The Annals of Probability
1996, Vol. 24, No. 1, 541–545

CORRECTION

ENTROPY AND THE CONSISTENT ESTIMATION OF JOINT DISTRIBUTIONS

BY KATALIN MARTON[1] AND PAUL C. SHIELDS[2]

The Annals of Probability (1994) **22** 960–977

Jeff Steif has brought to our attention an error on page 975 of our paper. Our argument that inequality (30) implies the immediately following inequality has a gap. A closer look shows an even more serious problem, namely, Lemma 6 as stated is probably not true, since nothing in the weak Bernoulli property precludes the possibility that splitting sets for x_1^n may depend on past coordinates $\{x_i: i \le 0\}$. With a modified definition of the splitting concept an alternative version of Lemma 6 is true and this is sufficient to prove our principal theorem, Theorem 4.

The following text replaces the discussion from the paragraph preceding Lemma 5 on page 973 to the end of Section 3 on page 976.

The ψ-mixing admissibility result is extended to the weak Bernoulli case as follows. The basic idea remains the same: replace the overlapping k-block distribution by a shifted nonoverlapping k-block distribution with a gap g between the blocks. Then replace the measure by the product measure on these k-blocks, a replacement that introduces only a small exponential error. Then apply the i.i.d. result. The weak Bernoulli property guarantees that only a small exponential error is introduced by replacing the measure by the product measure, at least for a large fraction of shifts, provided a small fraction of blocks are omitted and conditioning on the past is allowed. This will be enough to obtain the weak Bernoulli admissibility result.

Given positive integers k and g, $r \in [1, k+g]$ and $j \ge 1$, define

$$\tilde{x}_j(r) = x_{r+(j-1)(k+g)}^{r+(j-1)(k+g)+k-1}.$$

For $(t+1)(k+g) \le n < (t+2)(k+g)$ and $J \subset [1,t]$, define

$$\hat{\mu}_{k,g}^{r,J}(a_1^k | x_1^n) = \frac{|\{j \in J: \tilde{x}_j(r) = a_1^k\}|}{|J|}, \qquad a_1^k \in A^k,$$

that is, the empirical distribution of k-blocks obtained by looking only at those k-blocks $\tilde{x}_j(r)$ for which $j \in J$.

We will make use of the fact that if the overlapping k-block distribution is not close to the true distribution, then for a fixed fraction of shifts, $\hat{\mu}_{k,g}^{r,J}$ is not close to the true distribution, as long as J is a large subset of $[1,t]$. This

Received June 1995.
[1]Partially supported by Hungarian National Foundation for Scientific Research Grant OTKA 1906 and MTA-NSF Project 37.
[2]Partially supported by NSF Grant DMS-90-24240 and MTA-NSF Project 37.

sharper form of Lemma 4 is easy to prove. We state it as follows in the form that will be used.

LEMMA 5. *Given $\delta > 0$, there is a positive $\gamma < 1/2$ such that for any g there is a $K = K(g, \gamma)$ such that if $k \geq K$, if $k/n < \gamma$ and if $|\hat{\mu}_k(\cdot|x_1^n) - \mu_k| \geq \delta$, then $|\hat{\mu}_{k,g}^{r,J}(\cdot|x_1^n) - \mu_k| \geq \delta/4$ for at least $2\gamma(k+g)$ indices $r \in [1, k+g]$ for any subset $J \subset [1, t]$ of cardinality at least $(1-\gamma)t$.*

Given $\gamma > 0$, an index $j \geq 1$ will be called a (γ, r, k, g)-*splitting index* for the (doubly infinite) sequence $x \in A^Z$ if

$$\mu(\tilde{x}_j(r)|x_{-\infty}^{r+(j-1)(k+g)-g-1}) < (1+\gamma)\mu(\tilde{x}_j(r)).$$

The set of all x for which j is a (γ, r, k, g)-splitting index will be denoted by $B_j(\gamma, r, k, g)$ or by $B_{r,j}$ if γ, k and g are understood. Note that the set $B_{r,j}$ is measurable with respect to the past coordinates $i \leq r + (j-1)(k+g) + k - 1$.

LEMMA 6A. *Fix (γ, r, k, g) and fix a finite set J of positive integers. Then for any assignment $\{\tilde{x}_j(r): j \in J\}$ of k-blocks,*

$$\mu\Big(\bigcap_{j \in J}([\tilde{x}_j(r)] \cap B_{r,j})\Big) \leq (1+\gamma)^{|J|} \prod_{j \in J} \mu([\tilde{x}_j(r)]).$$

PROOF. Put $j_m = \max\{j: j \in J\}$ and condition on

$$B^* = \bigcap_{j \in J - \{j_m\}} ([\tilde{x}_j(r)] \cap B_{r,j})$$

to obtain

$$\begin{aligned} &\mu\Big(\bigcap_{j \in J}([\tilde{x}_j(r)] \cap B_{r,j})\Big) \\ &\quad = \mu\Big(\bigcap_{j \in J - \{j_m\}}([\tilde{x}_j(r)] \cap B_{r,j})\Big)\mu([\tilde{x}_{j_m}(r)] \cap B_{r,j_m}|B^*). \end{aligned} \tag{101}$$

The second factor $\mu([\tilde{x}_{j_m}(r)] \cap B_{r,j_m}|B^*)$ is an average of the measures

$$\mu([\tilde{x}_{j_m}(r)] \cap B_{r,j_m}|x_{-\infty}^{r+(j_m-1)(k+g)-g-1}),$$

each of which satisfies

$$\mu([\tilde{x}_{j_m}(r)] \cap B_{r,j_m}|x_{-\infty}^{r+(j_m-1)(k+g)-g-1}) \leq (1+\gamma)\mu(\tilde{x}_{j_m}(r)),$$

by the definition of B_{r,j_m}. Thus (101) yields

$$\mu\Big(\bigcap_{j \in J}([\tilde{x}_j(r)] \cap B_{r,j})\Big) \leq (1+\gamma) \cdot \mu(\tilde{x}_{j_m}(r)) \cdot \mu\Big(\bigcap_{j \in J - \{j_m\}}([\tilde{x}_j(r)] \cap B_{r,j})\Big),$$

and the proof follows by induction. □

The almost sure existence of a large density of splitting indices for most shifts r is established in the following lemma.

LEMMA 6B. *If μ is weak Bernoulli and $0 < \gamma < 1/2$, then there is a gap $g = g(\gamma)$, there are integers $k(\gamma)$ and $t(\gamma)$ and there is a sequence of measurable sets $\{G_n(\gamma)\}$, such that the following hold:*

(a) *$x \in G_n(\gamma)$ eventually a.s.*

(b) *If $k \geq k(\gamma)$, if $t \geq t(\gamma)$ and if $(t+1)(k+g) \leq n < (t+2)(k+g)$, then for $x \in G_n(\gamma)$, there are at least $(1-\gamma)(k+g)$ values of $r \in [1, k+g]$ for each of which there are at least $(1-\gamma)t$ indices j in the interval $[1, t]$ that are (γ, r, k, g)-splitting indices for x.*

PROOF. First we use the weak Bernoulli property to choose $g = g(\gamma)$ so large that for any k,

$$\int \mu(x_1^k | x_{-\infty}^{-g}) \left| 1 - \frac{\mu(x_1^k)}{\mu(x_1^k | x_{-\infty}^{-g})} \right| d\mu(x_{-\infty}^{-g}) < \frac{\gamma^4}{4}.$$

Fix g and for each k define

$$f_k(x) = \frac{\mu(x_1^k)}{\mu(x_1^k | x_{-\infty}^{-g})}$$

and let $\mathscr{B}_k$ denote the σ-algebra determined by the random variables

$$\{X_i : i \leq -g\} \cup \{X_i : 1 \leq i \leq k\}.$$

Direct calculation shows that each f_k has expected value 1 and that $\{f_k\}$ is a martingale with respect to the increasing sequence $\{\mathscr{B}_k\}$. Thus f_k converges almost surely to some f.

Fatou's lemma implies that

$$\int | 1 - f(x) | \, d\mu \leq \frac{\gamma^4}{4},$$

so there is an M such that if

$$C_M = \left\{ x\colon \ |1 - f_k(x)| \leq \frac{\gamma^2}{2}, \ \forall \, k \geq M \right\},$$

then $\mu(C_M) > 1 - \gamma^2/2$. The ergodic theorem implies that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \mathscr{I}_{C_M}(T^{i-1}x) > 1 - \frac{\gamma^2}{2} \quad \text{a.s.},$$

where $\mathscr{I}_{C_M}$ denotes the indicator function of C_M, so that if we define

$$G_n(\gamma) = \left\{ x\colon \frac{1}{n} \sum_{i=1}^{n} \mathscr{I}_{C_M}(T^{i-1}x) > 1 - \frac{\gamma^2}{2} \right\},$$

then $x \in G_n(\gamma)$ eventually almost surely.

Let us put $k(\gamma) = M$ and let $t(\gamma)$ be any integer larger than $2/\gamma^2$. Fix $k \geq M$, $t \geq t(\gamma)$ and $(t+1)(k+g) \leq n < (t+2)(k+g)$, and fix an $x \in G_n(\gamma)$. The definition of $G_n(\gamma)$ and the assumption $t \geq 2/\gamma^2$ imply that

$$\frac{1}{t(k+g)} \sum_{i=1}^{t(k+g)} \mathscr{I}_{C_M}(T^{i-1}x)$$

$$= \frac{1}{(k+g)} \sum_{r=1}^{k+g} \frac{1}{t} \sum_{j=1}^{t} \mathscr{I}_{C_M}\big(T^{r+(j-1)(k+g)-1}x\big) > 1 - \gamma^2,$$

so there is a subset $R = R(x) \subseteq [1, k+g]$ of cardinality $|R| \geq (1-\gamma)(k+g)$ such that for $x \in G_n(\gamma)$ and $r \in R(x)$,

$$\frac{1}{t} \sum_{j=1}^{t} \mathscr{I}_{C_M}\big(T^{r+(j-1)(k+g)-1}x\big) > 1 - \gamma.$$

In particular, if $r \in R(x)$, then $T^{r+(j-1)(k+g)-1}x \in C_M$ for at least $(1-\gamma)t$ indices $j \in [1, t]$. However, if $T^{r+(j-1)(k+g)-1}x \in C_M$, then

$$\mu\big(\tilde{x}_j(r) | x_{-\infty}^{r+(j-1)(k+g)-g-1}\big) < (1+\gamma)\mu\big(\tilde{x}_j(r)\big),$$

which implies that j is a (γ, r, k, g)-splitting index for x.

In summary, for $x \in G_n(\gamma)$ and $r \in R(x)$ there are at least $(1-\gamma)t$ indices j in the interval $[1, t]$ that are (γ, r, k, g)-splitting indices for x. Since $|R(x)| \geq (1-\delta)(k+g)$, this completes the proof of Lemma 6B. □

THEOREM 4. *If μ is WB and $k(n) \leq (\log n)/(H+\varepsilon)$, $n = 1, 2, \dots$, then $\{k(n)\}$ is admissible for μ.*

PROOF. Fix $\delta > 0$, choose a positive $\gamma < 1/2$ and then choose integers $g = g(\gamma)$, $k(\gamma)$ and $t(\gamma)$ and measurable sets $G_n = G_n(\gamma)$, $n \geq 1$, so that conditions (a) and (b) of Lemma 6B hold. Fix $t \geq t(\gamma)$ and $(t+1)(k+g) \leq n < (t+2)(k+g)$, where $k(\gamma) \leq k \leq (\log n)/(H+\varepsilon)$. For each $r \in [1, k+g]$ and $J \subset [1, t]$, let $D_n(r, J)$ be the set of those sequences x for which every $j \in J$ is a (γ, r, k, g)-splitting index.

We have

$$\bigcap_{j \in J} B_{r,j} = D_n(r, J),$$

so that Lemma 6A and the fact that $|J| \leq t$ yield

$$\mu\bigg(\bigcap_{j \in J} [\tilde{x}_j(r)] \cap D_n(r, J)\bigg) \leq (1+\gamma)^t \prod_{j \in J} \mu(\tilde{x}_j(r)). \tag{102}$$

If $x \in G_n(\gamma)$, then Lemma 6B implies that there are $(1-\gamma)(k+g)$ indices $r \in [1, k+g]$ for each of which there are at least $(1-\gamma)t$ indices j in the interval $[1, t]$ that are (γ, r, k, g)-splitting indices for x.

On the other hand, it can be assumed that γ is so small and t so large that Lemma 5 assures that if $\big|\hat{\mu}_k(\cdot|x_1^n) - \mu_k\big| \geq \delta$, then $\big|\hat{\mu}_{k,g}^{r,J}(\cdot|x_1^n) - \mu_k\big| \geq \delta/4$ for at least $2\gamma(k+g)$ indices r, for any subset $J \subset [1,t]$ of cardinality at least $(1-\gamma)t$. Thus for γ sufficiently small and $k \geq k(\gamma)$ and $t \geq t(\gamma)$ sufficiently large, for any $x \in G_n(\gamma)$ there exists at least one $r \in [1, k+g]$ and at least one $J \subset [1,t]$ of cardinality at least $(1-\gamma)t$, for which $x \in D_n(r,J)$ and $\left|\hat{\mu}_{k,g}^{r,J}(\cdot|x_1^n) - \mu_k\right| \geq \delta/4$. This means that

$$\{x\colon |\hat{\mu}_k - \mu_k| \geq \delta\} \cap G_n(\gamma)$$
$$\subseteq \bigcup_{r=1}^{k+g} \bigcup_{\substack{J\subseteq[1,t]\\|J|\geq(1-\gamma)t}} \left(\left\{x\colon |\hat{\mu}_{k,g}^{r,J}(\cdot|x_1^n) - \mu_k| \geq \delta/4\right\} \cap D_n(r,J)\right).$$

The proof of Theorem 4 can now be completed very much like the proof for the ψ-mixing case. Using the argument of that proof, we can bound $\mu\{x\colon |\hat{\mu}_{k(n)} - \mu_{k(n)}| \geq \delta\} \cap G_n(\gamma)$ above by

$$2^{-2t\gamma\log\gamma}(1+\gamma)^t[k(n)+g](t+1)^{2^{k(n)(H+\varepsilon/2)}}2^{-t(1-\gamma)C\delta^2/400} \tag{103}$$

for t sufficiently large. This bound is the counterpart of (25), but here we used (102) in place of (23), and an extra factor, $2^{-2t\gamma\log\gamma}$, appeared to bound the number of subsets $J \subseteq [1,t]$ of cardinality at least $(1-\gamma)t$. If γ is small enough, then, as in the ψ-mixing case, (103) will be summable in n. Since $x \in G_n(\gamma)$, eventually almost surely, this establishes Theorem 4. □

Acknowledgment. We wish to thank Jeff Steif, whose careful reading of our paper brought to light our mistake, as well as some other less critical errors, and who suggested several improvements in the presentation.

MATHEMATICS INSTITUTE
HUNGARIAN ACADEMY OF SCIENCES
P.O.B. 127
1364 BUDAPEST
HUNGARY

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TOLEDO
2801 WEST BANCROFT STREET
TOLEDO, OHIO 43606
E-mail: pshield2@uoft2.utoledo.edu