FELLER OPERATORS AND MIXTURE PRIORS IN BAYESIAN NONPARAMETRICS

Sonia Petrone and Piero Veronese

L. Bocconi University, Milano. Italy

Appendix

A.1 Proof of the results in Section 3.

Proof of Lemma 1.

For simplicity, we prove the lemma in terms of the natural parameter rather than in the mean parameter. The results follow since there is a one-to-one positive monotone correspondence between the two parameterizations.

The result i) is due to Cifarelli and Regazzini (1987).

Point ii). With no loss of generality assume $a = 0$, otherwise one can translate Y. Because Θ is not empty, for $\theta_0 \in \Theta$ we have

$$\infty > M(\theta_0) = \log \int \exp\{\theta_0\, y\}\nu(dy) \geq \log \int \exp\{\theta_1\, y\}\nu(dy) = M(\theta_1)$$

for each $\theta_1 < \theta_0$. As a consequence $\inf \Theta = -\infty$ and we have to compute

$$\lim_{\theta \to -\infty} p_\theta(0) = \lim_{\theta \to -\infty} \exp\{-M(\theta)\} = \lim_{\theta \to -\infty} \left[\nu(0) + \int_{\mathcal{Y}\setminus\{0\}} \exp\{\theta\, y\}\nu(dy) \right]^{(-1)} . \quad \text{(A.1)}$$

If $\theta < \theta_0 < 0$, then $\exp\{\theta\, y\} < \exp\{\theta_0\, y\}$, since $y > 0$, with $\exp\{\theta_0\, y\}$ integrable w.r.t. ν. Furthermore $\lim_{\theta\to-\infty} \exp\{\theta\, y\} = 0$ and so, by the dominated convergence theorem, we have $\lim_{\theta\to-\infty} \int_{\mathcal{Y}\setminus\{0\}} \exp\{\theta\, y\}\nu(dy) = 0$. As a consequence (A.1) converges to $\nu(0)^{-1}$ and $\lim_{\theta\to-\infty} p_\theta(0)\nu(0) = 1$. Consider now $\theta \to \sup\Theta$. Since $\exp\{\theta\, y\}$ is a monotone increasing function in θ, from the monotone convergence theorem it follows directly that

$\lim_{\theta\to\sup\Theta}\int_{\mathcal{Y}}\exp\{\theta\, y\}\nu(dy) = M(\sup\Theta) = \infty$, where the last equality follows by the definition of Θ in a regular NEF. It follows that

$$\lim_{\theta\to\sup\Theta} p_\theta(0)\nu(0) = \lim_{\theta\to\sup\Theta}\exp\{-M(\theta)\}\nu(0) = \exp\{-M(\sup\Theta)\}\nu(0) = 0.$$

Point iii) is analogous to point ii) assuming $b = 0$, and noticing that $\sup\Theta = \infty$. $\diamond$

Proof of Theorem 2.

With no loss of generality, set $k = 1$ and notice that

$$H(x; z) = P_x([z,\infty)) = 1 - P_x(z) + p_x(z)\nu(z), \quad \forall x \in (a, b). \tag{A.2}$$

Consider first $a < z < b$. By (A.2) and using i) of Lemma 1, we have

$$\lim_{x\to a^+} P_x([z,\infty)) = 0 \quad \text{and} \quad \lim_{x\to b^-} P_x([z,\infty)) = 1.$$

For $a < x < b$, we have: $H(x; z) = 1 - P_x(z)$ when P_x is absolutely continuous, while, when P_x is discrete with support points $\{a = z_1 < z_2 < \cdots < z_N = b\}$, $N \leq \infty$,

$$H(x; z) = P_x([z,\infty)) = P_x([z_{i+1},\infty)) = 1 - P_x(z_i) \tag{A.3}$$

for $z_i < z \leq z_{i+1}$, $i = 1, 2, \ldots, N-2$ and for $z_{N-1} < z < z_N = b$ if $i = N-1$. In both cases, $H(x; z)$ is clearly a continuous function in x, and since $P_x(z)$ has a monotone likelihood ratio in z (see Lehmann (1959, Chap.3, Lemma 2)), it is monotone non decreasing in x.

The density $h(x; z)$ on (a, b) can be computed by deriving under the sign of integral

$$h(x; z) = \frac{d}{dx}P_x([z,\infty)) = \frac{d\theta}{dx}\int_{[z,\infty)}\frac{d}{d\theta}\exp\left(\theta t - M(\theta)\right)d\nu(t) = \frac{1}{V(x)}\int_{[z,\infty)}(t-x)dP_x(t),$$

recalling that $x = dM(\theta)/d\theta$ and $Var_\theta(Z) = d^2M(\theta)/d\theta^2$.

For $z \leq a$, it is clearly $H(x; z) = 1$ for any $x \in (a, b)$ so that $H(x; z) = 0$ for $x < a$ and $H(x; z) = 1$ for $x \geq a$, that is $H(\cdot; z)$ is degenerate on a.

For $z = b$, $H(x; z) = P_x([b,\infty))$ for $a < x < b$. Therefore, if P_x is absolutely continuous, $H(x; z) = 0$ for $x < b$, so that it is degenerate on b because, by definition,

$H(\cdot;z)$ is right-continuous. If P_x is discrete with mass $p_x(b)\nu(b)$ on b, using iii) of Lemma 1, we have

$$H(x;z) = \begin{cases} 0 & x \le a \\ p_x(b)\nu(b) & a < x < b \\ 1 & x \ge b \end{cases}$$

with $p_x(b)$ monotone increasing in x, and $\lim_{x\to b^-} p_x(b)\nu(b) = 1$. The expression of the relative density is a clear specialization of (3.6).

◇

Proof of Theorem 3.

The expression (3.10) corresponds to (3.9) by the definition (3.5) of $H_k(x;z)$ (see also formula (3.4)). By Theorem 2, the kernels $H_k(x;z)$ are d.f.'s, so that $B_{k,U}$ is a d.f.. We now study the mass concentrated at the extreme points a and b. Using Lemma 1, it is easy to verify that $\lim_{x\to a^+} B_{k,U}(x) = U(a)$ and

$$\lim_{x\to b^-} B_{k,U}(x) = \lim_{x\to b^-}\left(H_k(x;a)U(a) + \int_{(a,b)} H_k(x;z)dU(z) + H_k(x;b)U(\{b\})\right)$$
$$= U(a) + \int_{(a,b)} dU(z) + U(\{b\}) \lim_{x\to b^-} H_k(x;b)$$
$$= \begin{cases} U(a) + (1-U(a)-U(\{b\})) + 0 = 1 - U(\{b\}) & \text{if } \nu(b) = 0 \\ U(a) + (1-U(a)-U(\{b\})) + U(\{b\}) \lim_{x\to b^-} p_{k,x}(b)\nu\{b\} = 1 & \text{if } \nu\{b\} > 0. \end{cases}$$

For the continuity points $a < x < b$,

$$\frac{dB_{k,U}(x)}{dx} = \frac{d}{dx}(U(a) + \int_{(a,b)} H_k(x;z)dU(z) + H_k(x;b)U(\{b\})).$$

By Theorem 2, $H_k(\cdot;z)$ is absolutely continuous for $z \in (a,b)$ and $H_k(x;b)$ is either zero for $x < b$, or it is absolutely continuous. Therefore $h_k(\cdot,\cdot)$ is measurable and the result follows applying Fubini's Theorem.

◇

Proof of Lemma 2.

As a preliminary step, we define $Q_k^r(x) = E(Z_{k,x}^r) = \int_{(a,b)} z^r\, h_k(x;z)dz$ for $z \in (a,b)$

and $r \geq 0$ and show that

$$Q_k^r(x) = \frac{k}{r+1} V(x)^{-1} \left(E(Z_{k,x}^{r+2}) - x E(Z_{k,x}^{r+1}) \right). \tag{A.4}$$

For a continuous ERS and recalling the definition (3.6) of $h_k(x;z)$ for $x \in (a,b)$, an integration by parts of Q_k^r leads to

$$Q_k^r(x) = \frac{k}{r+1} V(x)^{-1} \left[z^{r+1} \int_z^b (t-x) dP_{k,x}(t) \Big|_a^b + \int_a^b z^{r+1} (z-x) dP_{k,x}(z) \right]. \tag{A.5}$$

The first term in the right hand side of (A.5) is trivially zero for $z \to a$ and $z \to b$ with a and b finite. If a or b are not finite, the same result follows applying the l'Hospital rule and recalling that $z^m p_{k,x}(z)\nu(z) \to 0$ for $y \to \mp\infty$ and each positive integer m, since the NEF admits moments of any order. The second addend in (A.5) gives directly the expression (A.4).

For a discrete ERS with measure ν_k having support on $\{a = z_1, z_2, \dots b\}$ (omitting for simplicity the possible dependence on k) with $b \leq \infty$, from (A.3) and noting $h_k(x; z_{i+1}) = \frac{d}{dx}(1 - P_{k,x}(z_i))$ we have

$$\begin{aligned} Q_k^r(x) &= \int_{(a,b)} z^r h_k(x;z) dz = \sum_{i=1}^b \int_{(z_i, z_{i+1}]} z^r h_k(x; z_{i+1}) dz = \sum_{i=1}^b h_k(x; z_{i+1}) \frac{z_{i+1}^{r+1} - z_i^{r+1}}{r+1} \\ &= \frac{1}{r+1} \left[z_1^{r+1} \frac{d}{dx} P_{k,x}(z_1) + \sum_{i=2}^b z_{i+1}^{r+1} \frac{d}{dx} (P_{k,x}(z_{i+1}) - P_{k,x}(z_i)) \right] \\ &= \frac{1}{r+1} \left[\sum_{i=1}^b z_i^{r+1} \frac{d}{dx} p_{k,x}(z_i) \nu(z_i) \right] = \frac{k}{r+1} V(x)^{-1} \left[\sum_{i=1}^b z_i^{r+1} (z_i - x) p_{k,x}(z_i) \nu(z_i) \right] \end{aligned}$$

which gives (A.5).

Consider $r = 0$. From (A.4), we obtain

$$Q_k^0(x) = \int_{(a,b)} h_k(x;z) dz = k V(x)^{-1} \left(E(Z_{k,x}^2) - x E(Z_{k,x}) \right) = 1$$

thus $h_k(x;z)$, which is clearly a positive function, is a density in z.

Consider now point (2). From (A.4), setting $r = 1$, we have

$$Q_1(x) = E(Z^*_{k,x}) = \frac{k}{2} V(x)^{-1} \left(E(Z^3_{k,x}) - xE(Z^2_{k,x}) \right). \tag{A.6}$$

Using the cumulant transform properties of the NEF, it is not difficult to show that $E(Z^3_{k,x}) = V(x)V'(x)/k^2 + E(Z^2_{k,x})x + 2xV(x)/k$. Using this expression in (A.6) gives (3.12). Setting now $r = 2$ in (A.4), and arguing similarly, we obtain (3.13).

To prove point (3), consider first a continuous ERS. Because

$$h_k(x; z) = k/V(x) \left(\int_{[z,x)} (t - x) dP_{k,\theta(x)}(t) + \int_{[x,+\infty)} (t - x) dP_{k,\theta(x)}(t) \right)$$

where the first addend is negative, the result follows easily. The proof is analogous for a discrete ERS, recalling from that in this case $h_k(x; z)$ is piecewise constant. ◇

A.2 Properties of the kernel $h_k(x; z)$

The following lemma studies the behavior of the kernel density $h_k(x; z)$ in the tails of x.

With no loss of generality, we assume that the natural parameter space Θ of the NEF used in the ERS includes zero. Indeed, let $\{P_\theta, \theta \in \Theta\}$ be a NEF parametrized in the natural parameter. If $0 \notin \Theta = (\alpha, \beta)$, choose $q \in \mathbb{R}$ such that $0 \in (\alpha - q, \beta - q) = \Theta^*$ and define $\theta^* = \theta - q$. Then we can write the d.f. $P_\theta(y)$ as

$$\begin{aligned} P_\theta(y) &= \frac{\int_{-\infty}^{y} \exp(\theta t - qt) e^{qt} \nu(dt)}{\int_{-\infty}^{+\infty} \exp(\theta t - qt) e^{qt} \nu(dt)} = \frac{\int_{-\infty}^{y} \exp(\theta^* t) \nu^*(dt)}{\int_{-\infty}^{+\infty} \exp(\theta^* t) \nu^*(dt)} \\ &= \int_{-\infty}^{y} \exp(\theta^* t - M^*(\eta)) \nu^*(dt) = P_{\theta^*}(y), \end{aligned}$$

where $\nu^*(dt) = e^{qt}\nu(dt)$ and M^* denotes the cumulant transform of ν^*. Consequently the families $\{P_\theta, \theta \in \Theta\}$ and $\{P_{\theta^*}, \theta^* \in \Theta^*\}$ are equivalent.

Lemma 3. *Let $k_0 \in \Lambda$, and fix $d_1, d_2 \in (a, b)$ such that $d_1 < d_2$ and $\theta(d_1) < 0$; $\theta(d_2) > 0$. Then there exist values $c^*, c^{**} \in [a, b]$, with $c^* < d_1, c^{**} > d_2$, $\nu_{k_0}([c^*, d_1)) > 0$, $\nu_{k_0}([d_2, c^{**}]) > 0$, $z^* \in (a, b)$ and $\delta \geq 0$ such that for any $k \in [k_0 - \delta, k_0] \subset \Lambda$ we have* *(i) for* $a < x < d_1$,

$$\frac{h_{k_0}(x;d_1)}{h_k(x;d_2)} \leq D_1(k,k_0)\ \exp\{\theta(x)(k_0 d_1 - kc^{**} - z^*(k_0-k))\} \tag{A.7}$$

(ii) for $d_2 < x < b$,

$$\frac{h_{k_0}(x;d_2)}{h_k(x;d_1)} \leq D_2(k,k_0)\ \exp\{\theta(x)(k_0 d_2 - kc^{*} - z^*(k_0-k))\}\ .$$

The functions $D_1(\cdot,k_0)$ *and* $D_2(\cdot,k_0)$ *are bounded for* $k \in [k_0-\delta,k_0]$.

Proof. Remind that we assume, with no loss of generality, that $0 \in \Theta$. It follows that $M(0) < \infty$ and consequently

$$\nu_k([a,b]) = \int_a^b e^{k0z}\nu_k(dz) = e^{kM(0)} < \infty. \tag{A.8}$$

(i) $a < x < d_1$.

Note that, from (3.6), we can write

$$\frac{h_{k_0}(x;d_1)}{h_k(x;d_2)} \quad = \quad \frac{k_0}{k}\ \frac{E(Z_{k_0,x}|Z_{k_0,x}\geq d_1)-x}{E(Z_{k,x}|Z_{k,x}\geq d_2)-x}\ \frac{P_{k_0,x}([d_1,\infty))}{P_{k,x}([d_2,\infty))}. \tag{A.9}$$

First we show that the second ratio in the right hand side of (A.9) is bounded. Indeed $E(Z_{k,x}|Z_{k,x} > z)$ is an increasing function of x, since

$$\frac{\partial}{\partial x}E(Z_{k,x}|Z_{k,x} > z) = \frac{k}{V(x)}Var(Z_{k,x}|Z_{k,x} > z) > 0 \quad \forall x \in (a,b),$$

and consequently $\sup_{x\in(a,d_1)} E(Z_{k_0,x}|Z_{k_0,x} > d_1) = B_{k_0,d_1}$, where B_{k_0,d_1} is a finite constant. Thus

$$0 \leq \frac{E(Z_{k_0,x}|Z_{k_0,x}\geq d_1)-x}{E(Z_{k,x}|Z_{k,x}\geq d_2)-x} \leq \frac{B_{k_0,d_1}-x}{d_2-x} \leq \frac{B_{k_0,d_1}-d_1}{d_2-d_1} < \infty$$

where the last inequality holds since $(B_{k_0,d_1}-x)/(d_2-x)$ is an increasing function of x.

Consider now the ratio $P_{k_0,\theta}((d_1,\infty))/P_{k,\theta}((d_2,\infty))$ where, for simplicity, we have used the natural parameterization. Let $c^{**} > d_2$ such that $\nu_{k_0}([d_2,c^{**}]) > 0$. If k_0 is not an isolated point of Λ, fix $\delta > 0$ so that $\nu_k([d_2,c^{**}]) > 0$ for any $k \in [k_0-\delta,k_0]$. This is always possible since $\nu_k(\cdot)$ is continuous in k, as can be easily checked from the continuity of $P_{k,\theta}$ shown in the proof of the following Lemma 5.

Recalling that $\theta(x)$ is an increasing function of x and, by assumption, $\theta(d_1) < 0$, we have $\theta(x) < \theta(d_1) < 0$, for $a < x < d_1$, so that, for $k \in [k_0 - \delta, k_0]$

$$\frac{P_{k_0,\theta}([d_1,\infty))}{P_{k,\theta}([d_2,\infty))} \le e^{-(k_0-k)M(\theta)} \frac{\int_{[d_1,\infty)} e^{k_0\theta z}\, d\nu_{k_0}(z)}{\int_{[d_2,c^{**})} e^{k\theta z}\, d\nu_k(z)} \le e^{-(k_0-k)M(\theta)} \frac{e^{k_0\theta d_1}}{e^{k\theta c^{**}}} \frac{\nu_{k_0}([d_1,\infty))}{\nu_k([d_2,c^{**}])} .$$

By (A.8) and since $\nu_k([d_2,c^{**}]) > 0$ for $k \in [k_0-\delta,k_0]$, $\nu_{k_0}([d_1,\infty))/\nu_k([d_2,c^{**}])$ is finite. For $k < k_0$, by inequality (2.4) in Diaconis and Ylvisaker's (1979), there exists a set G with $\nu(G) > 0$ and a value $z^* \in G$ such that

$$\exp(-(k_0-k)\, M(\theta)) \le \frac{e^{-(k_0-k)\theta\, z^*}}{\nu(G)^{k_0-k}} .$$

Therefore (A.7) holds, with

$$D_1(k,k_0) \;=\frac{k_0}{k}\, \frac{B_{k_0,d_1}-d_1}{d_2-d_1}\, \frac{\nu_{k_0}([d_1,\infty))}{\nu_k([d_2,c^{**}])}\, \nu(G)^{-(k_0-k)} . \tag{A.10}$$

Finally, as noticed before, ν_k is continuous at k_0, therefore $D_1(k,k_0)$ has a finite maximum for $k \in [k_0-\delta,k_0]$.

(ii) $d_2 < x < b$.

From (3.6) we can write

$$\frac{h_{k_0}(x;d_2)}{h_k(x;d_1)} \;=\; \frac{k_0}{k}\, \frac{x - E(Z_{k_0,x}|Z_{k_0,x} < d_2)}{x - E(Z_{k,x}|Z_{k,x} < d_1)}\, \frac{P_{k_0,x}((-\infty,d_2))}{P_{k,x}((-\infty,d_1))},$$

where $E(Z_{k,x}|Z_{k,x} < z))$ is an increasing function of x. Setting $\inf_{x\in(d_2,b)} E(Z_{k_0,x}|Z_{k_0,x} < d_2)) = A_{k_0,d_2}$, we have

$$\frac{x - E(Z_{k_0,x}|Z_{k_0,x} < d_2)}{x - E(Z_{k,x}|Z_{k,x} < d_1)} \le \frac{x-A}{x-d_1} \le \frac{d_2-A}{d_2-d_1} ,$$

where the last inequality holds since $(x-a)/(x-d_1)$ is descreasing in x. Proceeding as in point (i), it can be shown that there exist $c^* < d_1$ with $\nu_{k_0}([c^*,d_1)) > 0$ and $z^* \in (a,b)$ such that

$$\frac{P_{k_0,\theta}((-\infty,d_2))}{P_{k,\theta}((-\infty,d_1))} \le \frac{\nu_{k_0}((-\infty,d_2))}{\nu_k([c^*,d_1))}\, \nu(G)^{-(k_0-k)} \quad \exp(\theta[k_0 d_2 - kc^* - z^*(k_0-k)])$$

for any $k \in [k_0 - \delta, k_0]$, and the thesis follows. $\diamond$

Part (i) of the following lemma is of autonomous interest. According with the discussion in Section 3.3, it shows that the fiducial distribution of the natural parameter of a NEF has finite moments of any order.

Lemma 4. *(i) For any $z \in I\!R$ and $k \in \Lambda$, $\int \mid \theta(x) \mid^r dH_k(x;z) < \infty$ for any $r > 0$.*

(ii) If U has support included in (a,b), $\int \mid \theta(x) \mid^r b_{k,U}(x)dx < \infty$ for any $r > 0$.

Proof. From Theorem 2, for $z \leq a$, $z > b$ and $\{z = b,\ \nu(b) = 0\}$ (with a and b finite), $H_k(x;z)$ is degenerate and consequently the result holds. Consider now $z \in (a,b)$ or $\{z = b,\ \nu(b) > 0\}$ and assume, with no loss of generality, $k = 1$. We have to prove that for any integer r

$$\int_{(a,b)} |\theta(x)|^r dH(x;z)$$
$$= \int_{\{x:\theta(x)<0\}} (-\theta(x))^r dH(x;z) + \int_{\{x:\theta(x)\geq 0\}} \theta(x)^r dH(x;z) < \infty \,. \tag{A.11}$$

First, reparameterize the d.f. $H(x;z)$, defined by (A.2), in terms of the natural parameter and let $\{\theta \in \Theta : \theta < 0\} = (\alpha, \beta)$, with $\beta = \min(\sup \Theta, 0)$. If (α, β) is non empty, an integration by parts of the first integral in the right hand side of (A.11) leads to

$$\int_\alpha^\beta (-\theta)^r \frac{d}{d\theta} P_\theta([z,\infty))d\theta = [(-\theta)^r P_\theta([z,\infty))]_\alpha^\beta + r \int_\alpha^\beta (-\theta)^{r-1} P_\theta([z,\infty))d\theta \tag{A.12}$$

which is clearly finite for $\alpha > -\infty$ (since β is finite). If $\alpha = -\infty$, by the inequality (2.4) in Diaconis and Ylvisaker (1979) there exists a set A with $\nu(A) > 0$ and $t_A \in A$ such that

$$0 \leq \lim_{\theta \to -\infty} \int_{[z,+\infty)} (-\theta)^r \exp(\theta t - M(\theta)) d\nu(t) \leq \lim_{\theta \to -\infty} \frac{1}{\nu(A)} \int_{[z,+\infty)} (-\theta)^r \exp(\theta(t - t_A)) d\nu(t) \,.$$

Choosing $t_A < z$, and applying the Lebesgue dominated convergence theorem we conclude that the first term on the right hand side of (A.12) is finite. Consider now the last integral in (A.12). This is obviously finite if $\alpha > -\infty$. If $\alpha = -\infty$, applying again Diaconis and

Ylvisaker's inequality, is it possible to show that

$$\lim_{\theta\to-\infty} \frac{\theta^{r-1}P_\theta([z,\infty])}{1/\theta^2} = \lim_{\theta\to-\infty} \theta^{r+1}P_\theta([z,\infty)) = 0\,,$$

which is a sufficient condition to prove that the integral is finite in this case, too.
A similar argument shows that also the second integral in the right hand side of (A.11) is finite.

(ii). Let d_1, d_2 such that $a < d_1 < c_1$, $c_2 < d_2 < b$. Consider $I_{(a,b)} = \int_{(a,b)} |\theta(x)|^r b_{k,U_0}(x)dx$ and decompose it as the sum $I_{(a,b)} = I_{(a,d_1)} + I_{[d_1,d_2]} + I_{(d_2,b)}$. From part 3) of Lemma 2, for $x < d_1$, $h_k(x;\cdot)$ is decreasing on $[c_1, c_2]$, therefore

$$I_{(a,d_1)} = \int_{(a,d_1)} |\theta(x)|^r \int_{[c_1,c_2]} h_k(x;z)dU_0(z)dx \le \int_{(a,d_1)} |\theta(x)|^r \, h_k(x;c_1)dx\ ,$$

which is finite since, as a consequence of Proposition 4, $\int |\theta(x)|^r h_k(x;z)dx < \infty$. Analogously, it can be shown that $I_{(d_2,b)} < \infty$. Finally, $I_{[d_1,d_2]} < \infty$ since $\theta(x)^r$ is bounded on $[d_1, d_2]$.

$\diamond$

Lemma 5. *Let k_0 be in an interval in Λ. The kernel function $h_k(x;z)$ is continuous at k_0, i.e. for $k \to k_0, h_k(x;z) \to h_{k_0}(x;z)$, for any x and $z \in (a,b)$.*

<u>Proof</u>. First, we write $h_k(x;z)$ as

$$h_k(x,z) = \frac{k}{V(x)}\,(x - E(Z_{k,x}|Z_{k,x} < z))\,P_{k,x}((-\infty,z))$$

Let $S_{k,x} = kZ_{k,x} \sim P^*_{k,x}$ (see Section 3.1) and let $m^*_{k,x} = E(e^{tS_{k,x}}) = e^{k[M(t+\theta(x))-M(\theta(x))]}$ be the moment generating function of $S_{k,x}$, for t in a neighborhood on the origin. Clearly $\lim_{k\to k_0} m^*_{k,x}(t) = m^*_{k_0,x}(t)$ for any t and consequently $P^*_{k,x}$ converges weakly to $P^*_{k_0,x}$ as $k \to k_0$.
Now, if ν is dominated by the Lebesgue measure, then $P^*_{k_0,x}$ is continuous, so that $P^*_{k,x}(s) \to P^*_{k_0,x}(s)$ for any s and, by Slustky theorem, $P_{k,x}(z) \to P_{k_0,x}(z)$ for $k \to k_0$ and any z. If ν is discrete with aritmetic support, then convergence of the moment generating

functions implies pointwise convergence of the probability mass functions. Consequently, in both cases $P_{k,x}((-\infty, z)) \to P_{k_0,x}((-\infty, z))$ as $k \to k_0$, for any x and z.

Let now $\bar{Z}_{k,x}$ be the random variable $Z_{k,x}$ truncated at z, i.e. $\bar{Z}_{k,x} \sim Q_{k,x}(\cdot) = P_{k,x}(\cdot)/P_{k,x}((-\infty, z))$. For the previous results, $\bar{Z}_{k,x}$ converges to $\bar{Z}_{k_0,x}$ in distribution as $k \to k_0$. Furthermore, for k in a neighborhood J of k_0, $\{\bar{Z}_{k,x}, k \in J\}$ is uniformly integrable (since $\sup_{k\in J} E((\bar{Z}_{k,x})^2) \leq \sup_{k\in J} E(Z^2_{k,x}) = \sup_{k\in J}(x^2 + V(x)/k) < \infty$) and therefore $E(\bar{Z}_{k,x}) \to E(\bar{Z}_{k_0,x})$ (see e.g. Serfling (1980, p.14)). The thesis follows easily. $\diamond$

A.3 Continuity of $B_{k,U}$

Here we prove some general properties of $B_{k,U}$. The following Lemma shows a boundness property of the density $b_{k,U}$. Then we give some results of continuity of $B_{k,U}$ in its parameters (k, U). We write $U_n \Rightarrow U$ for denoting weak convergence of U_n to U and $U_n \to^{TV} U$ for convergence in total variation.

Lemma 6. *For any U, $b_{k,U}(x)$ is bounded and bounded away from zero for x and k in compact sets.*

Proof. For a known property of a NEF (see e.g. Lehmann (1959, Chap.2, Thm.9)) we can compute the derivative of $B_{k,U}(x) = E(U(Z_{k,x}))$ w.r.t. the natural parameter under the integral sign, obtaining

$$\begin{aligned} b_{k,U}(x) = \frac{d}{dx}B_{k,U}(x) &= \frac{d\theta}{dx}\frac{d}{d\theta}E(U(Z_{k,x})) = \frac{k}{V(x)}\int_{-\infty}^{\infty} U(z)(z-x)dP_{k,x}(z) \\ &= \frac{k}{V(x)}Cov(Z_{k,x}, U(Z_{k,x})), \end{aligned} \tag{A.13}$$

where we use $d\theta/dx = (dM'(\theta)/d\theta)^{-1}|_{\theta=\theta(x)} = V(x)^{-1}$.

From the Cauchy - Schwarz inequality and since $Var(U(Z_{k,x})) \leq 1$, it follows that

$$b_{k,U}(x) = \frac{k}{V(x)}Cov(Z_{k,x}, U(Z_{k,x})) \leq \left(\frac{k}{V(x)}\right)^{1/2}.$$

Because $V(x)$ is a continuous function in x, we have

$$\sup_{k\in C}\sup_{x\in[d_1,d_2]} b_{k,U}(x) \le \bar{q} < \infty,$$

where C is a closed interval in Λ, so that $b_{k,U}$ is bounded above on compact sets.

For proving that $b_{k,U}(x)$ is bounded away form zero for $x \in [d_1, d_2]$ and $k \in C$, notice that

$$\begin{aligned}
&Cov(Z_{k,x}, U(Z_{k,x}))\\
&= \int_{[a,x)} (x-z)(U(x)-U(z))dP_{k,x}(z) + \int_{[x,b]} (z-x)(U(z)-U(x))dP_{k,x}(z)\\
&\ge \int_{[a,a^*]} (x-z)(U(x)-U(z))dP_{k,x}(z) + \int_{[b^*,b]} (z-x)(U(z)-U(x))dP_{k,x}(z)\\
&\ge (x-a^*)(U(x)-U(a^*))P_{k,x}(a^*) + (b^*-x)(U(b^*)-U(x))P_{k,x}([b^*,b]) \qquad \text{(A.14)}
\end{aligned}$$

where $a^* = a$ and $b^* = b$ if, respectively, $P_{k,x}(a) > 0$ or $P_{k,x}(\{b\}) > 0$, otherwise choose $a^* < d_1$ such that $P_{k,x}(a^*) > 0$ and $b^* > d_2$ such that $P_{k,x}([b^*, b]) > 0$ $\forall k \in C$. Letting $\min(A, B) = A \wedge B$ and using (A.14), we obtain

$$\begin{aligned}
b_{k,U}(x) &\ge \frac{k}{V(x)}[(x-a^*)\wedge(b^*-x)][P_{k,x}(a^*)\wedge P_{k,x}([b^*,b])][U(b^*)-U(a^*)]\\
&= \underline{q_k}[U(b^*)-U(a^*)],
\end{aligned}$$

which is positive for any U not degenerate on a. Furthermore, as shown in the proof of Lemma 5, $P_{k,x}$ is continuous in k, so that, setting $\min_{k\in C} \underline{q}_k = \underline{q}$, we have

$$\inf_{k\in C}\inf_{x\in[d_1,d_2]} b_{k,U}(x) = \underline{q}[U(b^*)-U(a^*)] > 0. \qquad \text{(A.15)}$$

◇

Remark. Lemma 6 shows that, for x and k in compact sets, $b_{k,U}(x)$ can be majorated independently of U. On the contrary, the lower bound in (A.15) depends on U. However, it is greater or equal to $\alpha\,\underline{q}$ for any U in a class of d.f.'s such that $U([c_1, c_2]) > \alpha$ for $[c_1, c_2] \subset (a^*, b^*)$.

For brevity, in the following we limit our attention to the case when U_n and U_0 have no mass concentrated at the extremes $\{a\}$ and $\{b\}$, so that B_{k,U_n} and B_{k,U_0} are absolutely continuous d.f., with densities b_{k,U_n} and b_{k,U_0}.

Proposition 4. *Let $B_{k,U}$ denote the completed Feller operator for a d.f. U, with ERS $\{P_{k,x}, k \in \Lambda, x \in (a,b)\}$ and let $(U_n, n \geq 1)$ be a sequence of d.f.'s.*

(a) Continuity in U. *If U_n converges weakly to a continuous d.f. U_0, then b_{k,U_n} converges to b_{k,U_0} pointwise and $B_{k,U_n} \to^{TV} B_{k,U_0}$.*

(b) Continuity in k. *Let k_0 be in an interval included in Λ. If $k \to k_0$, then $b_{k,U}$ converges to $b_{k_0,U}$ pointwise and $B_{k,U} \to^{TV} B_{k_0,U}$.*

(c) Continuity of $B_{k,U}$. *Let k_0 be in an interval included in Λ. Then*

(i) if $k \to k_0$, $U_n \Rightarrow U_0$ and U_0 is continuous, then $B_{k,U_n} \Rightarrow B_{k_0,U_0}$ uniformly in x;

(ii) if $k \to k_0$ and $U_n \to^{TV} U_0$, then $B_{k,U_n} \to^{TV} B_{k_0,U_0}$.

Proof. (a). We have $b_{k,U_n}(x) = \int h_k(x;z)dU_n(z)$. The functions $h_k(x,\cdot)$ and U_0 have no common discontinuity points, since U_0 is continuous. Moreover, we have $h_k(x,z) \leq h_k(x,x)$ from point 3) of Lemma 2, therefore $h_k(x,\cdot)$ is bounded for any $x \in (a,b)$. Thus, using the Helly-Bray theorem,

$$\lim_{n\to\infty} b_{k,U_n}(x) = \lim_{n\to\infty} \int h_k(x;z)dU_n(z) = \int h_k(x;z)dU_0(z) = b_{k,U_0}(x),$$

for $x \in (a,b)$. It follows, by Scheffé theorem, that $B_{k,U_n} \to^{TV} B_{k,U_0}$ for $n \to \infty$.
(b). By Lemma 6, $b_{k,U}(x)$ is bounded for k in a compact set, for any fixed x and U. Taking U degenerate on x it follows that $h_k(x,x)$ is bounded for $k \in (k_0 - c, k_0 + c) \subset \Lambda$, $c > 0$. Thus, since $h_k(x,z) \leq h_k(x,x)$, the dominated convergence theorem can be applied, and we have

$$\lim_{k\to k_0} b_{k,U}(x) = \lim_{k\to k_0} \int h_k(x;z)dU(z) = \int \lim_{k\to k_0} h_k(x;z)dU(z) = b_{k_0,U}(x).$$

By Scheffé theorem, it follows that $B_{k,U} \to^{TV} B_{k_0,U}$ for $k \to k_0$.

(c). For proving (i), notice that

$$| B_{k,U_n}(x) - B_{k_0,U_0}(x) | \leq | B_{k,U_n}(x) - B_{k,U_0}(x) | + | B_{k,U_0}(x) - B_{k_0,U_0}(x) | . \quad \text{(A.16)}$$

The first addend in the right hand side can be made smaller than $\epsilon/2$ for $n > \bar{n}(\epsilon/2)$, independently on k since

$$| B_{k,U_n}(x) - B_{k,U_0}(x) | \leq \int | U_n(z) - U_0(z) | \, dP_{k,x}(z) \leq \sup_z | U_n(z) - U_0(z) |, \quad \text{(A.17)}$$

and $U_n \Rightarrow U_0$ uniformly, because U_0 is a continuous d.f.. The second addend in (A.16) can be made smaller than $\epsilon/2$ for k sufficiently close to k_0, by point (b). Thus $B_{k,U_n}(x) \to B_{k_0,U_0}(x)$ for any x. The convergence is uniform in x since B_{k_0,U_0} is a continuous d.f..

For showing (ii), consider

$$\sup_A | B_{k,U_n}(A) - B_{k_0,U_0}(A) | \leq \sup_A | B_{k,U_n}(A) - B_{k,U_0}(A) | + \sup_A | B_{k,U_0}(A) - B_{k_0,U_0}(A) | . \quad \text{(A.18)}$$

For the first addend, notice that

$$\begin{aligned} \sup_A |B_{k,U_n}(A) - B_{k,U_0}(A)| &= \sup_A | \int H_k(A;z)dU_n(z) - \int H_k(A;z)dU_0(z)| \\ &\leq \sup_{|g|<1} | \int g dU_n - \int g dU_0| = d_{TV}(U_n, U_0), \end{aligned} \quad \text{(A.19)}$$

where $d_{TV}(U_n, U_0)$ denotes the total variation distance between U_n and U_0. Therefore it can be made smaller than $\epsilon/2$ for $n > \bar{n}(\epsilon/2)$, independently on k, since by assumption $U_n \to^{TV} U_0$. The second addend of (A.18) can be smaller than $\epsilon/2$ for $| k - k_0 | < \eta(\epsilon/2)$, by point (b), and the thesis follows. ◇

Remark. The assumption of continuity of U_0 in part (a) of the above proposition is not necessary if the ERS is continuous. In fact, for (a) it is enough to assume that U_0 and $P_{k,x}$ have no common discontinuity points and $U_n(a) \to U_0(a) = 0$ and $U_n(b) \to U_0(b) = 1$. For a discrete ERS, the kernel $h_k(x;z)$ is piecewise constant, and the result (a) holds providing that $U_n(z_{j,k}) \to U_0(z_{j,k})$ for any support point $z_{j,k}$ of $P_{k,x}$.

The following result provides an extension of Feller's Theorem 1 and of Theorem 4.

Proposition 5. (Approximation properties). *Let $B_{k,U}$ and U_n be defined as in Proposition 4.*

(i) If $k \to \infty$ and $U_n \Rightarrow U_0$, where U_0 is a continuous d.f., then $B_{k,U_n} \Rightarrow U_0$, uniformly in x.

(ii) If $k \to \infty$ and $U_n \to^{TV} U_0$, where U_0 is an absolutely continuous d.f. with continuous and bounded density u_0, then $B_{k,U_n} \to^{TV} U_0$.

Proof. (i). We have

$$|B_{k,U_n}(x) - U_0(x)| \leq |B_{k,U_n}(x) - B_{k,U_0}(x)| + |B_{k,U_0}(x) - U_0(x)|.$$

The first addend can be made smaller than $\epsilon/2$ for $n > \bar{n}(\epsilon/2)$, independently on k, as shown from (A.17). The second addend can be made smaller than $\epsilon/2$ for $k > \bar{k}(\epsilon/2)$ by Theorem 1. The convergence is uniform being U_0 a continuous d.f..
(ii). Write

$$\sup_A |B_{k,U_n}(A) - U_0(A)| \leq \sup_A |B_{k,U_n}(A) - B_{k,U_0}(A)| + \sup_A |B_{k,U_0}(A) - U_0(A)|.$$

The first addend can be made smaller than $\epsilon/2$ for $n > \bar{n}(\epsilon/2)$, independently on k, as shown from (A.19). The second addend can be made smaller than $\epsilon/2$ for $k > \bar{k}(\epsilon/2)$ by Theorem 4.

$\diamond$

Finally, we give a result of continuity in Kullback-Leibler of the density $b_{k,U}$.

Lemma 7. *Let k_0 be in an interval contained in Λ. If U_0 has support included in (a,b), then*

$$lim_{k \uparrow k_0} KL(b_{k_0,U_0}, b_{k,U_0}) = 0 \ .$$

Proof. For brevity, let $b_0(x) = b_{k_0,U_0}(x)$. Denote by $[c_1, c_2]$ the support of U_0. Fix $d_1 < c_1$ and $d_2 > c_2$ in (a,b) satisfying the assumptions of Lemma 3 , i.e. $\theta(d_1) < 0 < \theta(d_2)$. Then we can write

$$
\begin{aligned}
0 &\leq KL(b_0, b_{k,U_0}) \\
&= \int_{(a,d_1)} \log \frac{b_0(x)}{b_{k,U_0}(x)}\, b_0(x)dx + \int_{[d_1,d_2]} \log \frac{b_0(x)}{b_{k,U_0}(x)}\, b_0(x)dx + \int_{(d_2,b)} \log \frac{b_0(x)}{b_{k,U_0}(x)}\, b_0(x)dx \ .
\end{aligned}
$$

It suffices to consider the case when the integrals on the right hand side are positive (otherwise, we can majorize $KL(b_0, b_{k,U_0})$ by omitting the negative addends). For $a < x < d_1$, from part 3) of Lemma 2,

$$
h_k(x; d_2) \leq b_{k,U_0}(x) = \int_{[c_1,c_2]} h_k(x; z)dU_0(z) \leq h_k(x; d_1) \ .
$$

By Lemma 3, there exists a left neighborhood of k_0 such that for any $k \in [k_0 - \delta, k_0]$,

$$
\log \frac{b_0(x)}{b_{k,U_0}(x)} \leq \log \frac{h_{k_0}(x; d_1)}{h_k(x; d_2)} \leq \log D_1 + |\theta(x)|\, G \ ,
$$

where $D_1 = |\max_{k\in[k_0-\delta,k_0]} D_1(k, k_0)|$ and $G = \max_{k\in[k_0-\delta,k_0]} |k_0 d_1 - kc^{**} - z^*(k_0 - k)|$. Therefore,

$$
\int_{(a,d_1)} \log \frac{b_0(x)}{b_{k,U_0}(x)}\, b_0(x)dx \leq (\log D_1 + G) \int_{(a,d_1)} \max(1, |\theta(x)|)\, b_0(x)dx \ .
$$

By part (ii) of Lemma 4, $\int_{(a,b)} |\theta(x)|\, b_0(x)dx < \infty$. So, for any $\epsilon > 0$, there exists a value of d_1 sufficiently close to a such that $\int_{(a,d_1)} \max(1, |\theta(x)|)b_0(x)dx$ is smaller than $\epsilon/(3(\log D_1 + G))$, and $\int_{(a,d_1)} \log(b_0(x)/b_{k,U_0}(x))\, b_0(x)dx < \epsilon/3$.

The integral on the right hand tail can be treated analogously, using (ii) of Lemma 3.

Finally, consider $x \in [d_1, d_2]$. By Lemma 6, $|\log(b_0(x)/b_{k,U_0}(x))|$ is bounded $k \in [k_0 - \delta, k_0]$, so we can apply the dominated convergence theorem, obtaining

$$
\lim_{k\to k_0} \int_{[d_1,d_2]} \log \frac{b_0(x)}{b_{k,U_0}(x)}\, b_0(x)dx = \int_{[d_1,d_2]} \lim_{k\to k_0} \log \frac{b_0(x)}{b_{k,U_0}(x)}\, b_0(x)dx = 0 \ ,
$$

being $\lim_{k\to k_0} b_{k,U}(x) = b_{k_0,U}(x)$ by (b) of Proposition 4. Thus the integral on $[d_1, d_2]$ can be made smaller than $\epsilon/3$ for k in a (left)-neighborhood of k_0, and this concludes the proof.

◇

A.4 Support of the mixture prior

Proof of Theorem 5.

(i). By part (i) of Proposition 5, for any $\epsilon > 0$ we can choose $\bar{k} = \bar{k}(\epsilon/2)$ and a weak neighborhood $W(F_0) = W_{\epsilon/2}(F_0)$ of F_0 such that $B_{k,U}$ is in a weak neighborhood $W_\epsilon(F_0)$ for any $k \geq \bar{k}$ and $U \in W(F_0)$. Therefore

$$\pi_B(W_\epsilon(F_0)) \geq \int_{k \geq \bar{k}} \pi_U(W(F_0) \mid k) dp(k),$$

which is positive being $\pi_U(W(F_0) \mid k) > 0$ for any k and $p(k)$ positive by assumption.

(ii). For any $\omega \in \Omega$, we have

$$\sup_A |B_{k,F_n}(A) - F_0(A)| \leq \sup_A |B_{k,F_n}(A) - B_{k,F_0}(A)| + \sup_A |B_{k,F_0}(A) - F_0(A)|.$$

By Theorem 4 we can choose $\bar{k} = \bar{k}_{\epsilon/2}$ sufficiently large so that $\sup_A |B_{k,F_0}(A) - F_0(A)|$ is smaller than $\epsilon/2$ for $k > \bar{k}$. For any such k, by part (a) of Proposition 4 we can choose a weak neighborhood $W_{\epsilon/2,k}(F_0))$ such that the first addend is smaller than $\epsilon/2$ for any $U \in W_{\epsilon/2,k}(F_0)$. Therefore

$$\pi_B(V_\epsilon(F_0)) \geq \int_{k \geq \bar{k}} \pi_U(W_{\epsilon/2,k}(F_0) \mid k) dp(k)$$

which is positive being $\pi_U(W_{\epsilon/2,k}(F_0) \mid k) > 0$ for any k and $p(k)$ positive by assumption.

$\diamond$

Proof of Theorem 6.

Remind that, as shown in Appendix A.2, given a NEF with natural parameter $\theta \in \Theta$, we can assume that $0 \in \Theta$ with no loss of generality.

Let $f_0 = b_{k_0,U_0}$ be the density of $F_0 = B_{k_0,U_0}$ (by assumption, U_0 has support included in (a,b), so that $U_0(a) = 0$ and $U_0(\{b\}) = 0$ and b_{k_0,U_0} is a density). Denote by $KL(f,g) = \int \log(f(x)/g(x))\ f(x)dx$ the Kullback-Leibler divergence between two densities f and g.

We can write

$$\begin{aligned} KL(f_0, b_{k,U}) &= \int_{(a,b)} \log \frac{f_0(x)}{b_{k,U_0}(x)}\, f_0(x)dx + \int_{(a,b)} \log \frac{b_{k,U_0}(x)}{b_{k,U}(x)}\, f_0(x)dx \\ &= KL(f_0, b_{k,U_0}) + I(k,U)\ . \end{aligned} \tag{A.20}$$

If k_0 is an isolated point of Λ, fix $k = k_0$, so that $KL(f_0, b_{k_0,U_0}) = 0$. If k_0 is in an interval included in Λ, by Lemma 7 in Appendix A.3, for any $\epsilon > 0$ there exists $\delta = \delta(\epsilon)$ such that $KL(f_0, b_{k,U_0}) < \epsilon/2$ for $k \in [k_0 - \delta, k_0]$. If the integral $I(k,U)$ on the right hand side of (A.20) is negative, then $0 \leq KL(f_0, b_{k,U}) \leq KL(f_0, b_{k,U_0}) < \epsilon/2$. So it remains to consider the case when $I(k,U) > 0$, with $k \in [k_0 - \delta, k_0]$, $\delta \geq 0$ (with $\delta = 0$ if k_0 is an isolated point of Λ).

Let $[c_1, c_2]$ be the support of U_0. Fix $d_1 < c_1$, $d_2 > c_2$ in (a,b) satisfying the assumptions of Lemma 3. Then, decompose the integral $I(k,U)$ as the sum of the integrals on $(a, d_1), [d_1, d_2], (d_2, b)$.

For the left-tail integral on (a, d_1), using part 3) of Lemma 2 and (i) of Lemma 3, we have

$$\begin{aligned} \int_{(a,d_1)} \log \frac{b_{k,U_0}(x)}{b_{k,U}(x)} f_0(x)dx &\leq \int_{(a,d_1)} \log \frac{h_k(x; d_1)}{h_k(x; d_2)\, U([c_1, c_2])}\, f_0(x)dx \\ \leq \ \{\log D_1 + G - \log U([c_1, c_2])\} &\int_{(a,d_1)} \max(1, |\theta(x)|) f_0(x)dx \end{aligned}$$

where: $D_1 = |\sup_{k\in[k_0-\delta,k_0]} D_1(k,k)|$, with $D_1(k,k)$ defined as in (A.10); note that $D_1 < \infty$ since ν_k is continuous in k; and $G = \max_{k\in[k_0-\delta,k_0]} k(c^{**} - d_1)$. Now, $\log D_1 + G - \log U([c_1, c_2]) < \log D_1 + G + \log 2$ for U in a weak neighborhood of U_0 such that $U([c_1, c_2]) > \frac{1}{2}$. Then, by part (ii) of Lemma 4 we can choose d_1 sufficiently close to a so that $\int_a^{d_1} \max(1, |\theta(x)|)\, f_0(x)dx$ is sufficiently small, and the integral

$$\int_{(a,d_1)} f_0(x) \log(b_{k,U_0}(x)/b_{k,U}(x))dx$$

is smaller than $\epsilon/6$. The right tail can be treated analogously.

Let now consider $\int_{[d_1,d_2]} \log(b_{k,U_0}(x)/b_{k,U}(x))\, f_0(x)dx$. From Lemma 6 and the related Remark, it follows that, for $x \in [d_1, d_2]$, $k \in [k_0 - \delta, k_0]$ and any U such that

$U([c_1, c_2])) > 1/2$, $b_{k,U}(x)$ is bounded and bounded away from zero. Then, if $\{U_n\}$ is a sequence of d.f.'s converging weakly to U_0, with $U_n([c_1, c_2])) > 1/2$ for each n, we have $|\log(b_{k,U_0}(x)/b_{k,U_n}(x))| < constant$; by dominated convergence

$$\lim_{n\to\infty} \int_{[d_1,d_2]} \log \frac{b_{k,U_0}(x)}{b_{k,U_n}(x)} f_0(x)dx = \int_{[d_1,d_2]} \lim_{n\to\infty} \log \frac{b_{k,U_0}(x)}{b_{k,U_n}(x)} f_0(x)dx.$$

Assume for the moment that $U_0(\cdot)$ and $h_k(x;\cdot)$ have no common discontinuity points. Then, by part (a) of Proposition 4 and the related Remark, b_{k,U_n} converges pointwise to b_{k,U_0}, so that

$$\int_{[d_1,d_2]} \lim_{n\to\infty} \log \frac{b_{k,U_0}(x)}{b_{k,U_n}(x)} f_0(x)dx = 0.$$

Since the weak topology is metrizable, the above result is equivalent to say that for any $\epsilon > 0$ there exists a weak neighborhood $W_{\epsilon,k}(U_0)$ such that $\int_{[d_1,d_2]} \log(b_{k,U_0}(x)/b_{k,U}(x))\, f_0(x)dx$ is smaller than $\epsilon/6$ for any $U \in W_{\epsilon,k}(U_0)$. Thus we conclude that, given $k \in [k_0 - \delta, k_0]$, for any $\epsilon > 0$ we can choose a neighborhood $\mathcal{W}_{k,\epsilon}(U_0)$ of U_0 such that $I(k, U) < \epsilon/2$ for $U \in \mathcal{W}_{k,\epsilon}(U_0)$.

Finally, from (A.20), if k_0 is an isolated point of Λ, the prior probability of a Kullback-Leibler neighborhood of f_0 is

$$\begin{aligned}\pi_B(\{B_{k,U} : KL(f_0, b_{k,U})) < \epsilon\}) &\geq \pi_B(\{B_{k,U} : k = k_0, U \in \mathcal{W}_{k_0,\epsilon}(U_0)\}) \\ &= \pi_U(\mathcal{W}_{k_0,\epsilon}(U_0)|k_0)\, p(k_0),\end{aligned}$$

which is positive by the assumptions. If k_0 belongs to an interval included in Λ, then

$$\pi_B(\{B_{k,U} : KL(f_0, b_{k,U}) < \epsilon\}) \geq \int_{[k_0-\delta,k_0]} \pi_U(\mathcal{W}_{k,\epsilon}(U_0)|k)\, p(k)dk > 0\ ,$$

being the integral of a positive function on an interval with positive length.

The theorem is proved under the restriction that U_0 and $h_k(x;\cdot)$ have no common discontinuities. This is always true if the ERS is continuous, since in this case $h_k(x;z)$ is continuous in z. If the ERS is discrete, the assumptions on the measure ν imply that ν_k has a finite number of support points, $\{z_{j_1,k}, \ldots, z_{j_m,k}\}$ say, in the closed interval $[c_1, c_2]$, (see Ramachandran (1967, Chap.1 and 2)).

Let $z_{j_0,k} < c_1$ and $z_{j_{m+1},k} \geq c_2$, so that $U_0(z_{j_0,k}) = 0$ and $U_0(z_{j_{m+1},k}) = 1$. Then

$$b_{k,U_0}(x) = \sum_{i=0}^{m} \left(U_0(z_{j_{i+1},k}) - U_0(z_{j_i,k})\right)\ h_k(x; z_{i,k}).$$

Let $U_0^{(k)}$ be a continuous d.f. such that $U_0^{(k)}(z_{j_i,k}) = U_0(z_{j_i,k})$, $i = 0, 1, \ldots, m+1$. Then $b_{k,U_0^{(k)}}(x) = b_{k,U_0}(x)$ and, if U_n converges weakly to $U_0^{(k)}$ as $n \to \infty$, $b_{k,U_n}(x) \to b_{k,U_0^{(k)}}(x) = b_{k,U_0}(x)$, by (a) of Proposition 4. Therefore the integral $\int_{[d_1,d_2]} \log(b_{k,U_0}(x)/b_{k,U}(x)) f_0(x) dx$ can be made sufficiently small for U in a weak neighborhood $\mathcal{W}_{k,\epsilon}(U_0^{(k)})$ of $U_0^{(k)}$. It follows that

$$\pi_B(\{B_{k,U} : KL(f_0, b_{k,U}) < \epsilon\}) \geq \int_{(k_0-\delta,k_0]} \pi(\mathcal{W}_{k,\epsilon}(U_0^{(k)})|k)\ p(k)\, dk > 0,$$

which is positive since by assumption $\pi_U(\cdot|k)$ has full weak support for any k, so that $\pi(\mathcal{W}_{k,\epsilon}(U_0^{(k)})|k) > 0$, and $p(k)$ is positive.

◇

Proof of Theorem 7.

Denote with E the finite support of f_0 and write

$$KL(f_0, b_{k,U}) = \int_E \log \frac{f_0(x)}{b_{k,F_0}(x)} f_0(x) dx + \int_E \log \frac{b_{k,F_0}(x)}{b_{k,U}(x)} f_0(x) dx \ . \qquad \text{(A.21)}$$

Assume first that f_0 is bounded away from zero, so that there exist positive constants m and M such that $0 < m \leq f_0(x) \leq M < \infty$, for $x \in E$. Now, we have

$$b_{k,F_0}(x) = E(f_0(Z^*_{k,x})) = E(f_0(Z^*_{k,x})|Z^*_{k,x} \in E) Pr(Z^*_{k,x} \in E)\ ,$$

where $Z^*_{k,x}$ was defined in Lemma 2, and recalling that $f_0(z) = 0$ for $z \notin E$. Then by the internality of the expected value,

$$m Pr(Z^*_{k,x} \in E) \leq b_{k,F_0}(x) \leq M\ .$$

Now, assume that $x \in E$. Then, from Lemma 2, the density of $Z^*_{k,x}$ has a mode in a point of E and thus $Pr(Z^*_{k,x} \in E) > 0$, for each fixed x and k. Furthermore, from (3.12) and

(3.13), it follows that $Pr(Z^*_{k,x} \in E) \to 1$, for $k \to \infty$. Thus

$$|\log \frac{f_0(x)}{b_{k,F_0}(x)}| \leq \text{ constant } < \infty, \quad \forall x \in E, \text{ and } k > k^*, \ k^* \in \Lambda.$$

Therefore, by dominated convergence and using the fact that $b_{k,F_0}(x) \to f_0(x)$, for $k \to \infty$, by Theorem 4, we have

$$\lim_{k\to\infty} \int_E \log \frac{f_0(x)}{b_{k,F_0}(x)} f_0(x)dx = 0 \ .$$

Thus, for any $\epsilon > 0$ the first integral on the right side of (A.21) can be made $< \epsilon/2$ for $k > \bar{k} = \bar{k}(\epsilon/2)$. Now, if Λ contains an isolated point $k_0 > \bar{k}$, fix $k = k_0$ in the second integral in the right hand side of (A.21), otherwise fix k_0 and $\delta > 0$ such that $k_0 - \delta > \bar{k}$ and let $k \in [k_0 - \delta, k_0]$. For each such k's, the second integral on the right hand side of (A.21) can be treated as the integral $I(k, U)$ appearing in (A.20), with F_0 in place of U_0. Thus, following a similar proof as for Theorem 6, it can be shown that for any fixed $k \in [k_0 - \delta, k_0]$, it can be made smaller than $\epsilon/2$ for U in a weak neighborhood $\mathcal{W}_{\epsilon,k}(F_0)$. It follows that, if k_0 is an isolated point in Λ,

$$\pi_B(\{B_{k,U} : KL(f_0; b_{k,U}) < \epsilon\}) \geq \pi_U(W_{k_0,\epsilon}(F_0)|k_0)p(\{k_0\}) \ ,$$

which is positive by the assumptions; otherwise

$$\pi_B(\{B_{k,U} : KL(f_0; b_{k,U}) < \epsilon\}) \geq \int_{[k_0-\delta,k_0]} \pi_U(W_{k,\epsilon}(F_0)|k) \ p(k)dk \ ,$$

which is also positive. This proves the thesis under the assumption that f_0 is bounded away from zero.

In the general case, we can use Lemma 5.1 in Ghosal, Ghosh and Ramamoorthi (1999), which shows that, if f_0 and f_1 are densities such that $f_0 \leq C f_1$ for a constant $C > 0$, then for any density f

$$KL(f_0, f) \leq (C + 1) \log C + C\{KL(f_1, f) + \sqrt{KL(f_1, f)}\} \ .$$

Let $\alpha > 0$ and $f_\alpha(x) = \max(f_0(x), \alpha)/\int \max(f_0(x), \alpha)dx = \max(f_0(x), \alpha)/C_\alpha$. Being $f_0 \leq C_\alpha f_\alpha$, we have by the above Lemma

$$KL(f_0, b_{k,U}) \leq (C_\alpha + 1)\log C_\alpha + C_\alpha\{KL(f_\alpha, b_{k,U}) + \sqrt{KL(f_\alpha, b_{k,U})}\}.$$

Fix α sufficiently small, so that C_α is close to one and $(C_\alpha + 1)\log C_\alpha$ is small. Because the density f_α is bounded and bounded away from zero, we can use the result established above. Therefore, denoting with F_α the d.f. associated to f_α, we can find for any $\epsilon > 0$, weak neighborhoods $\mathcal{W}_{\epsilon,k}(F_\alpha)$ and a value k_0 such that, given $k \in [k_0 - \delta, k_0]$, $KL(f_\alpha, b_{k,U})$ is small for every $U \in \mathcal{W}_{\epsilon,k}(F_\alpha)$. Thus

$$\pi_B(\{B_{k,F} : KL(f_0; b_{k,U}) < \epsilon\}) \geq \int_{[k_0-\delta,k_0]} \pi_U(W_{k,\epsilon}(F_\alpha)|k)\; dp(k),$$

which is positive by the assumptions.

◇

Proof of Proposition 3.

Observe that $B_{k,U} = B_{k,Q}$ implies $U(a) = Q(a)$ and, for all $x \in (a, b)$

$$\int (U(z) - Q(z))dP_{k,x}(z) = 0.$$

Since the NEF $P_{k,x}$ is complete, it follows that

$$P_{k,x}(U(Z_{k,x}) - Q(Z_{k,x})) = 1. \tag{A.22}$$

Therefore, if $P_{k,x}$ is absolutely continuous, it must be $U(z) = Q(z)$ for all $z \in (a, b)$ except at most on subsets of Lebesgue measure zero. But, being U and Q d.f.'s, this implies that $U(z) = Q(z)$ for all $z \in (a, b)$. If $P_{k,x}$ is discrete, (A.22) implies that $U(z_{j,k}) = Q(z_{j,k})$ at any support point $z_{j,k}$ of $P_{k,x}$ and therefore the identifiability of the weights $w^U_{j,k}$.

◇

Additional references

Cifarelli, D.M. and Regazzini, E. (1987). Priors for exponential families which maximize the association between past and future observations. In *Probability and Bayesian Statistics* (Edited by R. Viertl), 83-95. Plenum Press, New York.

Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7**, 269-281.

Ghosal, S., Ghosh, J.K. and Ramamoorthi, R.V. (1999). Consistent semi-parametric Bayesian inference about a location parameter. *J. Statist. Plann. Inference* **77**, 181-193.

Ramachandran, B. (1967). *Advanced Theory of Characteristic Functions.* Statistical Publishing Society, Calcutta.

Serfling, J., R. (1980). *Approximation Theory of Mathematical Statistics.* J. Wiley, New York.

Department of Decision Sciences

E-mail: sonia.petrone@unibocconi.it

E-mail: piero.veronese@unibocconi.it