Z. Wahrscheinlichkeitstheorie verw. Gebiete
60, 139–141 (1982)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

Correction to "Entropy and Maximal Spacings for Random Partitions"

Eric V. Slud

Department of Mathematics, University of Maryland, College Park, MD 20742, USA

Lemma 4.1 of Slud (1978, p. 348) is incorrect, and therefore Theorems 4.2(a) and Theorem 4.3 are wrong as stated. Both J.-C. Lootgieter and R. Dudley point out that if $\{Y_i\}_{i=1}^3$ is the uniform-selection sequence from the Kakutani-uniform partition tree $\{X_k\}_{k=1}^\infty \equiv \{X_k^\lambda(\omega)\}_{k=1}^\infty$ (for terminology and notations see Slud (1978)), then the law $\mathscr{L}(\{Y_j\}_{j=1}^3 \setminus \{X_i\}_{i=1}^2 \mid \{X_i\}_{i=1}^2 \subset \{Y_j\}_{j=1}^3)$ is a mixture with weights $P(X_2 = Y_2 \mid \{X_i\}_{i=1}^2 \subset \{Y_j\}_{j=1}^3)$ and $P(X_2 = Y_3 \mid \{X_i\}_{i=1}^2 \subset \{Y_j\}_{j=1}^3)$ of the uniform law $\mathscr{L}(Y_3 \mid X_1 = Y_1, X_2 = Y_2)$ and the non-uniform $\mathscr{L}(Y_2 \mid X_1 = Y_1, X_2 = Y_3)$, contradicting Lemma 4.1.

In this Note we replace Theorem 4.3 by a slightly weaker result and supply a corrected proof, still using an embedding method.

Theorem 4.3′. *Let μ_n denote the empirical measure of the Kakutani-uniform random sequence $\{X_k^\lambda(\omega)\}_{k=1}^\infty$ defined in Slud* (1978). *Then for each $t \in [0, 1]$, almost surely as $n \to \infty$, $|\mu_n([0, t]) - t| = \mathcal{O}(n^{-1/3} \log^2 n)$.*

This Theorem gives an almost-sure rate of equidistribution for $\{X_k^\lambda\}_{k=1}^\infty$ the same (up to a slowly varying function) as the best rate which can be derived from the proofs of van Zwet (1978) and Lootgieter (1977).

We acknowledge here the priority of van Zwet's and Lootgieter's solutions to Kakutani's Conjecture of equidistribution. However, the embedding method of Slud (1978) and of the present note is completely different from their proof idea.

As in Slud (1978), let $\{Y_i\}_{i=1}^N$ denote the uniform-selection sequence on the Kakutani-uniform partition tree $\{X_k^\lambda(\omega)\}_{k=1}^\infty$ and write $\{r_i\}_{i=1}^n \equiv \{X_k^\lambda(\omega)\}_{k=1}^n$, where n will approach ∞ along a subsequence $\{m_j\}$ which with $N \equiv N(n)$ will be defined below. For each n, N we let $\{Y'_{i,n,N}\}_{i=n+1}^N$ (or simply $\{Y'_i\}_{i=n+1}^N$) enumerate the points Y_k which subdivide spacings of length $\leqq L_n$, where L_n is the maximal spacing length among the points $\{r_i\}_{i=1}^n \cup \{0, 1\}$. It is clear from this definition that $\{Y_i\}_{i=1}^N = \{r_i\}_{i=1}^n \cup \{Y'_i\}_{i=n+1}^N$ almost surely for sufficiently large n and $N(n) \geqq [2n \cdot \log n]$ (by Theorem 2.1 of Slud, p. 343, and Theorem 4.2(b) as before). However we cannot now assert that $\{Y'_i\}_{i=n+1}^N$ is conditionally i.i.d. uniform given $\{r_i\}_{i=1}^n \subset \{Y_i\}_{i=1}^N$. The important sense in which this is approximately true is given by the following Lemma and Corollary.

0044-3719/82/0060/0139/$01.00

Lemma 4.1′. *Let* $\{\ell_n\}_{n=1}^{\infty}$ *be any fixed decreasing sequence with* $\ell_n \geqq (n+1)^{-1}$, *and let* $N(n) > n$. *Define* $K_n(t, \ell_n) \equiv \operatorname{card}\{k: 1 \leqq k \leqq N(n),\ Y_k \leqq t,\ Y_k$ *subdivides spacing of length* $\leqq \ell_n\}$, *and* $K_n \equiv K_n(1, \ell_n)$. *Then* $K_n(t, \ell_n) - tK_n$ *has expectation* $\mathcal{O}(N\ell_n + \log^2 N)$ *and variance* $\mathcal{O}(N \log N + N^2 \ell_n)$.

Proof. By exchangeability of uniform spacings, the initial point of a spacing interval of given length $\leqq \ell_n$ has distribution function approximately uniform, with error (given the spacings) at most ℓ_n or the largest spacing. Letting $\zeta_k(t) = \zeta_k(t, n) \equiv I[Y_k \leqq t,\ Y_k$ hits spacing $\leqq \ell_n]$, it follows by exchangeability and Lévy's (1939) result on maximal uniform spacings, that for all $d \geqq 0$, $j_1 < j_2 < \ldots < j_d < k$, with k, n, and N sufficiently large, for all $t \in [0, 1]$

$$|E(\zeta_k(t, n) - t \cdot \zeta_k(1, n) \mid \zeta_{j_1}(t), \zeta_{j_1}(1), \ldots, \zeta_{j_d}(t), \zeta_{j_d}(1))| \\ \leqq 2(d+1)(k^{-1} \log k + \ell_n). \qquad (*)$$

Summing $(*)$ in k with $d=0$ yields

$$E(K_n(t, \ell_n) - tK_n) = \sum_{k=1}^{N} E(\zeta_k(t, n) - t\zeta_k(1, n)) \\ = \mathcal{O}\left(\sum_{k=1}^{N} k^{-1} \log k + N\ell_n\right) = \mathcal{O}(\log^2 N + N\ell_n).$$

Similarly, summing $(*)$ for $d=1$ in j_1 and k yields

$$E(K_n(t, \ell_n) - tK_n)^2 = \mathcal{O}(N \log N + N^2 \ell_n).$$

Proof of Theorem. Let $\{m_k\}_{k=1}^{\infty}$ be an increasing sequence with $m_{k+1} - m_k \sim 3m_k^{2/3}$, $m_k \sim k^3$, as $k \to \infty$. Define $N_k \equiv [2m_k \log m_k]$, and let n_k be the random index $\min\{n: L_n < m_k^{-1}\}$, where L_n is the largest spacing among $\{X_i^\lambda\}_{i=1}^n \cup \{0, 1\}$. It is easy to see (cf. Slud, 1978, Prop. 3.2) that L_n is stochastically smaller than the largest spacing M_n for $\{Y_i\}_{i=1}^n \cup \{0, 1\}$. By Lévy's (1939) result on maximal spacings, a.s. for all sufficiently large k, $m_k \leqq n_k < 2m_k \log m_k$, $n_k \leqq N_k < 2n_k \log n_k$, and as in Slud (1978), $\{X_j^\lambda\}_{j=1}^{n_k} \subset \{Y_i\}_{i=1}^{N_k}$. Now Lemma 4.1′ with ℓ_n replaced by m_k^{-1}, N by N_k, implies $K_{n_k}(t, m_k^{-1}) - tK_{n_k} \equiv \operatorname{card}\{i: 1 \leqq i \leqq N_k: Y_i \leqq t$ hits spacing $\leqq m_k^{-1}\} - t \cdot \operatorname{card}\{i \leqq N_k: Y_i$ hits spacing $\leqq m_k^{-1}\}$ has mean $\mathcal{O}(\log^2 m_k)$, variance $\mathcal{O}(m_k \log^2 m_k)$ as $k \to \infty$. Hence with probability 1 for all sufficiently large k, the Borel-Cantelli Lemma implies $|K_{n_k}(t, m_k^{-1}) - tK_{n_k}| \leqq m_k^{2/3} \log^2 m_k$. That is, $\operatorname{card}(\{Y_j\}_{j=1}^{N_k} \setminus \{X_i^\lambda\}_{i=1}^{n_k}) \cap [0, t] - (N_k - n_k)\, t = \mathcal{O}(m_k^{2/3} \log^2 m_k)$ a.s. By the Lévy-Hinčin Law of the Iterated Logarithm, $\operatorname{card}\{Y_i\}_{i=1}^{N_k} \cap [0, t] - N_k t = \mathcal{O}((N_k \log\log N_k)^{1/2})$ a.s. for $k \to \infty$, and we conclude $\operatorname{card}\{X_i^\lambda\}_{i=1}^{n_k} \cap [0, t] - n_k t = \mathcal{O}(m_k^{2/3} \log^2 m_k) = \mathcal{O}(n_k^{2/3} \log^2 n_k)$ a.s.

Finally, for $n_k \leqq n < n_{k+1}$, $\operatorname{card}\{X_i^\lambda\}_{i=n_k}^{n} \cap [0, t] \leqq n_{k+1} - n_k =$ total number of points Y_i $(1 \leqq i < \infty)$ subdividing spacings of length between m_{k+1}^{-1} and m_k^{-1}. It is not hard to estimate directly using results of Darling (1953) that $n_{k+1} - n_k$ has mean $\mathcal{O}(m_k^{2/3} \log m_k)$ and variance $\mathcal{O}(m_k)$, from which the Borel-Cantelli Lemma again implies a.s. when $k \to \infty$, $n_{k+1} - n_k = \mathcal{O}(m_k^{2/3} \log m_k)$. Then for $n \to \infty$, a.s. $\{X_i\}_{i=1}^{n} \cap [0, t] - tn = \mathcal{O}(n^{2/3} \log^2 n)$ and our Theorem is proved.

As an example of the estimates mentioned above for $n_{k+1}-n_k$, we observe from Darling (1953) that the expected number of spacings for $\{Y_i\}_{i=1}^n$ between m_{k+1}^{-1} and m_k^{-1} in length is

$$n(n+1)\int_{m_{k+1}^{-1}}^{m_k^{-1}}(1-r)^{n-1}\,dr=(n+1)[(1-m_{k+1}^{-1})^n-(1-m_k^{-1})^n],$$

so that the probability that Y_{n+1} hits a spacing with length between m_{k+1}^{-1} and m_k^{-1} is asymptotically (for large k) $(n+1)\,m_k^{-1}((1-m_{k+1}^{-1})^n-(1-m_k^{-1})^n)$. The total expected number of such n is asymptotically (in k) $m_k^{-1}\sum_{n=1}^{\infty}(n+1)(1-m_{k+1}^{-1})^n-(1-m_k^{-1})^n)=\mathcal{O}(m_{k+1}-m_k)$. However, we do not supply further detailed estimates because the published method of van Zwet (1978) can be used to prove that in fact $E(n_{k+1}-n_k)=2(m_{k+1}-m_k)$ and $\operatorname{Var}(n_{k+1}-n_k)=\mathcal{O}(m_{k+1}-m_k)$, which is stronger than what we need.

Remarks. (i) Since our proof depends on the exchangeability of uniform spacings, it does not automatically extend to general subdividing distributions $\mathscr{m}$, as originally asserted in Sect. 5 of Slud (1978).

(ii) An important feature of our Embedding Method of proof of equidistribution is that it allows extensions (for uniform subdividing-distribution) to equidistribution by d-tuples, $d\geqq 2$, that is, as $n\to\infty$ the empirical measures on $[0,1]^d$ of $\{(X_k^\lambda, X_{k+1}^\lambda, \ldots, X_{k+d-1}^\lambda)\}_{k=1}^n$ almost surely converge weakly to Lebesgue measure on $[0,1]^d$. Such an extension for $d=2$ proceeds by considering (for fixed m_k, $N_k=[2m_k\log m_k]$, $\ell_{n_k}=m_k^{-1}$ as in Lemma 4.1'; $1\leqq j_1<j_2\leqq N_k$)

$$\begin{aligned}\zeta_{j_1,j_2}(t_1,t_2,\ell_{n_k})\equiv I[&Y_{j_1}\leqq t_1 \text{ and } Y_{j_2}\leqq t_2 \text{ each hit spacings } \geqq \ell_{n_k};\\ &Y_{j_1+1},\ldots,Y_{j_2-1} \text{ do } \textit{not} \text{ hit spacings } \geqq \ell_{n_k}].\end{aligned}$$

Using the fact that Y_j belongs to $\{X_k\}_{k=1}^n$ or to $\{Y_i'\}_{i=n+1}^N$ according as it hits a spacing of length respectively $>$ or $\leqq L_n$, one estimates via (*) that

$$\sum_{1\leqq j_1<j_2\leqq N}(\zeta_{j_1,j_2}(t_1,t_2,\ell_{n_k})-t_1t_2\,\zeta_{j_1,j_2}(1,1,\ell_{n_k}))$$

has mean $\mathcal{O}(\log^2 m_k)$ and variance $\mathcal{O}(m_k\log^2 m_k)$. It follows as in the Theorem above that for $d=2$ (and similarly all $d\geqq 2$) $\{(X_k^\lambda, X_{k+1}^\lambda, \ldots, X_{k+d-1}^\lambda)\}_{k=1}^n$ are a.s. equidistributed as d-tuples with rate $\mathcal{O}(n^{-1/3}\log^2 n)$.

References

Darling, D.A.: On a class of problems related to the random division of an interval. Ann. Math. Statist. **24**, 239–253 (1953)

Lootgieter, J.-C.: Sur la répartition des suites de Kakutani (I). Ann. Inst. H. Poincaré XIII, 385–410 (1977)

Slud, E.V.: Entropy and maximal spacings for random partitions. Z. Wahrscheinlichkeitstheorie verw. Geb. **41**, 341–352 (1978)

Van Zwet, W.R.: A proof of Kakutani's conjecture on random subdivision of longest intervals. Ann. Probab. **6**, 133–137 (1978)

Received July 30, 1980; in revised form on October 22, 1981