*. **7** 903–916. MR3000019
[13] Eaton, M.L. (1989). *Group Invariance Applications in Statistics*. *NSF-CBMS Regional Conference Series in Probability and Statistics* **1**. Hayward, CA: IMS. MR1089423
[14] Hartigan, J.A. (1983). *Bayes Theory*. *Springer Series in Statistics*. New York: Springer. MR0715782
[15] Hartigan, J.A. (1996). Locally uniform prior distributions. *Ann*. *Statist*. **24** 160–173. MR1389885
[16] Jaynes, E.T. (2003). *Probability Theory*. Cambridge: Cambridge Univ. Press. MR1992316
[17] Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proc*. *Roy*. *Soc*. *London*. *Ser*. *A*. **186** 453–461. MR0017504
[18] Jeffreys, H. (1961). *Theory of Probability*, 3rd ed. Oxford: Clarendon Press. MR0187257
[19] Lane, D.A. and Sudderth, W.D. (1983). Coherent and continuous inference. *Ann*. *Statist*. **11** 114–120. MR0684869
[20] Lang, S. (1977). *Analyse Reelle*. Paris: InterEditions.
[21] Laplace, P.-S. (1995). *Théorie Analytique des Probabilités*. *Vol*. *II*. Paris: Éditions Jacques Gabay. Livre II: Théorie générale des probabilités. [Book II: General probability theory], Suppléments. [Supplements], Reprint of the 1820 third edition (Book II) and of the 1816, 1818, 1820 and 1825 originals (Supplements). MR1400403
[22] Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd ed. *Springer Texts in Statistics*. New York: Springer. MR1639875
[23] Rényi, A. (1970). *Foundations of Probability*. San Francisco, CA: Holden-Day. MR0264723
[24] Stone, M. (1963). The posterior t distribution. *Ann*. *Math*. *Statist*. **34** 568–573. MR0149586
[25] Stone, M. (1970). Necessary and sufficient condition for convergence in probability to invariant posterior distributions. *Ann*. *Math*. *Statist*. **41** 1349–1353. MR0266359
[26] Taraldsen, G. and Lindqvist, B.H. (2010). Improper priors are not improper. *Amer*. *Statist*. **64** 154–158. MR2757006
[27] Taraldsen, G. and Lindqvist, B.H. (2015). Improper priors and conditional probability spaces. *Comm*. *Statist*. *Theory Methods*. To appear.
[28] Tuyl, F., Gerlach, R. and Mengersen, K. (2009). Posterior predictive arguments in favor of the Bayes–Laplace prior as the consensus prior for binomial and multinomial parameters. *Bayesian Anal*. **4** 151–158. MR2486242

Received July 2014 and revised January 2015