HOW NON-UNIFORM CAN A UNIFORM SAMPLE BE: A HISTOGRAM APPROACH

by

Jeesen Chen, Burgess Davis[1] and Herman Rubin[2]

Technical Report #84-23 (Revised)

Department of Statistics

University of Cincinnati and Purdue University

November 1984

AMS 1980 Subject classification: Primary 60F10 Secondary G2G20

Key Words: Empirical measures, nonparametric density estimation

[1]Supported by NSF Grant MCS 8201128

[2]Supported by NSF Grant DMS-8401996

ABSTRACT

Let μ_n be the empirical probability measure associated with n i.i.d. random vectors each having a uniform distribution in the unit square S of the plane. After μ_n is known, take the worst partition of the square into $k \leq n$ rectangles R_i, each with its short side at least δ times as long as the long side, and let $Z = n\sum|\mu_n(R_j) - \mu(R_j)|$. We prove distribution inequalities for Z implying the right half of $c_{p,\delta}(nk)^{p/2} \leq EZ^p \leq C_{p,\delta}(nk)^{p/2}$, $p > 0$. (The left half follows easily by considering non-random partitions.) Similar results are obtained in other dimensions, and for population distributions other than uniform, and our results are related to data based histogram density estimation.

1. Introduction

There are a number of ways to approach the question asked by the title of this paper. Ours has its origins in data based histogram density estimation. Histogram density estimators should be data based, the partition cells being made smaller in regions with high concentrations of observations. This paper investigates how badly such estimators can perform when the population has a bounded density in the unit cube Q_d in $\mathbb{R}^d$ and the cells are controlled in a manner to be described below. We use $\int_{\mathbb{R}^d} |\hat{f}-f| d\lambda^d$ as our measure of performance of estimators $\hat{f}$, where λ^d is d-dimensional Lebesgue measure. Besides being natural, this measure has the desirable quality of being invariant under change of scale, a property the sometimes used $\int_{\mathbb{R}^d} (f-\hat{f})^2 d\lambda^d$ does not possess.

We begin by considering the case where the population distribution is uniform in the unit cube in $\mathbb{R}^d$. All the difficulties we surmount are present here. We focus on what happens as the sample size and the number of cells approach infinity. Let $X_1, X_2, \ldots$ be i.i.d. each uniformly distributed on Q_d, and let, for each Borel set A, $\lambda_n^d(A)$ be the empirical measure of A based on $X_1, X_2, \ldots X_n$, that is

$$\lambda_n^d(A) = \#\{j: 1 \leq j \leq n \text{ and } X_j \in A\}/n.$$

If $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$ is a partition of Q_d into disjoint Borel sets we define

$$Z_n(\pi) = n \sum_{j=1}^{k} |\lambda_n^d(\pi_j) - \lambda^d(\pi_j)| .$$

We note that $Z_n(\pi) = n\int_{Q_d} |f_n(\pi) - 1| d\lambda^d$, where $f_n(\pi)$ is the histogram density estimator based on λ_n^d and the partion π. To avoid trivialities, we always assume there are at least two cells in every partition.

Let $\mathbb{P}(k)$ be the collection of all partitions of [0,1] into k intervals. We prove the following theorem for d = 1.

<u>Theorem 1</u>: Given $\alpha > 0$ there is a constant $r_0(\alpha) = r_0$ (not depending on n or k) such that

$$P(\sup_{\pi \in \mathbb{P}(k)} Z_n(\pi) \geq r\sqrt{nk}) < \alpha^r \ \underline{\text{if}}\ r \geq r_0.$$

For $d \geq 2$ and $\delta > 0$, let $\mathbb{P}(d,k,\delta)$ denote the collection of all partitions of Q_d into k parallelopipeds oriented to the coordinate axes, such that the shortest edge of each parallelopiped is at least δ times as long as its longest edge. We assume δ is not so large that $\mathbb{P}(d,k,\delta)$ is empty. We prove

<u>Theorem 2</u> <u>For each</u> $d \geq 2$, $\delta > 0$ <u>and</u> $\beta > 0$ there is a constant $\gamma_0 = \gamma_0(d,\delta,\beta)$ <u>such that</u>

$$P(\sup_{\pi \in \mathbb{P}(d,k,\delta)} Z_n(\pi) \geq r\sqrt{nk}) < \beta^r \ \underline{\text{if}}\ r \geq \gamma_0.$$

Theorems 1.1 and 1.2 give the right hand sides of

(1.1) $c_p(nk)^{p/2} \leq E \sup_{\pi \in \mathbb{P}(k)} Z_n(\pi)^p \leq C_p(nk)^{p/2}$, $p > 0$, and

(1.2) $c_p(nk)^{p/2} \leq E \sup_{\pi \in \mathbb{P}(d,k,\delta)} Z_n(\pi)^p \leq C_{p,d,\delta}(nk)^{p/2}$, $p > 0$.

Here the constants do not depend on n or k. The left hand sides of these inequalities follow from the easily proved fact that for a non-random partition φ of Q_d into $k \leq n$ cells each of λ_d measure at least $(2k)^{-1}$ we have $P(Z_n(\varphi) > (nk)^{\frac{1}{2}}) > c$ where c is a constant independent of n and k. The proof of this is sketched at the end of Section 2. In terms of density estimation, (1.1) and (1.2) say that ratio of the worst case performance of our estimators to the best that can be hoped for, is not affected too much, i.e., infinitely much, by n and k.

We have three proofs of Theorem 1. One is based on square functions inequalities for BMO martingales. The second was communicated to us by Kenneth Alexander, who showed us how to use results from his recent paper [1] to not only prove Theorem 1, but also get better estimates on the extreme tail of the distribution of $\sup_{\pi \in \mathbb{P}(k)} Z_n(\pi)$ than those given here. We were surprised that Alexander's results were general enough to be applicable here, and somewhat relieved that they (apparently) do not give Theorem 2 of this paper. The third proof of Theorem 1 is the one we give, since it is the only one which extends to prove Theorem 2.

We now discuss the situation where the population density f has support in [0,1] but is otherwise arbitrary. If π is a partition, let $\bar{f}(\pi)(x)$ be the average density of the π cell containing x, and let $f_n(\pi)$ be the usual histogram density estimator based on π. Note that

$$\int_{\mathbb{R}^d} |f_n(\pi) - f(\pi)| d\lambda^d \leq \int_{\mathbb{R}^d} |f_n(\pi) - \bar{f}(\pi)| d\lambda^d + \int_{\mathbb{R}^d} |\bar{f}(\pi) - f(\pi)| d\lambda^d ,$$

The first term being the randomness error and the second the roughness error.

In dimension 1, the usual scaling argument gives that if $\Gamma(k)$ is the class of all k interval partitions of $(-\infty,\infty)$ then

$$\sup_{\pi \in \Gamma(k)} \int_{\mathbb{R}^d} |f_n(\pi) - \bar{f}(\pi)| d\lambda^1$$

has exactly the distribution of $\sup_{\pi \in \mathbb{P}(k)} Z_n(\pi)$. Thus Theorem 1 instantly translates to a theorem about the randomness error for data based histogram density estimators of arbitrary distributions with support in [0,1].

In dimensions higher than one, scaling is not available. However, the proof of Theorem 2 can be readily altered to yield the following.

Theorem 3. Let f be a density function which vanishes outside Q_d and which is bounded by the constant C. Given $\beta > 0$ there is a constant $r_o = r_o(\delta, d, C, \beta)$ such that

$$P(\sup_{\pi \in \mathbb{P}(d,k,\delta)} \int_{Q_d} |\bar{f}(\pi) - f_n(\pi)| d\lambda^d \geq r\sqrt{nk}) < \beta^r \text{ if } r \geq r_o.$$

Two of us (Chen and Rubin) have given further applications of the results of this paper to data based density estimation in [3].

2. Proof of Theorem 1

In this section, we prove Theorem 1. Without loss of generality, we can and do assume that $k < n/2$, since for all k $\sup_{\pi \in \mathbb{P}(k)} Z_n(\pi) \leq 2n$. Furthermore, we may and do assume that

$$k = 2^{i_o},$$

and

$$2^{i_1} \leq n < 2^{i_1+1}$$

for integers $i_o \leq i_1$. In this section we shall use λ to replace λ^1, and λ_n to replace λ_n^1.

Suppose we Poissonize the sample size n, i.e. we take a sample of size N, where N has Poisson distribution with mean n and independent of X_j's. Define for each set $A, N^*(A) = \#\{j: 1 \leq j \leq N \text{ and } X_j \in A\}$, and $N_i^* = N^*(\pi_i)$.

Notice that all of the $N_i^* - n\lambda_n(\pi_i)$, $i = 1,2,\ldots,k$, have the same sign and add up to N-n, so we have

$$\left|\sum_{i=1}^{k} |N_i^* - n\lambda(\pi_i)| - \sum_{i=1}^{k} |n\lambda_n(\pi_i) - n\lambda(\pi_i)|\right| \leq |N-n|.$$

Therefore, if $Z_n^* = \sum_{i=1}^{k} |N_i^* - n\lambda(\pi_i)|$,

<u>Proposition 1</u> $\qquad Z_n(\pi) = \leq |N-n| + Z_n^*(\pi)$.

Because of Proposition 1, Theorem 1 is an easy corollary of the following two propositions.

<u>Proposition 2</u> <u>For all</u> $\alpha > 0$ <u>there is a constant</u> $r_1 = r_1(\alpha)$ such that

$$P(|N-n| \geq r\sqrt{n}) \leq \alpha^r \text{ for } r \geq r_1.$$

<u>Proposition 3</u> <u>For any</u> $\alpha > 0$ <u>there is a constant</u> $r_2 = r_2(\alpha)$ such that

$$P(\sup_{\pi \in \mathbb{P}(k)} Z_n^*(\pi) \geq r\sqrt{nk}) \leq \alpha^r \text{ for } r \geq r_2.$$

Proof of Proposition 2:

$$(2.1) \qquad \begin{aligned} &P(|N-n| \geq r\sqrt{n}) \\ &\leq \inf_{s>0} e^{-rs\sqrt{n}} E(e^{s|N-n|}) \\ &\leq \inf_{s>0} e^{-rs\sqrt{n}} [E(e^{s(N-n)}) + E(e^{-s(N-n)})] \\ &\leq \inf_{s>0} e^{-rs\sqrt{n}} [\exp(n(e^{s}-s-1)) + \exp(n(e^{-s}+s-1))] \\ &\leq 2 \inf_{s>0} \exp(-r s\sqrt{n} + n(e^{s}-s-1)), \end{aligned}$$

since $N \sim$ Poisson (n) and $e^s - s - 1 \geq e^{-s} + s - 1$ for $s \geq 0$.

Let $\varphi_0(s) = -r s\sqrt{n} + n(e^s - s - 1)$. Notice that $\varphi_0(s)$ has its minimum value

$$\begin{aligned}(2.2) \qquad & \varphi_0(\log(1 + r/\sqrt{n})) \\ &= -n[(1+r/\sqrt{n})\log(1+r/\sqrt{n}) - r/\sqrt{n}] \\ &= -n\,\psi(r/\sqrt{n})\end{aligned}$$

where

$$\psi(x) = (1+x)\log(1+x) - x = \int_0^x \log(1+y)dy.$$

Elementary calculations show that $\psi(x) \geq \frac{1}{3}x^2$ for $x \leq 1$, and it is easily seen that $\psi(x)/x \to \infty$ as $x \to \infty$. Together with (2.1) and (2.2) this gives Proposition 2.

Before proving Proposition 3, we need the following notations and results. Define

$$T(u) = \#\{j\colon 0 \leq X_j \leq u,\ 1 \leq j \leq N\} - nu \quad \text{for } 0 \leq u \leq 1.$$

It is well-known that

Proposition 4. $T(u)$ has stationary independent increments and is a martingale.

For B an interval with endpoints $a < b$ we define

$$S(B) = \sup_{u,v \in B} |T(u) - T(v)|,$$

$$S_0(B) = \sup_{u \in B} |T(u) - T(a)|, \text{ and}$$

$$S_y(B) = \max(0, S_0(B) - y\sqrt{n/k}).$$

It is easy to see $S(B) \leq 2S_0(B)$.

Proposition 5. If $B = (a,b)$, then

$$E(e^{sS_0(B)}) \leq 4\,E(e^{s|T(b)-T(a)|}).$$

Proof: The process $Z(x) = \exp(s|T(x) - T(a)|/2)$ is, for each positive s, a convex function of the martingale T(x), and so is a submartingale. An inequality of Doob ([4], p. 317) says that if $M_1,\ldots,M_n$ is a nonnegative submartingale then

$$E \max_{1\le k\le n} M_k^p \le \left(\frac{p}{p-1}\right)^p E M_n^p, \quad p > 1.$$

The continuous time version of this inequality follows immediately, and applying it to the situation at hand we get

(2.3)
$$\begin{aligned} & E(e^{sS_0(B)}) \\ &= E((\exp(\sup_{x\in B} \tfrac{s}{2} |T(x)-T(a)|))^2) \\ &\le \left(\frac{2}{2-1}\right)^2 E((\exp \tfrac{s}{2} |T(b)-T(a)|)^2) \\ &= 4\, E(e^{s|T(b)-T(a)|}). \end{aligned}$$

Remark: By using Doob's inequality for $\exp(\lambda s|T(x)-T(a)|)$, $\lambda \to 0$, we can replace 4 with e.

Proposition 6. If $B = (a,b]$, and $y > 0$, then

$$E(e^{s\, S_y(B)}) \le 1 + 4\, e^{-s\, y\sqrt{n/k}}\, E(e^{s|T(b)-T(a)|}) .$$

Proof: $E(e^{s \max(0, S_o(B) - y\sqrt{n/k})})$

$$\begin{aligned} &\le 1 + E(e^{s(S_o(B) - y\sqrt{n/k})}) \\ &= 1 + e^{-sy\sqrt{n/k}}\, E(e^{s\, S_o(B)}). \end{aligned}$$

Proposition 6 now follows from Proposition 5.

In the following, we denote $[0,2^{-i}]$ by I_{io}, and put

$$I_{ij} = (2^{-i}(j-1), 2^{-i}j],\ j \neq 0,$$

and

$$\mathcal{B}(i_0;i_1) = \{I_{ij}:\ i_0 \leq i \leq i_1,\ j = 1,2,\ldots,2^i\}.$$

For each partition $\pi \in \mathbb{P}(k)$, if no internal division points of the partition π are binary rationals, (we may and do make this assumption without loss of generality), we associate dyadic intervals I_{ij} in $\mathcal{B}(i_0,i_1)$ with cells in $\mathbb{P}(k)$ in this way:

I. If cell π_ℓ satisfies $|\pi_\ell| > 2^{-i_0}$, associate to π_ℓ each dyadic interval of length 2^{-i_0} which has non-empty intersection with π_ℓ.

II. If $2^{-j} < |\pi_\ell| \leq 2^{-j+1}$ for some j, $i_1 > j > (i_0+1)$, associate to π_ℓ each dyadic interval of length 2^{-j} which has non-empty intersection with π_ℓ.

III. If $|\pi_\ell| \leq 2^{-i_1}$, associate to π_ℓ each dyadic interval of length 2^{-i_1} which has non-empty intersection with π_ℓ.

We note that each cell π_ℓ is covered by the dyadic intervals associated with it. Furthermore the collection of all those intervals associated with some cell of the partition has the following properties.

(P1) Each dyadic interval of length exceeding 2^{-i_1} is used at most twice.

(P2) At most $3k$ dyadic intervals are used.

Proposition 7. Let $w > 24$. Then

$$\sup_{\pi \in \mathbb{P}_k} Z_n^*(\pi) \leq 7w\sqrt{nk} + 4\sum_{i=i_0}^{i_1} \sum_{j=1}^{2^i} S_w(I_{ij}).$$

Proof: We observe that, for each $\pi \in \mathbb{P}(k)$, $Z_n^*(\pi)$ is bounded by the sum of all $N^*(\pi_\ell \cap I_{ij}) - n\lambda(\pi_\ell \cap I_{ij})$ where only terms with π_ℓ associated with I_{ij} are used. If $i < i_1$ only two π_ℓ can be associated with I_{ij}, and this term is bounded by $S(I_{ij}) < 2w\sqrt{n/k} + 2S_w(I_{ij})$. If $i = i_1$, we use the crude bound $N^*(\pi_\ell \cap I_{ij}) + n\lambda(\pi_\ell \cap I_{ij})$. Then for each I_{ij} used, the sum over all the π_ℓ associated with I_{ij} is bounded by

$$N^*(I_{ij}) + n \cdot 2^{-i_1} \leq S_0(I_{ij}) + 2n \cdot 2^{-i_1} < S_w(I_{ij}) + w\sqrt{n/k} + 4.$$

We get the result by using the bounds from (P1) and (P2) on the number of terms of the form $w\sqrt{n/k}(+4)$, and by including all the $S(I_{ij})$. q.e.d.

We are ready to prove Proposition 3.

Proof: Using Proposition 7, for $r \geq 24$ we have

$$(2.4) \qquad P(\sup Z_n^*(\pi) \geq 16r\sqrt{nk})$$

$$\leq P\left(\sum_{i=i_0}^{i_1} \sum_{j=1}^{2^i} S_r(I_{ij}) \geq r\sqrt{nk}/4\right)$$

$$\leq \sum_{i=i_0}^{i_1} P\left(\sum_{j=1}^{2^i} S_r(I_{ij}) \geq c_i\, r\sqrt{nk}\right)$$

for any sequence c_i such that $\sum_{i=i_0}^{i_1} c_i \leq 1/4$. We shall use $c_i = \xi 2^{(i_0-i)/4}$ with $\xi = (1-2^{-\frac{1}{4}})/4$.

Now

$$(2.5)\qquad P\left(\sum_{j=1}^{2^i} S_r(I_{ij}) \geq c_i\, r\sqrt{nk}\right)$$

$$\leq \inf_{s\geq 0} e^{-s\, c_i\, r\sqrt{nk}}\, E\left(e^{s\sum_{j=1}^{2^i} S_r(I_{ij})}\right)$$

$$= \inf_{s>0} \exp\left(-s\, c_i\, r\sqrt{nk} + 2^i \log E\left(e^{s\, S_r(I_{ij})}\right)\right),$$

because $T(x)$ is an additive process.

Let $\varphi_i(s) = -s\, c_i r\sqrt{nk} + 2^i \log E(e^{sS_r(I_{ij})})$

For $s > 0$ we have

$$\log E(e^{sS_r(I_{ij})})$$

$$\leq \log [1+e^{-sr\sqrt{n/k}}\, Ee^{S_o(I_{ij})}]$$

$$\leq \log [1+4e^{-sr\sqrt{n/k}} E\, e^{s|T(2^{-i}j) - T(2^{-i}(j-1))|}]$$

(using Proposition 5)

$$\leq 4e^{-sr\sqrt{n/k}}\, E\, e^{s|T(2^{-i}j) - T(2^{-i}(j-1))|}$$

Thus

$$(2.6)\qquad \varphi_i(s) \leq -s\, c_i r\sqrt{nk} + 2^i \cdot 4e^{-s\,r\sqrt{n/k}}\, Ee^{s\cdot|T(2^{-i}j) - T(2^{-i}(j-1))|}$$

$$\leq -s\, c_i r\sqrt{nk} + 8\cdot 2^i e^{-sr\sqrt{n/k}} \exp[n(e^s-s-1)/2^i],$$

this last step holding since, if Z is a Poisson random variable having mean μ and if $X = Z-\mu$, then for positive s, $Ee^{s|X|} \leq Ee^{sX} + Ee^{-sX} \leq 2Ee^{sX}$.

Let $A(r,s) = n(e^s-s-1)2^i-sr\sqrt{n/k}$ be the exponent in the last part of (2.6). Then if h is a positive number, which will be considered fixed for awhile, we have, since $2^i \leq n$, that there is a constant C_h depending only on h such that if $r \geq C_h$ then $A(r,h\sqrt{2^i/n}) \leq e^h-h-1-rh\sqrt{2^i/k} \leq -\sqrt{2^i/k}$.

Thus, plugging this estimate into (2.6), and putting $\lambda = 2^i/k$, we get

$$\begin{aligned}\varphi_i(h\sqrt{2^i/n}) &\leq -\xi rkh\lambda^{\frac{1}{4}} + 8\lambda ke^{-\sqrt{\lambda}}\\ &\leq k\lambda^{\frac{1}{4}}(-\xi rh+5)\\ &\leq -\xi hrk\lambda^{\frac{1}{4}}/2\\ &= -\xi hrk(2^i/k)^{\frac{1}{4}}/2,\end{aligned}$$

if $\xi rh \geq 10$.

Now if $hr \geq 2\log 2/\xi(2^{\frac{1}{4}}-1)$, we have

$$-\xi hrk(2^{i+1}/k)^{\frac{1}{4}}/2 \leq -\xi hrk(2^i/k)^{\frac{1}{4}}/2 - \log 2, \text{ so that}$$

$$\sum_{i=i_0}^{i_1} \exp(\varphi_i(h\sqrt{2^i/n})) \leq 2\exp(\xi hrk/2),$$

recalling that $2^{i_0} = k$. Plugging this into (2.4) and using the definition of φ_i, we get

$$\begin{aligned}P(\sup Z_n^*(\pi) \leq 16r\sqrt{nk}) &\leq 2\exp(-\xi hrk/2)\\ &= 2\exp(-\xi hk/2))^r\\ &\leq 2(\exp(-\xi h/2))^r,\end{aligned}$$

so that given $\theta > 0$ there is an $h = h(\theta)$ such that if $r > C_{h(\theta)}$ and $r \geq 10/\xi h(\theta)$ we have $P(\sup Z_n^*(\pi) \geq 16r\sqrt{nk}) \leq 2\theta^r$, which is readily seen to imply Proposition 3. q.e.d. Of course, better bounds in the above could theoretically be gotten by using the value for s which minimizes A(r,s) in place of $h\sqrt{2^i/n}$. This value is $\log(1 + 2^i r/\sqrt{nk})$. It is difficult to work with.

Now the left hand inequality in (1.1) will be proved. The proof of the left hand inequality in (1.2) is almost identical. Suppose X has a binomial distribution with parameters m and p. There is an absolute constant $c > 0$ such that

$$(2.7) \qquad P(|X-mp| \geq c\sqrt{mp}) \geq c, \quad \text{if } 1/2 \leq mp \leq m/2.$$

To see this let $Y = X-mp$. Routine calculations give the existence of a constant γ such that, if $1/2 \leq mp \leq m/2$, $(EY^4)^{\frac{1}{4}} \leq \gamma(EY^2)^{\frac{1}{2}}$. This implies there is positive constant ε such that $P(Y^2 > \varepsilon EY^2) > \varepsilon$, yielding (2.7). Now divide [0,1] into k intervals π_i each having length at least $(2k)^{-1}$ and at most 1/2. Since $\lambda_n(\pi_i)$ is binomial for each i, if we put $Y_i = |\lambda_n(\pi_i)-\lambda(\pi_i)|$ inequality (2.7) yields $P(Y_i > c\sqrt{n/k}) > c$. This implies the existence of a postive constant such that

$$P(\sum_{i=1}^{k} Y_i > \beta\sqrt{nk}) > \beta,$$

qed.

3. Proof of Theorem 2.

Now we indicate how to adapt the argument of the last section to prove Theorem 2. Each of the dimensions greater than one can be handled the same way, so we just treat the case d=2.

In this section λ will stand for Lebesgue measure on the unit square $[0,1]\times[0,1] = Q$, $X_1, X_2, \ldots$ will be a sequence of iid random vectors uniformly distributed in Q, and N will be a Poisson random variable, having mean n, which is independent of the sequence $X_1, X_2, \ldots$. For each measurable Set $A \subset Q$ we put

$$\Gamma(A) = \{\# j : X_j \in A,\ 1 \leq j \leq N\} - n\lambda(A),$$

and define $\Gamma(s,t) = \Gamma([0,s]\times[0,t])$. By a dyadic square we mean the product of the two dyadic intervals of equal lengths.

The proof of Theorem 2 in the d=2 case parallels the proof of Theorem 1, with dyadic squares used in place of dyadic intervals. We assume that $k = 2^{2i_0}$ for integer i_0, and define i_1 by $2^{2(i_1+1)} > n \geq 2^{2i_1}$. Poissonization works exactly as before, and we cover our partition rectangles with dyadic squares in the following way:
If the partition rectangle π_ℓ has smallest side length exceeding 2^{-i_0}, we associate to π_ℓ all dyadic squares of side length 2^{-i_0} which intersect π_ℓ, and continue the association algorithm in exact analogy to the one dimensional case, where now the smallest side length replaces the interval length we used in one dimension.

The properties (P1) and (P2) are now replaced with

(P1)' Each dyadic square of side length exceeding 2^{-i_1} is used at most $K_1(\delta)$ times.

(P2)' The total number of dyadic squares used is at most $K_2(\delta)k$.

Here $K_1(\delta)$ and $K_2(\delta)$ depend only on δ.

Once the covering argument had been made in the last section, it was necessary to get a bound on the variation of the process T over π in terms of its variation over the dyadic cover. The key for this was the inequality

(3.1) $$E \sup_{a \leq u \leq v \leq b} |f(T(u)) - f(T(v))|^p \leq C_p E|f(T(b)) - f(T(a))|^p, \quad p > 1,$$

whenever f is positive, increasing, and convex. This was needed since the intersection of a dyadic interval with a partition interval is an interval contained in the dyadic interval. Inequality (3.1) followed from the special case of Doob's inequality

(3.2) $$E \sup_{a \leq u \leq b} |f(T(u)) - f(T(a))|^p \leq \left(\frac{p}{p-1}\right)^p E|f(T(b)) - f(T(a))|^p.$$

Now, returning to the plane, let f remain a positive, convex increasing function, and define the variation of Γ over the rectangle $R=(a,b)\times(c,d)$ in the usual manner,

$$V_R(\Gamma) = \Gamma(b,d) + \Gamma(a,c) - \Gamma(a,d) - \Gamma(b,d).$$

The following holds.

Proposition 8. <u>There is a constant</u> C_p <u>such that if</u> G <u>is a square with sides oriented to the coordinate axes and if</u> $\mathbb{R}$ <u>is the collection of all oriented subrectangles of</u> G, then

$$\sup_{E\in\mathbb{R}} Ef(V_E(\Gamma))^p \le C_p Ef(V_G(\Gamma))^p.$$

This proposition is an immediate consequence of a two parameter martingale theorem of Cairoli ([2]). We will state the needed corollary of Cairoli's theorem and, for completeness, give a brief sketch of the proof.

Proposition 9. <u>Using the notation of Proposition 8, let</u> $\mathcal{B}$ <u>be the collection of all oriented subrectangles of</u> G <u>which have one corner at</u> (a,b). <u>Then</u>

$$\sup_{E\in\mathcal{B}} Ef(V_E(\Gamma))^p \le \left(\frac{p}{p-1}\right)^{2p} Ef(V_G(\Gamma))^p.$$

Proof. Define $D(s,t) = f(V_{(a,b)x(s,t)})$.

Put $L(s) = \sup_{b\le t\le d} D(s,t)$. Then $L(s)$, $a\le s\le c$, is a submartingale with respect to the σ-fields $\mathcal{O}_s = \sigma(D(x,t),\ a\le x\le s,\ b\le t\le d)$. Very roughly, here is why. Let $a \le s_0 < s_1 \le b$ and put $\delta > 0$.

Let $T = \inf\{t \geq b : D(s_0,t) \geq L(s_0)-\delta\}$, and note T is $\mathcal{O}_{s_0}$ measurable. Now, given $\mathcal{O}_{s_0}$, $D(s_1,T) - D(s_0,T)$ is conditionally independent of $D(s_0,T)$, and has conditional mean 0. Thus $E(D(s_1,T)|\mathcal{O}_{s_0}) = D(s_0,T)$ a.s., so that

$E(L(s_1)|\mathcal{O}_{s_0}) = E(\sup_t D(s_1,t)|\mathcal{O}_{s_0}) \geq D(s_0,T) > L(s_0) - \delta$ a.s., giving $E(L(s_1)|\mathcal{O}_{s_0}) \geq L(s_0)$ a.s. Since $L(s)$ is a submartingale, use Doob's inequality to get

$$(3.3) \qquad E \sup_{a \leq s \leq c} L(s)^p \leq \left(\frac{p}{p-1}\right)^p E\, L(c)^p.$$

It is also true that $D(c,t)$, $b \leq t \leq d$, is a submartingale. Again using Doob's inequality, we obtain

$$E\, L(c)^p = E \sup_{b \leq t \leq d} D(c,t)^p \leq \left(\frac{p}{p-1}\right)^p ED(c,d)^p$$

and putting this together with (3.3) we get

$$E \sup_{\substack{a \leq s \leq c \\ b \leq t \leq d}} D(s,t)^p = E \sup_{a \leq s \leq c} L(s)^p \leq \left(\frac{p}{p-1}\right)^{2p} E\, D(c,d)^p,$$

as was to be shown.

The proof of Theorem 2 can now be completed in exact parallel to the proof of Theorem 1.

References

[1] Alexander, K. S. (1984) Probability inequalities for empirical processes and a law of the iterated logarithm. Ann. Prob. 12 pp. 1041-1067.

[2] Cairoli, R. (1970) Une inegalité pour martingales á indices multiples et ses applications. Seminaré de Strasbourg; Springer, Berlin 1970, 1-27.

[3] Chen, J. and Rubin, H. (1984) On the consistency property of the data-based histogram density estimators, Purdue mimeograph series #84-11.

[4] Doob, J. L. (1953) Stochastic Processes. John Wiley & Sons, Inc. New York.

Jeesen Chen
Department of Mathematics
University of Cincinnati
Cincinnati, OH 45221

Burgess Davis
Department of Statistics
Purdue University
W. Lafayette, IN 47907

Herman Rubin
Department of Statistics
Purdue University
W. Lafayette, IN 47907