Statistica Sinica **18**(2008): Supplement, S47∼S57

ON THE CHOICE OF m IN THE n OUT OF n BOOTSTRAP AND CONFIDENCE BOUNDS FOR EXTREMA

Peter J. Bickel and Anat Sakov

University of California, Berkeley and Tanor Group

Supplementary Material

2. Appendix to Section 3

2.1. Proof of Theorem 1

For fixed $S < \infty$, by assumptions (A.1) and (A.4),

$$\min_{1\le s\ne k\le S}\Big\|L^*_{s,n}(\cdot)-L^*_{k,n}(\cdot)\Big\|_\infty \xrightarrow{p} \min_{1\le s\ne k\le S}\Big\|L_s(\cdot)-L_k(\cdot)\Big\|_\infty > 0.$$

On the other hand, let $\langle\sqrt{n}\rangle$ be the m_j closest to $\sqrt{n}$ and similarly define $\langle q\sqrt{n}\rangle$ then by (A.3):

$$\Big\|L^*_{\langle\sqrt{n}\rangle,n}(\cdot)-L^*_{\langle q\sqrt{n}\rangle,n}(\cdot)\Big\|_\infty \xrightarrow{p} 0. \tag{21}$$

Therefore,

$$\hat{m} \xrightarrow{p} \infty. \tag{22}$$

If $\tilde{L}^*_{m,n}$ corresponds to $\tilde{U}_n(\cdot)$, just as $L^*_{m,n}$ corresponds to $U_n(\cdot)$, define $\tilde{\hat{j}}$ in relation to $\{\tilde{L}^*_{m,n}\}$ as $\hat{j}$ is defined for $\{L^*_{m,n}\}$. Then, by construction of $\hat{m}$, for each J, and for n sufficiently large:

$$P(\tilde{\hat{j}}>J)\ge P\Big(\min_{1\le j\le J}\Big\|\tilde{L}^*_{m_j,n}(\cdot)-\tilde{L}^*_{m_{j+1},n}(\cdot)\Big\|_\infty > \Big\|\tilde{L}^*_{\langle\sqrt{n}\rangle,n}(\cdot)-\tilde{L}^*_{\langle q\sqrt{n}\rangle,n}(\cdot)\Big\|_\infty\Big). \tag{23}$$

If condition (a) (Appendix A), for convergence in law of U_n, holds in (A.5), then

$$\min_{1\le j\le J}\Big\|\tilde{L}^*_{m_j,n}(\cdot)-\tilde{L}^*_{m_{j+1},n}(\cdot)\Big\|_\infty \overset{\mathcal{L}}{\Rightarrow} \min_{1\le j\le J}\Big\|U(q^j)(\cdot)-U(q^{j+1})(\cdot)\Big\|_\infty. \tag{24}$$

Finally, by (21), (23) and (24),

$$\liminf_n P(\tilde{\hat{j}} > J) \ge P\Big(\min_{1\le j\le J}\Big\|U(q^j)(\cdot)-U(q^{j+1})(\cdot)\Big\|_\infty > 0\Big).$$

Therefore, for each J, $P(\tilde{\jmath} > J) \to 1$. But, if (a) holds for (A.5), by construction, $\tilde{\jmath}$ and $\hat{\jmath}$ have the same distribution. Hence,

$$\frac{\hat{m}}{n} \xrightarrow{p} 0. \tag{25}$$

The result follows from (22) and (25).

Similarly, if (a') holds for (A.5) (Appendix A), then $\|\tilde{U}_n(\cdot) - U_n(\cdot)\|_\infty = o_p(1)$. It follows that $P(\tilde{\jmath} \neq \hat{\jmath}) \to 0$, and (25) follows for this case as well.

This completes the proof.

2.2. Proof of Theorem 2

(a) From assumptions (A.1)–(A.4), it follows that under F, $\hat{m} \xrightarrow{p} \infty$ (Theorem 1). Under the additional assumptions:

$$\begin{aligned} &\left|\left|L^*_{n,n}(\cdot) - L_n(\cdot)\right|\right| \\ &\quad = \left|\left|A_0\left(\cdot;\hat{F}_n\right) - A_0(\cdot;F) + n^{-\frac{1}{2}}\left(A_1\left(\cdot;\hat{F}_n\right) - A_1(\cdot;F)\right) + o_p\left(n^{-\frac{1}{2}}\right)\right|\right| \\ &\quad = \Omega_p(n^{-\frac{1}{2}}). \end{aligned}$$

On the other hand,

$$\left|\left|L^*_{m_j,n}(\cdot) - L^*_{m_{j+1},n}(\cdot)\right|\right| = m_j^{-\frac{1}{2}}\left|\left|A_1\left(\cdot;\hat{F}_n\right)\left(1 - q^{-\frac{1}{2}}\right) + o_p(1)\right|\right|,$$

which is minimized by $\hat{m} = n(1 + o_p(1))$. Thus (8) follows.

(b) From the assumptions, $A_0(\cdot;F) = A_0(\cdot)$, and

$$\left|\left|L^*_{n,n}(\cdot) - L_n(\cdot)\right|\right| = n^{-\frac{1}{2}}\left|\left|A_1\left(\cdot;\hat{F}_n\right) - A_1(\cdot;F) + o_p(1)\right|\right| = O_p\left(n^{-1}\right).$$

As above,

$$\left|\left|L^*_{m_j,n}(\cdot) - L^*_{m_{j+1},n}(\cdot)\right|\right| = m_j^{-\frac{1}{2}}\left|\left|A_1\left(\cdot;\hat{F}_n\right)\left(1 - q^{-\frac{1}{2}}\right) + o_p(1)\right|\right|,$$

which is minimized, to first order, by $\hat{m} = n(1 + o_p(1))$, and to second order, by $\hat{m} = n + o(1)$. Then

$$\begin{aligned} &\left|\left|L^*_{\hat{m},n}(\cdot) - L_n(\cdot)\right|\right| \\ &\quad = \left|\left|\hat{m}^{-\frac{1}{2}}\left(A_1\left(\cdot;\hat{F}_n\right) - A_1(\cdot;F)\right) + \left(\hat{m}^{-\frac{1}{2}} - n^{-\frac{1}{2}}\right)A_1(\cdot;F) + o_p\left(\hat{m}^{-\frac{1}{2}}\right)\right|\right| \\ &\quad = O_p\left(n^{-1}\right) \end{aligned}$$

and (9) follows.

This completes the proof.

2.3. Proof of Theorem 3

Writing

$$\begin{aligned}L_n(\cdot) &- L^*_{m,n}(\cdot)\\ &= A_0(\cdot;F) - A_0(\cdot;\hat{F}_n) + A_1(\cdot;F)(n^{-\frac{1}{2}} - m^{-\frac{1}{2}}) + o_p(m^{-\frac{1}{2}}) + \Omega_p(m^\beta n^{-\gamma})\\ &= A_1(\cdot;F)m^{-\frac{1}{2}} + \Omega_p(m^\beta n^{-\gamma}) + o_p(m^{-\frac{1}{2}}) + O_p(n^{-\gamma}),\end{aligned}$$

yielding an (optimal) minimizing value of $m_{opt} = \Omega_p(n^{\gamma(1/2+\beta)^{-1}})$, and (11) follows. Similarly,

$$L^*_{m_{j+1},n}(\cdot) - L^*_{m_j,n}(\cdot) = A_1(\cdot;F)m_j^{-\frac{1}{2}}\left(q^{-\frac{1}{2}} - 1\right) + \Omega_p\left(m_j^\beta n^{-\gamma}\right) + o_p\left(m^{-\frac{1}{2}}\right),$$

yielding $\hat{m} = \Omega_p(n^{\gamma(1/2+\beta)^{-1}})$, and (12) follows. The last claim follows from the above.

This completes the proof.

3. Appendix to Section 4

3.1. Lemma 1

Lemma 1. *Under* (vM)(I)–(III), *if* $m \to \infty$, $m/n \to 0$, *then*

$$a_m(X_{([n-\frac{n}{m}],n)} - b_m) = o_p(1).$$

Proof. From Theorem 5.1.7 (p. 164) of Reiss (1989), under (vM)(I)–(III) (and identifying $k(n)$ with $n - [n - n/m]$, and Reiss's b_n is our b_m) it follows that, $(nm)^{1/2} f(b_m)(X_{([n-n/m],n)} - b_m)$ converges weakly to the standard normal distribution.

In case (vM)(III), $a_m = mf(b_m)$ and the lemma follows.

In case (vM)(I), $a_m = 1/b_m$, and from (5.1.24) of Reiss (1989), $mb_m f(b_m) = O(1)$. Hence,

$$\frac{a_m}{\sqrt{nm}f(b_m)} = \frac{1}{b_m\sqrt{m}f(b_n)\sqrt{n}} = o(1),$$

and the lemma follows.

In case (vM)(II), and using (5.1.24) of Reiss (1989), the result follows using a similar argument.

3.2. Condition (A.5)

The proof of this condition uses the Poisson approximation to the Binomial distribution. Some notation and definitions are needed.

Define, $\hat{p}_n(x,\lambda) = 1 - \hat{F}_n(x/a_m + b_m)$, and $V_n(\lambda, x) = N(m\hat{p}_n(x,\lambda))$, where N denotes a standard Poisson process, independent of $X_1, X_2, \ldots$. Hence, $P^*(V_n(\lambda, x) = 0) = \exp(-m\hat{p}_n(x,\lambda))$. Define, further,

$$V_n^*(\lambda, x) = \begin{cases} \sum_{i=1}^m 1(X_i^* > \frac{x}{a_m} + b_m),\ m = [n\lambda] + 1,\ 0 \le \lambda < 1 - \frac{1}{n} \\ \sum_{i=1}^n 1(X_i^* > \frac{x}{a_n} + b_n),\quad 1 - \frac{1}{n} \le \lambda \le 1. \end{cases}$$

Lemma 2. *If* (13) *holds then for all* $\lambda_1, \ldots, \lambda_k$, $k < \infty$

$$\rho(\mathcal{L}^*(V_n^*(\lambda_1), \ldots, V_n^*(\lambda_k)), \mathcal{L}^*(V_n(\lambda_1), \ldots, V_n(\lambda_k))) \xrightarrow{p} 0,$$

where ρ *is the Prohorov metric on probabilities on* $D(\bar{R})$, *and* $\mathcal{L}^*$ *is the joint conditional law.*

Proof. Denote by $\|\cdot\|_{BV}$ the total variation norm between two probability measures. In the proof of Lemma 2, we use the following proposition which is an extension of a result of Hodges and LeCam (1960) (for a proof see Barbour (1988), Wang (1986), SintesBlanc (1991)).

Proposition 1. *Let* $\mathbf{X}_i = (X_{i1}, \ldots, X_{ik})$, $i = 1, \ldots, n$, *be independent and distributed according to the multinomial distribution with parameters* $(1, q_1, \ldots, q_k)$, *such that* $q_j \ge 0$ *and* $\sum_{j=1}^k q_j < 1$. *Let* $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ik})$, $i = 1, \ldots, n$, *where* Y_{ij} *are independent Poisson random variables with means* q_j. *Then* $\|\mathcal{L}(\mathbf{X}_1, \ldots, \mathbf{X}_n) - \mathcal{L}(\mathbf{Y}_1, \ldots, \mathbf{Y}_n)\|_{BV} \le n(\sum_{i=1}^k q_i)^2$, *where* $\mathcal{L}(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$ *is the joint law of* $(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$.

Continue proof of Lemma 2. For all positive integers K, L define $A_{K,L} = \{(k,l) \mid 1 \le k \le K,\ 1 \le l \le L\}$. Since for each λ, $V_n^*(\lambda, \cdot)$, $V_n(\lambda, \cdot)$ are monotone decreasing in x, it is suffices to establish for all $\lambda_1 < \cdots < \lambda_I$, $x_J < \cdots < x_1$, I, J:

$$R \equiv \rho(\mathcal{L}^*(\{V_n^*(\lambda_i, x_j) : (i,j) \in A_{I,J}\}),\ \mathcal{L}^*(\{V_n(\lambda_i, x_j) : (i,j) \in A_{I,J}\})) \xrightarrow{p} 0.$$

Let m_i be the m associated with λ_i, and set

$$C_{ij} = \begin{cases} (\frac{x_1}{a_{m_i}} + b_{m_i}, \infty), & j = 1 \\ (\frac{x_j}{a_{m_i}} + b_{m_i}, \frac{x_{j-1}}{a_{m_i}} + b_{m_i}], & j = 2, \ldots, J. \end{cases}$$

For a given i, the intervals C_{ij} are disjoint, but they are not necessarily so for different i's. We use the end-points of the C_{ij}, when $1 \le i \le I$ and $1 \le j \le J$, to create a partition of the real line into disjoint intervals D_r, where $1 \le r \le \beta$, and $1 \le \beta \le IJ$. Denote the end-points of the D's by $D_r = (z_{r-1}, z_r]$, where $-\infty < z_1 \cdots < z_\beta < z_{\beta+1} = \infty$. Each z_r is associated with a pair (x_j, λ_i), for some i and j.

Finally, set $\hat{p}_n(z_r)$ to be $\hat{p}_n(x_j, \lambda_i)$, with the corresponding i and j. With this preparation, rewrite,

$$V_n^*(\lambda_i, x_j) = \sum_{k=1}^{m_i} 1\left(X_k^* > \frac{x_j}{a_{m_i}} + b_{m_i}\right) = \sum_{k=1}^{m_i}\sum_{l=1}^{j} 1(X_k^* \in C_{il})$$
$$= \sum_{k=1}^{n}\sum_{r=1}^{\beta} \delta_{ijk} 1(X_k^* \in D_r),$$

where

$$\delta_{ijk} = \begin{cases} 1, \text{ if } 1 \le k \le m_i \text{ and } D_r \subset \bigcup_{l=1}^{j} C_{il} \\ 0, \text{ otherwise.} \end{cases}$$

For $r = 1, \ldots, \beta$, $P^*\left(X_k^* \in D_r\right) = \hat{p}_n(z_{r-1}) - \hat{p}_n(z_r)$.

We can similarly write, $V_n(\lambda_i, x_j) = \sum_{k=1}^{n}\sum_{r=1}^{\beta} \delta_{ijk} P_{kr}$, where the δ_{ijk} are as above, and the P_{kr} are independent Poisson with $E(P_{kr}) = P^*(X_k^* \in D_r)$. Therefore,

$$||\mathcal{L}^*\left(\{V_n^*(\lambda_i, x_j) : (i,j) \in A_{I,J}\}\right) - \mathcal{L}^*\left(\{V_n(\lambda_i, x_j) : (i,j) \in A_{I,J}\}\right)||_{BV}$$
$$\le ||\mathcal{L}^*(\{(1(X_k^* \in D_r) : (k,r) \in A_{n,\beta}\}) - \mathcal{L}^*(\{(P_{kr} : (k,r) \in A_{n,\beta}\})||_{BV}$$
$$\le n\Big(\sum_{r=1}^{\beta} P^*(X_1^* \in D_r)\Big)^2 \le nIJ \max_{i,j}\{\hat{p}_n^2(x_j, \lambda_i)\}.$$

To obtain the last statement, we have used the bound from Proposition 1.

Since $n\hat{p}_n$ has the Binomial distribution, $\mathrm{Var}(n\hat{p}_n(x_j, \lambda_i)) \le nE(\hat{p}_n(x_j, \lambda_i)) = n(1 - F(x_j/a_{m_i} + b_{m_i}))$. Note that assumption (13) is equivalent to

$$n\Big(1 - F(\frac{x}{a_n} + b_n)\Big) \to -\log G(x). \tag{26}$$

Combining the above, and since $n/m_i = O(1)$, we conclude that $n \max_{i,j} \hat{p}_n(x_j, \lambda_i) = O_p(1)$. Thus,

$$R \le \|\mathcal{L}^*(\{V_n^*(\lambda_i, x_j) : (i,j) \in A_{I,J}\}) - \mathcal{L}^*(\{V_n(\lambda_i, x_j) : (i,j) \in A_{I,J}\})\|_{BV} = o_p(1),$$

and the lemma follows.

Set, $W_n(\lambda) = a_m(X_{([n-n/m],n)} - b_m)$, then from (2):

$$U_n(\lambda, x) = P^*\left(V_n^*(\lambda, x + W_n(\lambda)) = 0\right).$$

Define, $S_n(x, \lambda) = n\hat{p}_n(x, \lambda)$, and $\tilde{S}_n(x, \lambda) = S_n(x + W_n(\lambda), \lambda)$. Using the last lemma, and the definitions of the processes, it follows that,

$$\rho\left(\left(U_n(\cdot, \lambda_1), \ldots, U_n(\cdot, \lambda_k)\right), (e^{-\lambda_1 \tilde{S}_n(\cdot, \lambda_1)}, \ldots, e^{-\lambda_k \tilde{S}_n(\cdot, \lambda_k)})\right) = o_p(1). \tag{27}$$

Let N' be a standard Poisson process, independent of N, and of $X_1, X_2, \ldots$, and let $\tau_1 < \tau_2 < \cdots$ be the consecutive jumps of $N'(\cdot)$. That is, $N'(\tau_r) = r$ and $N'(\tau_r -) = r - 1$. Denote $\psi(x) = -\log G(x)$.

Lemma 3. *Set* $\underline{S_n}(\cdot) = (S_n(\cdot, \lambda_1), \cdots, S_n(\cdot, \lambda_k))$, *and similarly for* $\underline{\tilde{S}_n}(\cdot)$ *and* $\underline{W_n}$. *Then,*

(i) *The process* $(\underline{S_n}(\cdot), \underline{W_n})$ *converges weakly to a limit*

$$(\underline{S}(\cdot), \underline{W}) \equiv (S(\cdot, \lambda_1), \ldots, S(\cdot, \lambda_k), W(\lambda_1), \ldots, W(\lambda_k)).$$

Hence, $\underline{\tilde{S}_n}(\cdot)$ *converges weakly to* $\underline{\tilde{S}}(\cdot) \equiv (\tilde{S}(\cdot, \lambda_1), \ldots, \tilde{S}(\cdot, \lambda_k))$, *where* $\tilde{S}(x, \lambda) \equiv S(x + W(\lambda), \lambda)$ *for* $1 \le j \le k$.

(ii) *Moreover, we can identify the distribution of* $S(\cdot, \lambda_j)$: *for* (vM)(III), *set* $\sigma_j = 1$, *and* $\mu_j = \log(\lambda_j)$. *For* (vM)(I)-(II), *set* $\sigma_j = \lambda_j^{-2}$ *and* $\mu_j = 1 - \lambda_j^{-2}$. *For* $\underline{u} = (u_1, \ldots, u_k)$, *let* $N'(\cdot|\underline{u})$ *denote the conditional distribution of* $N'(\cdot)$, *given* $\tau_{r_j} = \psi(\sigma_j u_j + \mu_j)$ *for* $1 \le j \le k$. *Then,* $\{\tilde{S}(\cdot, \lambda_j) : 1 \le j \le k\}$ *is distributed as* $\{N'(\psi(\cdot + \sigma_j u_j + \mu_j)) : 1 \le j \le k\}$, *given* $W(\lambda_j) = u_j$ *for* $1 \le j \le k$. σ_j *and* μ_j *are defined as follows.*

Assumption (A.5) follows from combining (27) and the first part of Lemma 3. The second part of the lemma is needed for Lemma 5).

Agenda for the proof of Lemma 3. To prove Lemma 3, we shall argue that:

(a) $\underline{W_n}$ has a weak limit $\underline{W}$.
(b) The conditional distribution of $\underline{S_n}(\cdot)$, given $\underline{W_n} = \underline{u}$, convergences weakly to a measure on $D^k[-\infty, \infty]$, say, $Q_{\underline{u}}$. Then, $(\underline{S_n}(\cdot), \underline{W_n})$, necessarily, converges weakly to $(\underline{S}(\cdot), \underline{W})$, where $\underline{S}(\cdot)$ has conditional distribution $Q_{\underline{W}}$.
(c) With these identifications, it follows that $\underline{\tilde{S}_n}(\cdot)$, given $W_n(\lambda_j) = u_j$, for $1 \le j \le k$, converges weakly to $(S(\cdot + u_1, \lambda_1), \ldots, S(\cdot + u_k, \lambda_k))$. This completes the first part of the lemma.
(d) The second part of the lemma is being proved while proving the first part.

We begin with an auxiliary lemma.

Lemma 4. *Under the von Mises conditions, if* $m/n \to \lambda > 0$ *then:*
1. *For* (vM)(I) *and* (vM)(II), $a_m/a_n \to \lambda^2$ *and* $a_n(b_m - b_n) \to \lambda^{-2} - 1$.
2. *For* (vM)(III), $a_m/a_n \to 1$ *and* $a_n(b_m - b_n) \to -\log(\lambda)$.

Proof. To simplify the notation, we denote m/n by λ rather then $\lambda^{(n)} \to \lambda$. The result is unaffected.

We start with (vM)(III). Here, $a_m = mf(b_m)$ and $a_n = nf(b_n)$. Noting that $1 - F(b_m) = 1/\lambda n$, and using (15):

$$\frac{\partial \log(f(b_m))}{\partial \lambda} = \frac{\partial \log f}{\partial \lambda}\left(F^{-1}\left(1 - \frac{1}{\lambda n}\right)\right) = \frac{f'(b_m)}{\lambda m f^2(b_m)}$$

$$= -\frac{1}{\lambda}\left[\left(\frac{1-F(b_m)}{f(b_m)}\right)' + 1\right] \to -\frac{1}{\lambda}.$$

Hence,

$$\log\left(\frac{a_n}{a_m}\right) = \log\left(\frac{f(b_n)}{\lambda f(b_m)}\right) = -\int_\lambda^1 \frac{dz}{z} + o(1) - \log(\lambda)$$
$$= -\log(1) + \log(\lambda) - \log(\lambda) + o(1) \to 0,$$

It follows that, $a_n/a_m \to 1$. Similarly,

$$\frac{\partial}{\partial\lambda} a_n(b_m - b_n) = \left[nf(b_n)\left(F^{-1}\left(1-\frac{1}{\lambda n}\right) - F^{-1}\left(1-\frac{1}{n}\right)\right)\right]'$$
$$= \frac{1}{\lambda^2}\frac{f(b_n)}{f(b_m)} = \frac{1}{\lambda}\frac{a_n}{a_m} \to \frac{1}{\lambda}.$$

Consider now (vM)(I). Then, $a_n = 1/b_n$ and $a_m = 1/b_m$. From (15):

$$\frac{\partial \log(f(b_m))}{\partial\lambda} \to -\frac{1}{\lambda}\left(\frac{1}{\lambda}+1\right). \tag{28}$$

Using (5.1.24) of Reiss (1989), which is an alternative set of sufficient conditions, it follows that

$$\frac{b_m f(b_m)}{1-F(b_m)} \cdot \frac{1-F(b_n)}{b_n f(b_n)} = \frac{b_m}{b_n} \cdot \frac{f(b_m)}{f(b_n)} \cdot \frac{m}{n} \to 1.$$

Using (28), it follows that, $a_n/a_m = b_m/b_n \to \lambda^{-2}$. Hence, $a_m/a_n \to \lambda^2$ and $a_n(b_m - b_n) = b_m/b_n - 1 \to \lambda^{-2} - 1$.

The argument for type (vM)(II) is very similar.

Proof of Lemma 3. From (5.1.28) in Reiss (1989), it follows that for $m = \lambda^{(n)}n$, and $\lambda^{(n)} \to \lambda$, $\tilde{W}_n(\lambda^{(n)}) = a_n(X_{([n-n/m],n)} - b_n)$, converges weakly to a limiting distribution, which depends on λ and G, and is given in (5.1.29) of Reiss (1989). To simplify notation, from now on, we drop the superscript in $\lambda^{(n)}$.

Note that, $W_n(\lambda) = (a_m/a_n)\tilde{W}_n(\lambda) + a_m(b_n - b_m)$. In view of Lemma 4, $W_n(\lambda)$ converges weakly, say to $W(\lambda)$. To show that $\underline{\tilde{W}_n}$, and hence $\underline{W_n}$, converges weakly to some $\underline{W}$, is a slight extension of (5.1.28) in Reiss (1989): to show this we only need to translate joint statements about extrema to joint statements about empirical distribution functions. This completes the proof of part (a) from the agenda.

To proceed with part (b) of the agenda we define, $N_n(u) \equiv S_n(\psi^{-1}(u), 1)$, a nondecreasing counting process, with jump points, $\tau_{1,n} < \cdots < \tau_{n,n}$, where,

$\tau_{j,n} = \psi(a_n(X_{(n-j+1,n)} - b_n))$. Define,

$$r_j = \begin{cases} [\frac{1}{\lambda_j}] + 1, & \text{if } \frac{1}{\lambda j} \text{ is not an integer} \\ \frac{1}{\lambda_j}, & \text{otherwise} \end{cases} \quad \text{for} \quad j = 1, \ldots, k.$$

Since, $m_1 > \cdots > m_k$, and hence $\tau_{r_k,n} \geq \cdots \geq \tau_{r_1,n}$, it follows that, $\{r_j\}$ is a non-decreasing sequence of positive integers ($1 \leq r_j \leq n$). Using the jump points rewrite:

$$\tau_{r_j,n} = \psi\left(a_n/a_{m_j} W_n(\lambda_j) - a_n(b_n - b_{m_j})\right). \tag{29}$$

This implies that, conditioning $\{S_n(\cdot, \lambda_j) : 1 \leq j \leq k\}$, on $\{W_n(\lambda_j) : 1 \leq j \leq k\}$, is equivalent to conditioning on $\{\tau_{r_j,n} : 1 \leq j \leq k\}$.

Note that, $N_n(\cdot)$ converges weakly to a standard Poisson process N'. Since

$$S_n(x, \lambda) = N_n(\psi(x\frac{a_n}{a_m} + (b_m - b_n)a_n)), \tag{30}$$

it follows that

$$S(x, \lambda_j) = N'(\psi(\sigma_j x - \mu_j)), \tag{31}$$

where $\sigma_j = a_n/a_m$ and $\mu_j = a_n(b_n - b_m)$. Using Lemma 4, and according to the situation (i.e., (vM)(I),(II) or (III)), we obtain the format as given in the lemma. The jump points of $S(x, \lambda_j)$ are

$$\tau_{r_j} = \psi(\sigma_j W(\lambda_j) + \mu_j). \tag{32}$$

From (29)$-$(32), it follows that, to complete the argument we need only to check that the weak limit of the conditional distribution of $N_n(\cdot)$, given $W_n(\lambda_j) = u_j$ for $j = 1, \ldots, k$, is that of $N'(\cdot|\underline{u})$, or, equivalently, that the conditional distribution of $N_n(\cdot)$, given $\tau_{r_j,n} = \psi(u_j a_n/a_{m_j} - a_n(b_n - b_{m_j}))$, converges to that of $N'(\cdot|\underline{u})$.

We start with the latter. Given $\tau_{r_j} = \psi(\sigma_j u_j + \mu_j)$ for $1 \leq j \leq k$, the remaining $n - k$ jump points of $N'(x)$, $0 \leq x \leq \tau_{r_k}$ are distributed as the concatenation of the following k blocks of order statistics of independent samples: the first block correspond to the first $r_1 - 1$ jump points, which are distributed like the order statistics of a sample of size $r_1 - 1$ from the $U(0, \tau_{r_1})$ distribution; the second block correspond to the $r_2 - r_1 - 1$ jump points between τ_{r_1} and τ_{r_2}, and are distributed like the order statistics of a sample of size $r_2 - r_1 - 1$ from $U(\tau_{r_1}, \tau_{r_2})$, and so on until the kth block, which correspond to the $r_k - r_{k-1} - 1$ jump points between $\tau_{r_{k-1}}$ and τ_{r_k}, and are distributed like the order statistics of a sample of size $r_k - r_{k-1} - 1$ from $U(\tau_{r_{k-1}}, \tau_{r_k})$. Finally, for $x > \tau_{r_k}$, $N'(x) - N'(x - \tau_{r_k})$ is again a standard Poisson process.

We now consider the conditional distribution of $N_n(\cdot)$, given $\tau_{r_j,n}$. From an obvious extension of Theorem 2.7 of David (1981), it follows that, given order statistics, $Y_{i_1},\dots,Y_{i_k}$ of a sample $Y_1,\dots,Y_n$ iid from a density $g\equiv G'$, the remaining $n-k$ order statistics are distributed like the $k+1$ blocks of order statistics of independent samples: $Y_1^{(1)},\dots,Y_{i_1-1}^{(1)}$ from density g_1; $Y_1^{(2)},\dots,Y_{i_2-i_1-1}^{(2)}$ from density g_2; and so on until $Y_1^{(k+1)},\dots,Y_{n-i_k}^{(k+1)}$ from density g_{k+1}, where:

$$
\begin{aligned}
&g_1(y)=g(y)1(y<Y_{(i_1)})/G(Y_{(i_1)});\\
&g_j(y)=g(y)1(Y_{(i_{j-1})}<y<Y_{(i_j)})/(G(Y_{i_j})-G(Y_{i_{j-1}})), \qquad j=2,\dots,k;\\
&g_{k+1}(y)=g(y)1(y>Y_{(i_k)})/\left(1-G\left(Y_{(i_k)}\right)\right).
\end{aligned}
$$

For $1\le j\le k$, let $Y_{in}^{(j)}$ be iid with conditional distribution functions (conditioned on $\tau_{r_j,n}$):

$$
\begin{aligned}
G_n^{(1)}(x)&=\frac{\bar F(X_{(n-r_1+1,n)})-\bar F(\frac{x}{a_n}+b_n)}{\bar F(X_{(n-r_1+1,n)})}, \quad x>\psi^{-1}(\tau_{r_1,n}),\\
G_n^{(j)}(x)&=\frac{\bar F(X_{(n-r_{j-1}+1,n)})-\bar F(\frac{x}{a_n}+b_n)}{\bar F(X_{(n-r_{j-1}+1,n)})-\bar F(X_{(n-r_j+1,n)})}, \ \psi^{-1}(\tau_{r_j,n})<x<\psi^{-1}(\tau_{r_{j-1},n}),\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad j=2,\dots,k,\\
G_n^{(k+1)}(x)&=\frac{F(\frac{x}{a_n}+b_n)}{F(X_{(n-r_k+1,n)})}, \quad x<\psi^{-1}(\tau_{r_k,n}).
\end{aligned}
$$

From (26) it follows that,

$$
\begin{aligned}
G_n^{(1)}(x)&\sim 1-\frac{\psi(x)}{\tau_{r_1,n}}\\
G_n^{(j)}(x)&\sim \frac{\tau_{r_{j-1},n}-\psi(x)}{\tau_{r_{j-1},n}-\tau_{r_j,n}} \qquad \text{for} \quad 1\le j\le k\\
G_n^{(k+1)}&\sim F\left(\frac{x}{a_n}+b_n\right).
\end{aligned}
$$

Set,

$$
\begin{aligned}
S_n^{(1)}(x)&=\sum_{i=1}^{r_1}1\left(Y_{in}^{(1)}>x\right),\\
S_n^{(j)}(x)&=\sum_{i=1}^{r_j-r_{j-1}-1}1\left(Y_{in}^{(j)}>x\right), \qquad 2\le j\le k,\\
S_n^{(k+1)}(x)&=\sum_{i=1}^{n-r_k-1}1\left(Y_{in}^{(k+1)}>x\right).
\end{aligned}
$$

Combining the above remarks with this set-up, we see that given $\tau_{r_j,n}$ for $1 \le j \le k$, $S_n(x) \equiv S_n(x,1)$ is distributed as the concatenation of $k+1$ independent processes, $S_n^{(j)}(\cdot)$,

$$S_n(x) = \begin{cases} S_n^{(1)}(x), & x > \psi^{-1}(\tau_{r_1,n}) \\ S_n^{(j)}(x), & \psi^{-1}(\tau_{r_j,n}) < x < \psi(\tau_{r_{j-1},n}), \quad 2 \le j \le k \\ S_n^{(k+1)}(x), & x < \psi^{-1}(\tau_{r_k,n}). \end{cases}$$

It follows that, if $\tau_{r_j} = \psi(\sigma_j u_j - \mu_j)$, then given $\tau_{r_j,n} = \psi(u_j a_n/a_{m_j} - a_n(b_n - b_{m_j}))$ for $1 \le j \le k$, $G_{j,n}$ converges weakly to G_j, where

$$G_j(x) = \frac{1 - \frac{\psi(x)}{\tau_{r_j}}}{1 - \frac{\psi(\tau_{r_j})}{\psi(\tau_{r_{j-1}})}}.$$

Therefore, $N_n^{(j)}(x)$, for $0 < x < \psi(\sigma_k u_k + \mu_k)$, has the limiting distribution of $N'(\cdot, \underline{u})$, since $G_j\left(\psi^{-1}(\cdot)\right)$ is the Uniform distribution on $\tau_{r_{j-1}} < x < \tau_{r_j}$.

Finally, since $n\bar{F}(a_n\psi^{-1}(x) + b_n) \to \psi\psi^{-1}(x) = x$, we can show that for $x > \tau_{r_k,n}$, $N_n^{(k+1)}(\cdot) - N_n(\tau_{r_k,n}) = \sum_{j=1}^{n} 1(X_j > \psi^{-1}(x)/a_n + b_n) - r_k$, converges weakly to a standard Poisson process, by simply checking finite dimensional joint distributions. The lemma follows when using the above and the usual Poisson convergence theorem.

3.3. Condition (A.6)

Lemma 5.
(i) *Define*

$$r \equiv r(\lambda) = \begin{cases} [\frac{1}{\lambda}] + 1, & \text{if } \frac{1}{\lambda} \text{ is not an integer} \\ \frac{1}{\lambda}, & \text{otherwise.} \end{cases}$$

then the marginal distribution of $\tilde{S}(x,\lambda) + r(\lambda)$ *is Poisson with parameter* $\sigma(\lambda)x$.

(ii) *Suppose* $U(\lambda,\cdot)$ *is the limit law of* $U_n(\lambda)$*, and* $U(\lambda,\cdot)$ *has the distribution specified in Lemma* 3. *That is,* $U(\lambda, x)$ *has the distribution of* $\exp(-\lambda\tilde{S}(x,\lambda))$. *Reparametrize* $U(\lambda,\cdot)$ *by* r *given above, say* $U(\lambda,\cdot) = \tilde{U}(r,\cdot)$. *Then,* $r \to \tilde{U}(r,\cdot)$ *is* $1-1$, $r = 1, 2, \ldots$.

Proof. For a fixed $x, \lambda = 1/m$, $\tilde{S}(x,\lambda)$, given $W(\lambda) = u$, is distributed as $N'(\psi(\sigma(\lambda)(x+u) + \mu(\lambda)))$ given $\tau_r = \psi(\sigma(\lambda)u + \mu(\lambda)$. But, $N'(\psi(\sigma(\lambda)(x+u) + \mu(\lambda))) - N'(\psi(\sigma(\lambda)u + \mu(\lambda)))$ is independent of $N'(\psi(\sigma(\lambda)u + \mu(\lambda))) = r$ given $\tau_r = \psi(\sigma(\lambda)u + \mu(\lambda))$, and has a Poisson $\psi(\sigma(\lambda)x)$ distribution and claim (i) follows.

Therefore, $\mathrm{E}(\exp(-\lambda\tilde{S}(x,\lambda))) = \exp(-\lambda r)\exp(\psi(\sigma(\lambda)x)\cdot(e^{-\lambda}-1))$.

For (vM)(III), $\psi(u) = e^{-u}$ and $\sigma(\lambda) = 1$. Therefore, if $r \to \tilde{U}(r)$ is not 1-1, then the function $\lambda \to \exp\{-r(\lambda)\lambda\}\exp\{e^{-x}(e^{-\lambda}-1)\}$ is not 1-1, as a map from $(0,1)$ to functions of x. But this is evidently false.

The argument for (vM)(I) and (vM)(II) is the same for the function $\lambda \to exp(-\lambda r(\lambda))\exp(\exp((x/\lambda^2)^\beta)(\exp(-\lambda)-1))$, for $\beta = \pm\gamma$.

Comment. This weakening of (A.6) suffices, since $\hat{m}$ is obtained by a search over $m_j = [q^j n]$, where $[n/m_j]$ ranges over distinct integers.

Department of Statistics, University of California, Berkeley, CA, 94720-3860, U.S.A.

E-mail: bickel@stat.berkeley.edu

Tanor Group, 4 Simtat Hashezif, Zichron Yaacov 30900, Israel.

E-mail: anat@sakov.com

(Received September 2006; accepted April 2007)