Z. Wahrscheinlichkeitstheorie verw. Gebiete
58, 139–167 (1981)

Zeitschrift für
Wahrscheinlichkeitstheorie
und verwandte Gebiete

On the Maximum Deviation Between the Histogram and the Underlying Density

David Freedman[1]* and Persi Diaconis[2]**

[1] Statistics Department, University of California, Berkeley, CA, USA
[2] Statistics Department, Stanford University, Palo Alto, CA, USA

Summary. The limiting joint distribution of the location and size of the maximum deviation between the historgram and the underlying density is derived. For smooth densities, the location and size of the maximum are asymptotically independent. The size has a limiting double-exponential distribution and the location has a limiting normal distribution.

1. Introduction

A sample of size k is drawn from a density f.

$$f \geqq 0 \quad \text{and} \quad \int_{-\infty}^{\infty} f(x)\,dx = 1. \tag{1.1}$$

A histogram of cell width h is used to estimate f. How far off is the estimate? Where does the maximum discrepancy occur? The object of this paper is to describe the asymptotic joint distribution of these two variables.

To state a precise result, assume that

$$f \text{ has a unique maximum at } x_0. \tag{1.2}$$

Assume too that f is locally quadratic at x_0:

$$f(x_0+x) = f(x_0) + \tfrac{1}{2}\alpha x^2 + o(x^2) \qquad \text{as } x \to 0, \tag{1.3}$$

where α is negative; write $f''(x_0) = \alpha$. This does not assume any differentiability; however, if f is smooth, then α is the ordinary second derivative at x_0. Finally, assume

$$\sup_x \{f(x_0+x) \colon |x| \geqq \delta\} < f(x_0) \qquad \text{for any } \delta > 0. \tag{1.4}$$

* Research partially supported by NSF grant MCS-80-02535
** Research partially supported by NSF grant MCS-77-16974

0044-3719/81/0058/0139/$05.80

For ordinary functions, (1.2) is equivalent to (1.3) and (1.4).

To define the histogram, choose a point λ_0 with $\lambda_0 \leqq x_0 < \lambda_0 + h$. By definition, cell j of the histogram will run from $\lambda_j = \lambda_0 + hj$ to $\lambda_{j+1} = \lambda_0 + h(j+1)$:

$$\text{(1.5)} \qquad \text{cell } j = [\lambda_j, \lambda_{j+1}) = [\lambda_0 + hj, \lambda_0 + h(j+1)), \quad j = 0, \pm 1, \pm 2, \ldots$$

The data consists of k independent random variables $X_1, X_2, \ldots, X_k$ with common probability density f. By definition, N_j is the number of data points falling in cell j. Formally,

$$\text{(1.6)} \qquad N_j \text{ is the number of indices } i = 1, \ldots, k \text{ with } \lambda_j \leqq X_i < \lambda_{j+1}.$$

By definition, the histogram is

$$\text{(1.7)} \qquad H(x) = N_j/(kh) \quad \text{for } x \in [\lambda_j, \lambda_{j+1}).$$

This definition forces the area under $H(x)$ to be 1. Let p_j be the probability of the j^{th} cell:

$$\text{(1.8)} \qquad p_j = \int_{\lambda_j}^{\lambda_{j+1}} f(x)\, dx.$$

Define $f_h(x)$ to be p_j/h for x between λ_j and λ_{j+1}.

The difference between the histogram and the density can be decomposed as:

$$\text{(1.9)} \qquad H(x) - f(x) = H(x) - f_h(x) + f_h(x) - f(x).$$

The term $H(x) - f_h(x)$ represents sampling error; $f_h(x) - f(x)$ represents bias. When h is small, sampling error dominates and the distribution of $\sup_x [H(x) - f(x)]$ is the same as the distribution of $\frac{1}{kh} \sup_j (N_j - kp_j)$. For this reason, it is useful to derive the distribution of the location and size of $\sup_j (N_j - kp_j)$. The following growth condition will be needed:

$$\text{(1.10)} \qquad k \to \infty \quad \text{and} \quad h \to 0 \quad \text{in such a way that } k \Big/ \left[\frac{1}{h}\left(\log \frac{1}{h}\right)^3\right] \to \infty.$$

In the absence of this condition, large-deviations corrections to the central limit theorem become relevant: see [2] for a related discussion. A final burst of notation:

$$\text{(1.11)} \qquad w_h(x) = \left[2 \log \frac{1}{h} - 2 \log \log \frac{1}{h} + x\right]^{1/2},$$

$$\text{(1.12)} \qquad \Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-u^2/2}\, du,$$

$$\text{(1.13)} \qquad 2\rho^2 = |f''(x_0)|/f(x_0).$$

The first result can now be stated:

(1.14) **Theorem.** *Assume* (1.1–13). *With probability approaching one,* $M_{kh} = \max_j (N_j - kp_j)$ *is taken on at a unique index* L_{kh}. *Moreover,* M_{kh} *and* L_{kh} *are asymptotically independent,* M_{kh} *being asymptotically double exponential and* L_{kh} *asymptotically normal. More precisely, the chance that*

$$\rho\sqrt{2\log\frac{1}{h}}\cdot h L_{kh} < y$$

and

$$M_{kh}/\sqrt{kh} < \sqrt{f(x_0)}\, w_h(x)$$

converges to

$$\Phi(y)\cdot\exp\left\{-\frac{1}{2\rho}e^{-x/2}\right\}.$$

In particular, the maximum discrepancy between N_j and kp_j occurs on the order of $h^{-1}\left(\log\frac{1}{h}\right)^{-1/2}$ cells away from the place x_0 where the density is maximum. Since the cell width is h, the maximum discrepancy occurs at a distance on the order of $\left(\log\frac{1}{h}\right)^{-1/2}$ from x_0. A final comment on the scaling in (1.14): for $x \in$ cell j,

$$(kh)^{1/2}[H(x) - f_h(x)] = (kh)^{-1/2}(N_j - kp_j) \tag{1.15}$$

so

$$(kh)^{1/2}\max(H - f_h) = (kh)^{-1/2}\max_j (N_j - kp_j). \tag{1.16}$$

The theorem will be proved in Sect. 2, and the bias term will be discussed in Sect. 3. The strategy is to derive the results from a general theorem in [4]. Section 4 describes the limiting behavior when some of the assumptions are violated: examples include uniform, exponential, and beta densities.

The theory developed here leads to rules for choosing h which seem to work well for real histograms; this will be explored in another paper. The methods of this paper can also be applied to frequency polygons, but we do not pursue this. Likewise, the method applies when f is defined on a half-line or a finite interval, provided no class interval crosses the boundary. We do not pursue this either. We focus on the maximum (positive) discrepancy. The method can be used to study the minimum (negative) discrepancy or the maximum absolute discrepancy. We do not pursue this either.

There has been some previous work related to Theorem (1.14). Smirnov (1944) considered the maximum normalized deviation

$$\sup_x |H(x) - f(x)|/\sqrt{H(x)}.$$

While Smirnov did not publish proofs of this theorems, he assumed the density was defined on a finite interval and bounded away from zero there. He used the growth condition (1.10). He found that the maximum normalized deviation has a limiting double-exponential distribution. This would follow from [4]. Similar theorems, with slightly different onditions, are proved by Tumanjan (1955), Woodroofe (1967) and Revesz (1972). The latter also considers the same maximum deviation we do. He works on a finite interval, assumes one bounded derivative for the density, and the growth condition $k \Big/ \left[\frac{1}{h}\log\frac{1}{h}\right] \to \infty$. He proves a strong law of large numbers for the maximum deviation. For example, when $h = 1/k^{1/3}$, Revesz shows that with probability one,

$$[k^{1/3}/\log k] \cdot \sup_x |H(x) - f(x)| \to 0$$

as k tends to infinity. Some of the authors just mentioned also give results for the maximum normalized error of kernel estimators for f. There is a recent paper on this topic by Bickel and Rosenblatt (1973). One novelty in the present paper is the treatment of the location of the maximum. Furthermore, as far as we know, this paper is the first to give the asymptotic distribution of $\sup_x H(x) - f(x)$.

Later, the following two calculus estimates will be needed.

(1.17) **Lemma.** *Suppose f is absolutely continuous on the interval $[a, b]$, with a.e. derivative g such that $|g| \leqq K < \infty$. Then*

$$\left| f - \frac{1}{b-a} \int_a^b f \right| \leqq \tfrac{1}{2} K(b-a),$$

and the inequality is sharp.

Proof. It is enough to do this without the absolute-value sign.

Case 1. The max of f is at b. Suppose without loss of generality that $f(a) = 0$. For $a \leqq x \leqq b$,

$$f(x) = \int_a^x g(u)\, du.$$

Integration by parts shows that

$$\begin{aligned}
\int_a^b f(x)\, dx &= \int_a^b (b-u)\, g(u)\, du \\
&= \int_a^b [(b-a) - (u-a)]\, g(u)\, du \\
&= (b-a) \int_a^b g(u)\, du - \int_a^b (u-a)\, g(u)\, du \\
&= (b-a) f(b) - \int_a^b (u-a)\, g(u)\, du.
\end{aligned}$$

Then

$$(b-a)f(b)-\int_a^b f(x)\,dx=\int_a^b (u-a)\,g(u)\,du$$
$$\leqq K\int_a^b (u-a)\,du=\tfrac{1}{2}K(b-a)^2.$$

Case 2. The max of f is at a. Use Case 1 on $f(a+b-x)$.

Case 3. The max of f is at ξ, with $a<\xi<b$.

Use Case 1 on (a,ξ) and Case 2 on (ξ,b):

$$f(\xi)\leqq\frac{1}{\xi-a}\int_a^\xi f+\tfrac{1}{2}K(b-a)$$

and

$$f(\xi)\leqq\frac{1}{b-\xi}\int_\xi^b f+\tfrac{1}{2}K(b-a).$$

Combining these two inequalities with the indicated weights,

$$f(\xi)=\frac{\xi-a}{b-a}f(\xi)+\frac{b-\xi}{b-a}f(\xi)\leqq\frac{1}{b-a}\int_a^b f+\tfrac{1}{2}K(b-a).$$

To see that the inequality is sharp, take $g=K$. □

(1.18) **Lemma.** *Suppose f is absolutely continuous on the interval $[a,b]$, as is f'. Let g be the a.e. derivative of f', and suppose $|g|\leqq K<\infty$. Let $c=\frac{1}{2}(a+b)$ and $\beta=|f'(c)|$. Then*

$$\left|\max f-\frac{1}{b-a}\int_a^b f-\tfrac{1}{2}\beta(b-a)\right|\leqq\tfrac{1}{6}K(b-a)^2$$

Proof. Without loss of generality, suppose $a=-1$ and $b=1$ so $c=0$. Likewise, take $f(0)=0$. By mapping x into $-x$, it is also permissible to assume that $\beta=f'(0)\geqq 0$.

Repeated integration by parts shows that

$$\bar{f}=\tfrac{1}{2}\int_{-1}^{1} f=\tfrac{1}{4}\left[\int_0^1 (1-u)^2\,g(u)\,du+\int_{-1}^{0}(1+u)^2\,g(u)\,du\right]$$

so $|\bar{f}|\leqq\frac{1}{6}K$. Now

$$f(x)-\bar{f}-\beta=\beta(x-1)+\tau(x)-\bar{f}$$

where

$$\tau(x)=\int_0^x (x-u)\,g(u)\,du$$

so $|\tau(x)|\leqq\frac{1}{2}K$. And

$$f(x)-\bar{f}-\beta\leqq\beta(x-1)+\tfrac{2}{3}K\leqq\tfrac{2}{3}K$$

while

$$f(1)-\bar{f}-\beta \geqq -\tfrac{2}{3}K. \quad \square$$

The constant 1/6 can be improved to 1/8; we omit the details. The 1/8 is sharp: take

$$f(x)=(\operatorname{sign} x)\cdot \tfrac{1}{2}Kx^2.$$

(1.19) **Notation.** $a_n \approx b_n$ means $a_n/b_n \to 1$ as $n\to\infty$, and $a_n \sim b_n$ means

$$0<\liminf_{n\to\infty} |a_n|/|b_n| \leqq \limsup_{n\to\infty} |a_n|/|b_n| < \infty.$$

In heuristic argument, we write $a_n \doteq b_n$ to mean "nearly equal."

2. The Proof of the Main Theorem

Fix $\delta>0$. It is convenient to split the cells into

(2.1) zone I, where $|hj|<\delta$,

(2.2) zone II, where $|hj|\geqq\delta$.

Only zone I contributes to the maximum: inside that zone, [4] can be used. To make contact with [4], imagine that k and h are functions of a hidden integer variable n tending to infinity; but continue to index by h, rather than n. Now

$$(N_j-kp_j)/(kh)^{1/2}=\alpha_{hj}Z_{hj}$$

where

$$\alpha_{hj}=(p_j/h)^{1/2}$$
$$Z_{hj}=(N_j-kp_j)/(kp_j)^{1/2}.$$

The object of study is

$$\max_j \alpha_{hj}Z_{hj}.$$

The first step is to estimate α_{hj}.

(2.3) **Lemma.** *Fix $\eta>0$. Then for some sufficiently small positive δ, and all sufficiently small positive h, if $|hj|<\delta$ then*

$$|\alpha_{hj}-\sqrt{f(x_0)}\,[1-\tfrac{1}{2}\rho^2h^2j^2]|<\eta h^2j^2+h.$$

Proof. Start from (1.3). Let $\tilde{\eta}$ be a small multiple of η, let $\delta>0$ be small, and set $\beta=-\alpha>0$. Then

$$f(x_0+x)<f(x_0)-\tfrac{1}{2}(1-\tilde{\eta})\,\beta x^2 \qquad \text{for } |x|<2\delta.$$

Abbreviate $\theta=(x_0-\lambda_0)/h$, so $0\leqq\theta\leqq 1$, and integrate over x with $x_0+h(j-\theta)\leqq x\leqq x_0+h(j+1-\theta)$:

$$p_j<f(x_0)h-\tfrac{1}{6}(1-\tilde{\eta})\,\beta h^3\,[(j+1-\theta)^3-(j-\theta)^3]$$

so

$$\begin{aligned} p_j/f(x_0)h &< 1-(1-\tilde{\eta})\rho^2h^2(j-\theta)^2-\tfrac{1}{3}(1-\tilde{\eta})\rho^2h^2(3j-3\theta+1)\\ &= 1-(1-\tilde{\eta})\rho^2h^2j^2+(1-\tilde{\eta})(2\theta-1)\rho^2h^2j-(1-\tilde{\eta})\rho^2h^2(\theta^2-\theta+\tfrac{1}{3})\\ &\leqq 1-(1-\tilde{\eta})\rho^2h^2j^2+\rho^2h^2|j|+\tfrac{1}{3}\rho^2h^2. \end{aligned}$$

The rest of the argument is omitted as routine: note that

$$h^2|j|=h|j|\cdot h<\delta h. \quad \square$$

The α_{hj} in [4, (1.1)] have already been introduced; now set $\beta_{hj}\equiv 0$. Take the scale parameter ε in [4, (1.2)] to be h itself, and set the centers c to 0. So $t_{hj}=hj$ in [4, (1.3)]. The interval I is $[-\delta,\delta]$. Turn now to [4, Sect. 4]; write t_0 for t_∞, α_0'' for α_∞'', β_0'' for β_∞'', index by the subscript h rather than n, and set

$$\begin{aligned} &t_h=t_0=0,\\ &\alpha_h=\sqrt{f(x_0)}, \quad \alpha_h'=0, \quad \alpha_0''=-\rho^2\sqrt{f(x_0)},\\ &\beta_h=\beta_h'=\beta_0''=0. \end{aligned}$$

The present ρ^2 coincides with the ρ^2 of [4, (4.6)]. Conditions [4, (4.1-6)] are easy to check, using the present (2.3). Conditions [4, (1.16–23)] are also easy: [4, (1.19)] is the present assumption (1.10). Theorem [4, (4.7)] now establishes the present (1.14), provided j is confined to zone I, i.e., $|hj|<\delta$.

What remains is to show that the j's in zone II do not contribute to the maximum. This is somewhat tedious. The next lemma does the job, plus a little bit more that will be useful later.

(2.4) **Lemma.** *Assume* (1.10). *Let* $y>1$ *and* $0<a<\infty$. *Let*

$$m=y\left[akh\cdot 2\log\frac{1}{h}\right]^{1/2}.$$

Then

$$\sum_j P\{N_j>kp_j+m\}\to 0,$$

where the sum extends over j's such that $p_j\leqq ah$.

Note. Here, $\{N_j\}$ can be any multinomial variables, with k trials and underlying cell probabilities $\{p_j\}$, and $h>0$ arbitrary. Let J be a set of indices, with $p_j\leqq ah$ for all $j\in J$. Let $y>1$. Then with probability tending to one as $h\to 0$,

$$\max_{j\in J}(N_j-kp_j)\leqq y\left[akh\cdot 2\log\frac{1}{h}\right]^{1/2}. \tag{2.5}$$

This is almost immediate from (2.4).

Proof. It is convenient to handle four types of j's separately.

Zone A, where $bh<p_j\leqq ah$ and $kp_j>Nm$.

Zone B, where $bh<p_j\leqq ah$ and $kp_j\leqq Nm$.

Zone C, where $p_j \leqq bh$ and $kp_j > \beta m$.
Zone D, where $p_j \leqq bh$ and $kp_j \leqq \beta m$.

The parameters b, β and N defining these zones must now be chosen. Begin by choosing $N < \infty$ but large, so that

$$\frac{y^2}{1+1/N} > 1. \tag{2.6}$$

Next, choose β. Clearly,

$$G(x) = \left(\frac{x}{1+x}\right)^{1+x} \approx x \quad \text{as } x \to 0. \tag{2.7}$$

Choose β to satisfy

$$0 < \beta < 1/4 \quad \text{and} \quad eG(x) < 3x \quad \text{for } 0 < x < \beta. \tag{2.8}$$

Finally, choose $b > 0$ so small that

$$\frac{y^2}{1+1/\beta} \cdot \frac{a}{b} > 1. \tag{2.9}$$

The argument for zones A–B–C involves a variant of Bernstein's inequality [3, Theorem 4b]:

$$P\{N_j > kp_j + m\} \leqq \exp\left[-\frac{1}{2}\frac{m^2}{kp_j + m}\right]. \tag{2.10}$$

Zone A. In (2.10), replace the m in the denominator on the right by its upper limit kp_j/N, and then p_j by its upper limit ah:

$$\begin{aligned} P\{N_j > kp_j + m\} &\leqq \exp\left[-\frac{1}{1+1/N}\frac{1}{2}\frac{m^2}{kp_j}\right] \\ &= \exp\left[-\frac{1}{1+1/N}\frac{1}{2}\frac{m^2}{akh}\right] \\ &\leqq \exp\left[-A \log\frac{1}{h}\right] \\ &= h^A \end{aligned}$$

where

$$A = \frac{y^2}{1+1/N} > 1$$

by (2.6). Since $\sum_j p_j = 1$, there are at most $1/bh$ terms in zone A. So

$$\sum_j P\{N_j > kp_j + m\} = O(h^{A-1}) = o(1),$$

the sum being extended over j's in zone A.

Zone B. In (2.10), replace the kp_j in the denominator on the right by its upper limit Nm:

$$P\{N_j > kp_j + m\} \leqq \exp\left[-\frac{1}{2}\frac{m}{N+1}\right].$$

But $m \sim \left(kh \log\frac{1}{h}\right)^{1/2}$ is much larger than $\log\frac{1}{h}$ by condition (1.10), and then the argument can be completed as in zone A.

Zone C. Let $i \geqq 1$. Let C_i be the set of j's in zone C with

$$bh/(i+1) < p_j \leqq bh/i.$$

So zone $C = \bigcup_{i=1}^{\infty} C_i$. Suppose $j \in C_i$. In (2.10), replace the m in the denominator by its upper limit kp_j/β. Then replace p_j by its upper limit bh/i:

$$P\{N_j > kp_j + m\} \leqq \exp\left[-Bi \log\frac{1}{h}\right]$$

where

$$B = \frac{y^2}{1+1/\beta} \cdot \frac{a}{b} > 1$$

by (2.9). There are at most $(i+1)/bh$ terms in C_i, so

$$\sum_{j \in C_i} \mathrm{P}\{N_j > kp_j + m\} \leqq \frac{i+1}{bh} h^{Bi}$$

and

$$\sum_{j \in \text{zone } C} P\{N_j > kp_j + m\} \leqq \frac{1}{bh} \sum_{i=1}^{\infty} (i+1) h^{Bi}$$

This is $O(h^{B-1}) = o(1)$ by an elementary argument.

Zone D. Here the bound [3] can be used again:

$$P\{N_j > kp_j + m\} \leqq [eG(kp_j/m)]^m, \tag{2.11}$$

where G was defined by (2.7). For j in zone D, condition (2.8) implies

$$\begin{aligned} P\{N_j > kp_j + m\} &\leqq (3kp_j/m)^m \\ &= \frac{3k}{m}(3kp_j/m)^{m-1} p_j \\ &\leqq \frac{3k}{m}(3\beta)^{m-1} p_j. \end{aligned}$$

But $\beta < 1/4$ by (2.8), so

$$\sum_{j \in \text{zone } D} P\{N_j > kp_j + m\} \leqq \frac{3k}{m}\left(\frac{3}{4}\right)^{m-1}. \tag{2.12}$$

To complete the proof, it is enough to show that the right side of (2.12) tends to 0. Taking logs, it is enough to prove

$$\log k - \log m - (\log \tfrac{4}{3}) m \to -\infty. \tag{2.13}$$

Now $k = \phi \frac{1}{h}\left(\log \frac{1}{h}\right)^3$ where $\phi \to \infty$ by (1.10), so

$$\log k = \log \phi + \log \frac{1}{h} + 3 \log\log \frac{1}{h}. \tag{2.14}$$

But $m \sim \left(kh \log \frac{1}{h}\right)^{1/2} \sim \phi^{1/2}\left(\log \frac{1}{h}\right)^2$ dominates the right side of (2.13). □

3. The Bias Term

The reults of this section can be summarized as follows. If $h \ll k^{-1/3}$, then bias is negligible and $\sup(H-f)$ behaves like $\sup(H-f_h)$, as determined in Theorem (1.14). See Corollary (3.20).

If $h \gg k^{-1/3}(\log k)^{1/3}$, then bias dominates and $\sup(H-f)$ behaves like $\sup(f_h - f)$. Proposition (3.23) shows that, suitably normalized, $\sup(H-f)$ tends to $\max|f'|$ in probability. Theorem (3.47) proves that $\sup(H-f)$ is taken on at unique location, in a neighborhood of the location of $\max|f'|$. The joint limiting distribution of the location and size of the maximum deviation is determined.

If h is between $k^{-1/3}$ and $k^{-1/3}(\log k)^{1/3}$, then bias and sampling error both contribute to $\sup(H-f)$. Suppose h is of order $k^{-1/3}(\log k)^{1/3}$. Theorem (3.14) determines the joint limiting distribution of the location and size of the maximum deviation. In particular, the maximum deviation is taken on at a unique location, in a neighborhood of the location of $\max(\sqrt{f} + \gamma|f'|)$, for γ a suitable constant.

The idea is to use [4] again, and some effort is needed to bring the present problem into that form. First, some heuristics. On cell j,

$$kh(H-f) = (N_j - kp_j) + kh(f_h - f) \tag{3.1}$$

so

$$(kh)^{1/2} \max(H-f) = \max_j \{\alpha_{hj} Z_{hj} + \gamma_{hj}\} \tag{3.2}$$

where

$$\begin{aligned} Z_{hj} &= (N_j - kp_j)/(kp_j)^{1/2} \\ \alpha_{hj} &= (p_j/h)^{1/2} \\ \gamma_{hj} &= (kh)^{1/2} \max_{\text{cell } j} (f_h - f). \end{aligned} \tag{3.3}$$

This is not quite in the form [4, (1.1)]. But letting

$$\beta_{hj} = \gamma_{hj}\left(2 \log \frac{1}{h}\right)^{-1/2} \tag{3.4}$$

gives

$$(3.5) \qquad (kh)^{1/2} \max (H-f) = \max_j \left\{\alpha_{hj} Z_{hj} + \beta_{hj} \left(2 \log \frac{1}{h}\right)^{1/2}\right\}.$$

This is in the form [4, (1.1)], and the coefficients must now be estimated.

Let I be some long (but finite) closed interval. The scale factor ε is taken as h itself; the center c, as 0. Thus

$$(3.6) \qquad t_{hj} = hj.$$

Conditions (1.1-2-3-4) are in force. To avoid tedious difficulties, assume f is smooth:

(3.7) f has three continuous derivatives; f and f' vanish at infinity; f is positive everywhere.

Condition (3.7) is discussed in a remark after the proof of (3.14), and again Sect. 4. Under this smoothness condition, $\alpha_{hj} \doteq \alpha(hj)$ and $\beta_{hj} \doteq \gamma_h \beta(hj)$, where

$$(3.8) \qquad \begin{aligned} \alpha(t) &= f(x_0+t)^{1/2} \\ \beta(t) &= |f'(x_0+t)| \\ \gamma_h &= \tfrac{1}{2} k^{1/2} h^{3/2} \left(2 \log \frac{1}{h}\right)^{-1/2} \end{aligned}$$

Suppose

(3.9) γ_h converges to a finite positive limit γ as $h \to 0$. In particular,

$$k = O\left(h^{-3} \log \frac{1}{h}\right), \qquad \text{i.e., } h = O[k^{-1/3} (\log k)^{1/3}].$$

As suggested by [4], consider the functions $\alpha + \gamma\beta$ and $\alpha + \gamma_h \beta$. Suppose

(3.10) $\alpha + \gamma\beta$ has a unique global maximum, say at t_0. Require I to include t_0 as an interior point.

As will be argued,

$$(3.11) \qquad f'(x_0+t_0) \neq 0.$$

Indeed, suppose by way of contradiction that $f'(x_0+t_0)=0$. Then $t_0=0$, for otherwise

$$f(x_0+t_0)^{1/2} + \gamma |f'(x_0+t_0)| = f(x_0+t_0)^{1/2} < f(x_0)^{1/2} \leqq f(x_0) + \gamma |f'(x_0)|.$$

Expanding around x_0,

$$f(x_0+h) = f(x_0) + O(h^2)$$

and

$$f'(x_0+h) = hf''(x_0) + O(h^2)$$

so

$$f(x_0+h)^{1/2} + \gamma |f'(x_0+h)| = f(x_0)^{1/2} + \gamma |h| \, |f''(x_0)| + O(h^2).$$

But $\gamma>0$ by assumption, and $f''(x_0)\neq 0$ by (1.3), so x_0 cannot be the location of the global maximum of $f(x)^{1/2}+\gamma|f'(x)|$. This completes the proof of (3.11).

In particular, β is f' or $-f'$ with the same choice of sign over some neighborhood of t_0. Therefore, β has a continuous second derivative.

The further assumption is needed, that

(3.11) $$\alpha''(t_0)+\gamma\beta''(t_0)<0.$$

Set

(3.12) $$\rho^2=-[\alpha''(t_0)+\gamma\beta''(t_0)]/\alpha(t_0).$$

If $\gamma=0$, then $\rho^2=-f''(x_0)/f(x_0)$ as before.

In view of (3.7–11), it is not hard to see that

(3.13) $\alpha+\gamma_h\beta$ has a unique global maximum, say at t_h; and $t_h\to t_0$ as $h\to 0$. Indeed, $t_h-t_0=O(\gamma_h-\gamma)$.

The main result of this section can now be stated.

(3.14) **Theorem.** *Suppose* (1.1–13) *and* (3.7–13). *In particular,* $\alpha(t)=f(x_0+t)^{1/2}$ *and* $\beta(t)=|f'(x_0+t)|$. *Let* $h\to 0$. *Then with probability approaching one,* $M'_{kh}=\max(H-f)$ *is attained in a unique cell; call its index* L'_{kh}. *Furthermore, the chance that*

$$\rho\sqrt{2\log\frac{1}{h}}\,(hL'_{kh}-t_h)<y$$

and

$$M'_{kh}\cdot\sqrt{kh}<\alpha(t_h)\,w_h(x)+\gamma_h\beta(t_h)\sqrt{2\log\frac{1}{h}}$$

converges to

$$\Phi(y)\cdot\exp\left\{-\frac{1}{2\rho}e^{-x/2}\right\}.$$

Note. In particular, $\max(H-f)$ occurs in the vicinity of the maximum of $\sqrt{f(x)}+\gamma|f'(x)|$. As far as the scaling is concerned, $M'_{kh}\cdot\sqrt{kh}=M_{kh}/\sqrt{kh}$: see (1.16).

Proof. The first step is to estimate α_{hj}. As is easily verified from (3.7),

(3.15) $$\alpha_{hj}=\alpha(hj)^{1/2}+O(h)=\alpha(hj)^{1/2}+o\left(1/\log\frac{1}{h}\right)$$
as $h\to 0$, uniformly in j with $hj\in I$.

Likewise for β_{hj}. Indeed, by (1.18) and Taylor's theorem,

(3.16) $$\max_x\{f_h(x)-f(x)\colon \lambda_0+hj\leqq x\leqq\lambda_0+h(j+1)\}=\tfrac{1}{2}h|f'(x_0+hj)|+O(h^2)$$

as $h\to 0$, uniformly in j with $hj\in I$.

(3.17) If $k \ll h^{-5}\left(\log\frac{1}{h}\right)^{-1}$, as is implied by (3.9), then $\beta_{hj}=\gamma_h\beta(hj)+o\left(1/\log\frac{1}{h}\right)$ as $h\to 0$, uniformly in j with $hj\in I$.

Thus, conditions [4, (1.1–5)] are satisfied. The remaining conditions for [4, (1.24)] have all been assumed: [4, (1.11)] follows from the continuity of f'''. As a result, the conclusions of (3.14) apply provided $\max_x [H(x)-f(x)]$ is taken over any long but finite closed interval including x_0+t_0 as an interior point. In other words, the j in $\max_j \{\alpha_{hj}Z_{hj}+\gamma_{hj}\}$ is constrained so $hj\in I$, where I is any long but finite closed interval including t_0 as an interior point.

In particular, the max of $\alpha_{hj}Z_{hj}+\gamma_{hj}$ over such j's is of order

$$[\alpha(t_h)+\gamma_h\beta(t_h)]\left(2\log\frac{1}{h}\right)^{1/2}.$$

Note that

$$\alpha(t_h)+\gamma_h\beta(t_h)\geqq\alpha(t_0)+\gamma_h\beta(t_0)\geqq\alpha(t_0)>0,$$

because $\beta(t)=|f'(x_0+t)|\geqq 0$. It must now be shown that the remaining j's do not contribute to the max: namely, if I sufficiently long, then

$$\max_{hj\notin I}\{\alpha_{hj}Z_{hj}+\gamma_{hj}\}$$

is only a small multiple of $\alpha(t_0)\sqrt{2\log\frac{1}{h}}$.

Since f vanishes at infinity, $p_j\leqq ah$ where a is small for $hj\notin I$ long. By (2.4), with overwhelming probability,

$$N_j-kp_j<2\left[akh\cdot 2\log\frac{1}{h}\right]^{1/2}$$

for all j with $hj\notin I$. Refer back to Definition (3.3) of α_{hj} and Z_{hj}. With that same overwhelming probability,

$$\begin{aligned}\alpha_{hj}Z_{hj}&=(N_j-kp_j)/(kh)^{1/2}\\&<2a^{1/2}\left(2\log\frac{1}{h}\right)^{1/2}\end{aligned}$$

for $hj\notin I$.

This leaves the job of estimating

(3.18) $$\max_j\{\gamma_{hj}\colon hj\notin I\}$$

where γ_{hj} was defined in (3.3). By (1.17),

(3.19) $$\max_{\text{cell}\,j}(f_h-f)\leqq\tfrac{1}{2}h\max_{\text{cell}\,j}|f'|.$$

Since f' vanishes at infinity, (3.18) is only a small multiple of

$$\tfrac{1}{2}k^{1/2}h^{3/2}=\gamma_h\left(2\log\frac{1}{h}\right)^{1/2}. \quad \square$$

Remark. Condition (3.7) is hardly minimal. The existence and continuity of f'' and f''', as well as the positivity of f, are needed only in the vicinity of x_0+t_0, the unique (by assumption) location of the global maximum of $f(x_0+t)^{1/2}+\gamma|f'(x_0+t)|$. With these weaker conditions, one can still eliminate the j's with $0<\delta\leqq|hj|\leqq 1/\delta$, as candidates for the location L_{kh} of the max. This interval can be represented as a finite union of closed intervals J so short that

$$\max_J \alpha+\gamma\max_J \beta<\max(\alpha+\gamma\beta).$$

On J, control over

$$\max \alpha_{hj}Z_{hj}=\max(N_j-kp_j)/(kh)^{1/2}$$

is obtained by (2.4). Compare [4, (2.5) or (3.3)]. Likewise, the conditions that f and f' vanish at infinity can be weakened to

$$f(x_0+t_0)>\limsup_{|x|\to\infty} f(x)$$

$$|f'(x_0+t_0)|\geqq\limsup_{|x|\to\infty}|f'(x)|.$$

Control over γ_{hj} is obtained from (1.17); and control over $\alpha_{hj}Z_{hj}$, as before, from (2.4).

A Second Remark. As Richard Olshen points out, condition (3.10) is almost bound to fail for symmetric unimodal densities, like the normal or the Cauchy; then $x_0=0$ and two global maxima for $\alpha+\gamma\beta$ can be anticipated, at $\pm t_0$ say. Suppose the regularity conditions hold at both places. Under such circumstances, (3.14) describes $M^+=\max_{x>0}[H(x)-f(x)]$, which occurs near t_0. It also describes $M^-=\max_{x<0}[H(x)-f(x)]$, which occurs near $-t_0$. The two maxima, and their locations, are asymptotically independent. So

$$\max_{\text{all } x}[H(x)-f(x)]=M^+\vee M^-$$

is still double exponential. The location of the overall maximum, however, is no longer normal, for it is near t_0 with probability 1/2 and near $-t_0$ with probability 1/2. Given that the location is near t_0, its conditional distribution does become asymptotically normal, and likewise for $-t_0$.

A Third Remark. Apparently, the situation is different when $\gamma_h\to 0$. Condition (1.11) of [4] is violated; β is locally wedgeshaped, not parabolic. In effect, we are trying to maximize

$$\sqrt{f(x_0)}\,[1-\tfrac{1}{2}a^2j^2h^2]\,Z_{hj}+\sqrt{f(x_0)}\,\gamma_h b|jh|\bigg/\sqrt{2\log\frac{1}{h}}$$

where $a^2 = 1/2|f''(x_0)|/f(x_0)$ and $b = |f''(x_0)|/\sqrt{f(x_0)}$: compare (1.13). There are two distinct places where the maximum can occur, one for positive j near $c\gamma_h/h$, and one for negative j near $-c\gamma_h/h$, where $c = b/a^2 = 2\sqrt{f(x_0)}$. The respective maxima in these two places are asymptotically independent, each being more or less as described in (3.14), although ρ must be computed from a and b. Also, the location of the maximum over positive j is in effect a truncated normal. So the global maximum is still double-exponential, but its location is a mixture of truncated normals. If $\gamma_h = o(1/\sqrt{2\log 1/h})$, the truncation point is at 0, so the mixture is itself normal: see (3.20). If γ_h is of order $1/\sqrt{2\log 1/h}$, the truncation point stabilizes away from 0. For larger γ_h, the truncation point drifts off to $\pm\infty$, and we get a mixture of two normals. Note that $t_h = O(\gamma_h)$. However, when γ_h is, e.g., of order $1/(2\log 1/h)^{1/4}$, the coefficients $\alpha(t_h)$ and $\beta(t_h)$ differ sufficiently from $\alpha(0)$ and $\beta(0)$: for

$$\alpha(t_h)/\alpha(0) - 1 \sim \gamma_h^2,$$
$$\beta(t_h)/\beta(0) - 1 \sim \gamma_h^2.$$

The next corollary shows that if h is small, sampling error dominates.

(3.20) **Corollary.** *Suppose* (1.1–13) *and* (3.7–13). *Suppose that* $k = o(h^{-3})$, *so* $\gamma_h \to 0$. *Then the asymptotic behavior of the location and size of* $\max(H-f)$ *coincides with that for* $(H-f_h)$.

Proof. This is easiest to argue from [4, (4.1)] with $\beta_{nj} = \beta_{hj} = \gamma_h \beta(t_{hj}) + o(1/\log 1/h)$ by (3.17). However, $t_h = O(\gamma_h)$ so $\gamma_h \beta(t_h) = O(\gamma_h^2) = o(1/\log 1/h)$. □

Now consider the case where

(3.21) $$kh^3/\log\frac{1}{h} \to \infty,$$

so bias dominates. In essence,

$$(kh)^{1/2} \max(H-f) \doteq \max_j \left\{\alpha(hj) Z_{hj} + \gamma_h \beta(hj) \sqrt{2\log\frac{1}{h}}\right\}$$

and $\gamma_h \to \infty$. Thus, the term $\gamma_h \beta(hj)\sqrt{2\log\frac{1}{h}}$ dominates, and what counts is the behavior of $\beta(t) = |f'(x_0 + t)|$ at its maximum. Assume

(3.22) $|f'|$ has a unique global maximum, say at $x_0 + t_1$.

Clearly, f' cannot vanish at $x_0 + t_1$. (If the domain of f is a finite interval, an extra assumption is called for here.)

(3.23) **Proposition.** *Suppose* (1.1–13) *and* (3.7–8) *and* (3.21–22). *Then*

$$(kh)^{1/2}\gamma_h^{-1}\left(2\log\frac{1}{h}\right)^{-1/2} \max(H-f) = 2 \cdot h^{-1} \max(H-f) \to \beta(t_1) = \max|f'|$$

in probability. Furthermore, for any δ positive, with probability approaching one, the max is taken on only for j's with $|hj-t_1|<\delta$.

Proof. Refer back to (3.5). In view of (1.14),

$$\max_j \{\alpha_{hj}Z_{hj}\} = \max_j \{(N_j - kp_j)/(kh)^{1/2}\}$$
$$\sim [f(x_0)]^{1/2}\left[2\log\frac{1}{h}\right]^{1/2}.$$

This is negligible by comparison with $\gamma_h\left[2\log\frac{1}{h}\right]^{1/2}$. Likewise, if $|hj-t_1|\geqq\delta$, then (3.19) entails

$$\max_j \beta_{hj}\left[2\log\frac{1}{h}\right]^{1/2} \leqq \theta\gamma_h\left[2\log\frac{1}{h}\right]^{1/2}$$

where

$$\theta = \max_t \{|f'(x_0+t)|: |t-t_1|\geqq\delta\} < |f'(x_0+t_1)|. \quad \square$$

Proposition (3.23) is a "weak law" for the maximum. To get a distributional result, assume

(3.24) $$f'''(x_0+t_1)\neq 0$$

and

$$h^{-3}\log\frac{1}{h} \ll k \ll h^{-5}\log\frac{1}{h}.$$

If k is of order $h^{-5}\log\frac{1}{h}$ or more, the behavior changes: finer estimates than (3.17) are needed for β_{hj}. If k is of order h^{-7} or more, the location L'_{kh} of the max changes its character, becoming discrete. We do not pursue these issues; for a related discussion, see [8, Sect. 3].

The heuristics will now be indicated. Proposition (3.23) shows that

$$\max(\alpha_{hj}Z_{hj}+\gamma_{hj})$$

is essentially

$$\gamma_h\beta(t_1)\left(2\log\frac{1}{h}\right)^{1/2},$$

which is blowing up as $h\to 0$. Subtract this lead term off, getting

(3.25) $$(kh)^{1/2}(H-f)-\gamma_h\beta(t_1)\left(2\log\frac{1}{h}\right)^{1/2} = \alpha_{hj}Z_{hj}+\tilde{\gamma}_{hj}$$

where

(3.26) $$\tilde{\gamma}_{hj} = \gamma_{hj}-\gamma_h\beta(t_1)\left(2\log\frac{1}{h}\right)^{1/2}$$
$$= \tfrac{1}{2}k^{1/2}h^{3/2}\left[2\frac{1}{h}\max_{\text{cell}\,j}(f_h-f)-\beta(t_1)\right]$$

from (3.3) and (3.8). Now from (3.16)

$$2\frac{1}{h}\max_{\text{cell } j}(f_h - f) \doteq \beta(hj).$$

For hj close to t_1, where β is maximum,

$$\beta(hj) - \beta(t_1) \doteq \tfrac{1}{2}\beta''(t_1)h^2(j - h^{-1}t_1)^2.$$

Parenthetically, this last can be made rigorous if $k \gg h^{-3}\left(\log\frac{1}{h}\right)^5$, but is too aggressive for smaller k's. To sum up, the right side of (3.25) is nearly

$$\alpha_{hj}Z_{hj} + \tfrac{1}{4}k^{1/2}h^{7/2}\beta''(t_1)(j - h^{-1}t_1)^2.$$

Now it will be possible to use [4] again, but with a new scale factor ε_h, chosen to satisfy

$$\varepsilon_h^2\left(2\log\frac{1}{\varepsilon_h}\right)^{1/2} \doteq k^{1/2}h^{7/2}.$$

To make this rigorous, set

(3.27) $$m = (\tfrac{1}{2}k^{1/2}h^{7/2})^{-1/2}.$$

(The factor 1/2 here is almost accidental.) So $m \to \infty$ as $h \to 0$; set

(3.28) $$\varepsilon_h = m^{-1}(2\log m)^{-1/4}.$$

Now (3.25) can be studied, in the guise

$$\max_j\left\{\alpha_{hj}Z_{hj} + \tilde{\beta}_{hj}\sqrt{2\log\frac{1}{\varepsilon_h}}\right\},$$

where

(3.29) $$\tilde{\beta}_{hj} = \tilde{\gamma}_{hj}\Big/\sqrt{2\log\frac{1}{\varepsilon_h}}.$$

This is in the form of [4, (1.1)]. The center called for in [4, (1.3)] is defined as follows:

(3.30) $$c_h = t_1/h.$$

Write

(3.31) $$\theta_{hj} = \varepsilon_h(j - c_h)$$

to avoid confusion with the t_{hj} previously used. Clearly,

(3.32) $$t_{hj} = hj = \delta_h\theta_{hj} + t_1$$

where

(3.33) $$\delta_h = h\varepsilon_h^{-1} \to 0.$$

This latter is easily verified, because for small h, using the growth condition (3.24),

(3.34) $$\tfrac{1}{2}\log\frac{1}{h} - \tfrac{1}{4}\log\log\frac{1}{h} \leqq \log m \leqq \log\frac{1}{h}.$$

By (3.28),

(3.35) $$\log\frac{1}{\varepsilon_h} \approx \log m.$$

Then

(3.36) $$\log\frac{1}{\varepsilon_h} \sim \log\frac{1}{h}.$$

A useful identity:

(3.37) $$\gamma_h\left(2\log\frac{1}{h}\right)^{1/2}\delta_h^2 = (2\log m)^{1/2}.$$

For the function α_n of [4, (1.4)], take

(3.38) $$\alpha_h(\theta) = \alpha(t_1 + \delta_h\theta) = f(x_0 + t_1 + \delta_h\theta)^{1/2} \qquad \text{by (3.8)}$$

so that

$$\alpha_h(\theta_{hj}) = \alpha(hj).$$

It is convenient to prove something a bit stronger than [4, (1.4)]:

(3.39) $$\alpha_{hj} = \alpha_h(\theta_{hj}) + o\left(1/\log\frac{1}{\varepsilon_h}\right)$$

uniformly in j with t_{hj} confined to a compact interval.

This is immediate from (3.15) and (3.36).

For the functions β_n of [4, (1.5)], take

(3.40) $$\beta_h(\theta) = \left(\log m/\log\frac{1}{\varepsilon_h}\right)^{1/2}\delta_h^{-2}\cdot[\beta(t_1+\delta_h\theta) - \beta(t_1)]$$

where $\beta(t) = |f'(x_0+t)|$. By (3.32),

(3.41 $$\beta_h(\theta_{hj}) = \left(\log m/\log\frac{1}{\varepsilon_h}\right)^{1/2}\delta_h^{-2}[\beta(hj) - \beta(t_1)].$$

For [4, (1.5)], it is claimed that

(3.42) $\tilde{\beta}_{hj} = \beta_h(\theta_{hj}) + o\left(1/\log\frac{1}{\varepsilon_h}\right)$ uniformly in j with t_{hj} confined to a compact interval.

This follows from (3.17) by tedious algebra, using the growth condition $k \ll h^{-5}\left(\log\frac{1}{h}\right)^{-1}$. Indeed,

$$\tilde{\beta}_{hj} = \left(2\log\frac{1}{\varepsilon_h}\right)^{-1/2} \tilde{\gamma}_{hj} \qquad \text{by (3.29)}$$

$$= \left(2\log\frac{1}{\varepsilon_h}\right)^{-1/2} \left[\gamma_{hj} - \gamma_h \beta(t_1)\left(2\log\frac{1}{h}\right)^{1/2}\right] \qquad \text{by (3.26)}$$

$$= \left(\log\frac{1}{h} \Big/ \log\frac{1}{\varepsilon_h}\right)^{1/2} [\beta_{hj} - \gamma_h \beta(t_1)] \qquad \text{by (3.4)}$$

$$= \gamma_h \left(\log\frac{1}{h} \Big/ \log\frac{1}{\varepsilon_h}\right)^{1/2} [\beta(hj) - \beta(t_1)] + o\left(1/\log\frac{1}{\varepsilon_h}\right) \qquad \text{by (3.17) and (3.36)}$$

$$= \gamma_h \left(2\log\frac{1}{h}\right) \delta_h^2 \cdot \left(2\log\frac{1}{\varepsilon_h}\right)^{-1/2} \delta_h^{-2}[\beta(hj) - \beta(t_1)] + o\left(1/\log\frac{1}{\varepsilon_h}\right)$$

$$= \left(\log m/\log\frac{1}{\varepsilon_h}\right)^{1/2} \delta_h^{-2}[\beta(hj) - \beta(t_1)] + o\left(1/\log\frac{1}{\varepsilon_h}\right) \qquad \text{by (3.37)}$$

$$= \beta_h(\theta_{hj}) + o\left(1/\log\frac{1}{\varepsilon_h}\right) \qquad \text{by (3.41)}$$

This completes the argument for (3.42).

Condition [4, (1.6)] is clear: for I, take any long (but finite) closed interval, with 0 as an interior point. For [4, (1.7)], let

$$\alpha_0(\theta) = \alpha(t_1) = f(x_0 + t_1)^{1/2} \tag{3.43}$$

for all θ. Then $\alpha_h \to \alpha_0$ as $h \to 0$ because α is continuous and $\delta_h \to 0$: see (3.38) and (3.33). Let

$$\beta_0(\theta) = \tfrac{1}{2}\beta''(t_1)\,\theta^2. \tag{3.44}$$

Then

$$\beta_h(\theta) \to \beta_0(\theta) \quad \text{uniformly over } \theta \in I. \tag{3.45}$$

Indeed, the normalizing factor $\left(\log m/\log\frac{1}{\varepsilon_h}\right)^{1/2}$ tends to 1 by (3.35). Next, recall that t_1 is the location of the maximum of $\beta(t) = |f'(x_0+t)|$, so $\beta(t_1) > 0$. Suppose, e.g., that $f'(x_0+t_1) > 0$. Then $\beta(t) = f'(x_0+t)$ in some neighborhood of t_1, and

$$\beta'(t) = f''(x_0+t), \qquad \beta''(t) = f'''(x_0+t).$$

In particular, $\beta'(t_1) = 0$ and $\beta''(t_1) < 0$: see (3.24). Now expand:

$$\delta_h^{-2}[\beta(t_1 + \delta_h\theta) - \beta(t_1)] = \tfrac{1}{2}\beta''(\xi)\theta^2$$

where $\xi \to t_1$ as $h \to 0$. The assumed continuity of f''' at t_1 completes the argument for (3.45).

The remaining conditions for [4, (1.24)] are quite easy to verify; [4, (1.11)] follows from the continuity of f''', as in the argument for (3.45). Let θ_h be the (unique) location of the global maximum of $\alpha_h+\beta_h$. Then $\theta_h\to 0$; indeed, $\theta_h = O(\delta_h)$. Let

$$\begin{aligned}\tilde\rho^2 &= -\beta''(t_1)/\alpha(t_1)\\ &= |f'''(x_0+t_1)|/f(x_0+t_1)^{1/2}\end{aligned} \tag{3.46}$$

(3.47) **Theorem.** *Suppose* (1.1–13), (3.7–8), (3.21–22), *and* (3.24). *With probability approaching one,* $M'_{kh}=\max(H-f)$ *is taken on in a unique cell, of index* L_{kh}. *The chance that*

$$\tilde\rho\sqrt{2\log\frac{1}{\varepsilon_h}}\,[\delta_h^{-1}(hL'_{kh}-t_1)-\theta_h]<y$$

and

$$\sqrt{kh}\,M'_{kh}-\gamma_h\beta(t_1)\sqrt{2\log\frac{1}{h}}<\alpha_h(\theta_h)\,w_h(x)+\beta_h(\theta_h)\sqrt{2\log\frac{1}{\varepsilon_h}}$$

converges to

$$\Phi(y)\cdot\exp\left\{-\frac{1}{2\tilde\rho}e^{-x/2}\right\}.$$

Proof. If j is confined so $\theta_{hj}\in I$, the result follows from [4, (1.24)]. It is only necessary to show that the remaining j's do not matter. If $|t_{hj}-t_1|\geqq\delta>0$, this is immediate from (3.23). Next, consider the j's such that

$$A\leqq|\theta_{hj}|\quad\text{but}\quad|t_{hj}-t_1|\leqq\delta. \tag{3.48}$$

Refer to (3.25) and (3.29): it must be shown that the max over j's satisfying (3.48) of

$$\alpha_{hj}Z_{hj}+\tilde\beta_{hj}\sqrt{2\log\frac{1}{\varepsilon_h}} \tag{3.49}$$

is of smaller order than

$$[\alpha_h(\theta_h)+\beta_h(\theta_h)]\sqrt{2\log\frac{1}{\varepsilon_h}}.$$

Now

$$\alpha_h(\theta_h)+\beta_h(\theta_h)\geqq\alpha_h(0)+\beta_h(0)=\alpha(t_1)>0,$$

since $\alpha_h+\beta_h$ is maximized at θ_h. Refer to (3.38) and (3.40) for the definitions of α_h and β_h.

Apply [4, (3.1)] to Z_{hj}, but on the h-scale: i.e., put h for the ε in [4, (3.1)]. The conclusion is that with overwhelming probability,

$$\max_j Z_{hj}\leqq 2\left[2\log\frac{1}{h}\right]^{1/2}<4\left[2\log\frac{1}{\varepsilon_h}\right]^{1/2}$$

for h small, by (3.36). Hence

$$\max_j \left\{\alpha_{hj} Z_{hj} + \tilde{\beta}_{hj}\left[2\log\frac{1}{\varepsilon_h}\right]^{1/2}\right\}$$
$$\leqq \left[2\log\frac{1}{\varepsilon_h}\right]^{1/2} [4\max_j \alpha_{hj} + \max_j \tilde{\beta}_{hj}]$$
$$\leqq \left[2\log\frac{1}{\varepsilon_h}\right]^{1/2} \left[4\max_j \alpha_h(\theta_{hj}) + \max_j \beta_h(\theta_{hj}) + o\left(1/\log\frac{1}{\varepsilon_h}\right)\right]$$

by (3.39) and (3.42), where the max is taken over all j's satisfying (3.48).

From the definition (3.38) of α_h it is clear that $\alpha_h(\theta) \leqq f(x_0)^{1/2}$. Refer now to the definition (3.40) of $\beta_h(\theta)$. Recall that t_1 is the location of the global maximum of $\beta(t) = |f'(x+t)|$, and β is locally quadratic at t_1. If δ is small, and $0 < \delta' < \delta$, then

$$\max_u \{\beta(t_1+u)\colon \delta' \leqq u \leqq \delta\} = \beta(t_1+\delta'),$$
$$\max_u \{\beta(t_1-u)\colon \delta' \leqq u \leqq \delta\} = \beta(t_1-\delta').$$

So, confining j to satisfy (3.48), and writing $\vee$ for max,

$$\max_j \beta_h(\theta_{hj}) \leqq \beta_h(A) \vee \beta_h(-A).$$

And so

$$\limsup_{h\to 0} \max_j \beta_h(\theta_{hj}) \leqq \tfrac{1}{2}\beta''(t_1)A^2.$$

Recall $\beta''(t_1) < 0$. Now choose A so large that

$$\lambda = 4f(x_0)^{1/2} + \tfrac{1}{2}\beta''(t_1)A^2 < 0.$$

With overwhelming probability, the max of the variables in (3.49), over the j's satisfying (3.48), is smaller than

$$[2\log\frac{1}{\varepsilon_h}\Bigg]^{1/2} \left[\lambda + o\left(1/\log\frac{1}{\varepsilon_h}\right)\right]$$

where $\lambda < 0$. Such j's do not matter. □

Remark. If $k \gg h^{-3}\left(\log\frac{1}{h}\right)^3$, the scaling can be simplified: the θ_h can be set to 0, and $\alpha_h(\theta_h)$ to $\alpha(t_1)$, and $\beta_h(\theta_h)$ to zero.

4. Examples

Our object in this section is to indicate what happens when the regularity conditions are violated. Some arguments are only sketched, for the focus is on qualitative features. Define M_{kh} and L_{kh} as in (1.14).

(4.1) *Example.* The density f is uniform on $[a,b]$. This violates the condition (1.2) that f have a unique maximum. In this case, L_{kh} is uniform, rather than asymptotically normal; L_{kh} and M_{kh} are independent; M_{kh} is asymptotically double-exponential, but with a different scaling than in (1.14). To state a more general result, do not assume f uniform. Instead, suppose $\max f=\mu$, attained on a whole interval of length λ. If x is bounded away from this interval, suppose $\sup_x f(x)<\mu$. Let

$$w_h^*(x)=\left[2\log\frac{1}{h}-\log\log\frac{1}{h}+x\right]^{1/2}. \tag{4.2}$$

Then the chance that

$$M_{kh}/\sqrt{\mu kh}<w_h^*(x)$$

converges to

$$\exp[-ce^{-x/2}]$$

where $c=\frac{1}{2}\lambda\pi^{-1/2}$. The relevant growth condition is (1.10).

Sketch of Proof. In the critical interval, $p_j=\mu h$, so the chance that

$$N_j-kp_j>(\mu kh)^{1/2}w_h^*(x)$$

is nearly

$$(2\pi)^{-1/2}w_h^*(x)^{-1}\exp\{-\tfrac{1}{2}w_h^*(x)^2\}\doteq\tfrac{1}{2}\pi^{-1/2}h\exp(-x/2).$$

See [5, (3.17)]. The part of the line bounded away from the critical interval can be handled by (2.4). We omit further details. □

(4.3) *Example.* Consider the beta density $f(x)=\frac{1}{2}x^{-1/2}$ for $0\leqq x\leqq 1$. Here, f has a unique maximum at 0, but $f(0)=\infty$. Now L_{kh} and M_{kh} are no longer independent; L_{kh} converges in law, without any rescaling, to a probability on the nonnegative integers; and $k^{-1/2}h^{-1/4}M_{kh}$ converges in law to something which is not double-exponential. The scaling $k^{-1/2}h^{-1/4}$ of M_{kh} here is quite different from the $k^{-1/2}h^{-1/2}$ in (1.14).

To describe the limit in more detail, let

$$c_j=[(j+1)^{1/2}-j^{1/2}]^{-1/2}\approx 2^{1/2}j^{1/4}. \tag{4.4}$$

Let L^* and M^* be the index and size respectively of the maximum of W_1/c_1, $W_2/c_2,\ldots$, the W's being independent $N(0,1)$ variables. Then (L_{kh},M_{kh}) converges in law to (L^*,M^*). The relevant growth condition is that

$$\lambda=k^{1/2}h^{1/4}\to\infty. \tag{4.5}$$

Sketch of Proof. Clearly,

$$p_j=h^{1/2}/c_j^2, \tag{4.6}$$

where c_j was defined in (4.4). For any fixed J, we claim that the joint distribution of

$$(N_j-kp_j)/\lambda:\ 0\leqq j\leqq J$$

converges to the joint distribution of

$$W_j/c_j\colon 0 \leqq j \leqq J.$$

The scaling-factor λ is defined in (4.5). For $J=0$, this follows from the central limit theorem with a uniform error bound. We now do $J=1$. Given $N_0=n_0$, conditionally N_1 is binomial; the number of trials is

$$k'=k-n_0=k(1-p_0)+\zeta_0$$

where ζ_0/λ is (unconditionally) almost $N(0,1)$. The success probability is $p'=p_1/1-p_0$. Thus, N_1 has conditional mean $k'p'$ and variance $kp'(1-p')$. But $kp_1=\lambda^2/c_1^2$ by (4.5–6), and

$$k'p'-kp_1=p_1\zeta_0/(1-p_0)$$

is of order $\lambda h^{1/2}$ which can be ignored. Thus, with high probability, given N_0 the conditional law of $(N_1-kp_1)/\lambda$ will be close to the law of W_1/c_1. General J is done by induction. For a more efficient argument, see [8, (3.17)].

The final step is to show that large j's don't count. We found this a bit difficult, and indicate the main steps. Fix $x>0$ but small. Set $m=\lambda x$, where λ was defined in (4.5). Let

$$q_j=P\{N_j>kp_j+m\}.$$

Then $\sum_{j=J}^{\infty} q_j$ is an upper bound to the chance that $\lambda^{-1}(N_j-kp_j)>x$ for some $j \geqq J$. We estimate this sum in two parts, defined by a parameter A with $A>9/x^2$.

If $J \leqq j \leqq A\lambda^2$, we use (2.10), replacing p_j by its upper bound $\frac{1}{2}h^{1/2}j^{-1/2}$, valid for $j \geqq 1$:

$$q_j \leqq \exp\{-x^2 j^{1/2}/[1+2\lambda^{-1}j^{1/2}x]\}.$$

Next, replace the j in the denominator by its upper bound $A\lambda^2$ to get

$$q_j \leqq \exp\{-x^2 j^{1/2}/[1+2A^{1/2}x]\}.$$

So $\sum_j \{q_j\colon J \leqq j \leqq A\lambda^2\}$ is small for J large: how large depends, of course, on x.

If $A\lambda^2 \leqq j$, we use (2.11). Again, replace p_j by $\frac{1}{2}h^{1/2}j^{-1/2}$, and note that $eG(u) \leqq 3u$ for small u:

$$q_j \leqq (\tfrac{9}{4}\lambda^2/x^2 j)^{\lambda x/2}$$

Now $\sum_j \{q_j\colon A\lambda^2 \leqq j\}$ can be bounded above by an integral, and it is very small. □

We next take up the role of condition (1.3), that f be locally quadratic at its unique global maximum x_0. It is this assumption which makes the location L_{kh}

asymptotically normal. To make this clear, suppose (1.1–10), except (1.3) is replaced by the condition that as $x \to 0$,

(4.7)
$$f(x_0+x)=f(x_0)-A|x|^a+o(|x|^a)$$
A and a are positive.

Then L_{kh} and M_{kh} are still asymptotically independent. And M_{kh} is asymptotically double-exponential, although the scale w_h of (1.11) must be replaced by

(4.8)
$$w_h^a(x)=\left[2\log\frac{1}{h}-\left(\frac{2}{a}+1\right)\log\log\frac{1}{h}+x\right]^{1/2}.$$

It is perhaps worth noting that as $a\to\infty$, condition (4.7) forces f to get flatter and flatter; while the scale w_h^a converges to the scale w_h^* for the uniform: see (4.2). Likewise, L_{kh} should be scaled not proportionally to $h\left(\log\frac{1}{h}\right)^{1/2}$, as in (1.14), but to $h\left(\log\frac{1}{h}\right)^{1/a}$. Its asymptotic density is then proportional to $\exp\{-|u|^a\}$. A familiar example covered by (4.7) is the density $f(x)=\frac{1}{2}\exp(-|x|)$, with $a=1$.

Suppose (4.7) is weakened further, breaking the symmetry: as $x\to 0^+$,

(4.9)
$$f(x_0+x)=f(x_0)-Ax^a+o(x^a)$$
$$f(x_0-x)=f(x_0)-Bx^b+o(x^b)$$
A, B, a, b positive.

If e.g. $a<b$, then the maximum occurs just to the right of x_0, i.e., L_{kh} is asymptotically positive. Again, L_{kh} and M_{kh} are asymptotically independent. If $a=b$ but e.g. $A<B$, then L_{kh} and M_{kh} are no longer independent: if L_{kh} is positive, then M_{kh} is bigger.

To handle this sort of situation, let

(4.10)
$$M_{kh}^+=\max_j\{N_j-kp_j: j\geqq 0\}.$$

With overwhelming probability, this is attained at a unique index L_{kh}^+. Likewise for M_{kh}^- and L_{kh}^-. Let

(4.11)
$$K_a=\int_0^\infty \exp(-u^a)\,du$$
$$\Phi_a(y)=K_a^{-1}\int_0^y \exp(-u^a)\,du$$
$$c_a=\tfrac{1}{2}\pi^{-1/2}K_aA^{-1/a}f(x_0)^{1/a}.$$

(4.12) **Theorem.** *Suppose* (1.1–10), *except that* (1.3) *is replaced by* (4.9). *Then the four variables* M_{kh}^+, M_{kh}^+, L_{kh}^+, L_{kh}^- *are asymptotically independent. Furthermore,*

$$P\left\{[A/f(x_0)]^{1/a}\left(\log\frac{1}{h}\right)^{1/a}hL_{kh}^+<y\right\}\to\Phi_a(y),$$
$$P\{M_{kh}^+/(kh)^{1/2}\leqq f(x_0)^{1/2}\,w_h^a(x)\}\to\exp\{-c_ae^{-x/2}\}.$$

Likewise for L^- and M^-.

Sketch of Proof. Only j's with hj near x_0 matter. Suppose $hj > x_0$, but close. Then,

$$p_j/f(x_0)\,h \doteq 1 - \frac{A}{f(x_0)}(hj)^a$$

so the chance that

$$N_j - kp_j > [f(x_0)\,kh]^{1/2}\,w_h^a(x)$$

is essentially

$$(2\pi)^{-1/2} w_h^a(x)^{-1} \exp\{-\tfrac{1}{2} w_h(x)^2 f(x_0)\,h/p_j\}$$
$$\doteq c_a e^{-x/2}\,\delta K_a^{-1} \exp\{-(\delta j)^a\}$$

where

$$\delta = \left[\frac{A}{f(x_0)} \log \frac{1}{h}\right]^{1/a} h.$$

The argument can be done as in [4]. □

The conclusions about M_{kh}^+ and L_{kh}^+ continue to hold, even if e.g. $B=0$; and the theorem then handles the case of an exponential density.

We turn next to the regularity conditions assumed in section 3 to deal with the bias term. One such was (3.7), which required the existence of three continuous derivatives.

(4.13) *Example.* There is a C_2 density f, which is positive everywhere, and locally quadratic at its unique global maximum; f and f' and f'' vanish at ∞; and f is C_∞ except at 0. Furthermore, $\sqrt{f(t)} + |f'(t)|$ has a unique global maximum at 0, but is not locally quadratic there, because f''' does not exist at 0. The conclusions of Theorem (3.14) fail, for this f.

Construction. In a small neighborhood of 0, set

$$f(x) = 1 + 2x - \tfrac{1}{2}x^2 - \tfrac{2}{5}C|x|^{5/2}$$

for $C>0$. Then

$$f'(x) = 2 - x - C|x|^{3/2} > 0$$

and

$$f(x)^{1/2} = 1 + x + O(x^2)$$

so

$$f(x)^{1/2} + |f'(x)| = 3 - C|x|^{3/2} + O(x^2).$$

In particular, $f(x)^{1/2} + |f'(x)|$ has a strict local maximum at 0. Continue f over the whole line so as to satisfy the conditions of the example, with 0 being the unique location of the global maximum of $f^{1/2} + |f'|$. To show that (3.14) fails, start at (3.8). Choose k and h so $\gamma_h \equiv 1$. Then

$$\alpha(t) + \gamma_h \beta(t) = f(x_0 + t)^{1/2} + |f'(x_0 + t)|$$

will have its global maximum at $t_h = t_0 = -x_0$. Define M'_{kh} and L'_{kh} as in (3.14).

Their joint asymptotic distribution will now be computed: L'_{kh} is not asymptotically normal, and the scale for M'_{kh} changes

Following the notation in (4.8–11), let

$$K = \int_{-\infty}^{\infty} \exp(-|u|^{3/2})\, du$$

$$\Psi(y) = K^{-1} \int_{-\infty}^{y} \exp(-|u|^{3/2})\, du \tag{4.14}$$

$$c = \tfrac{1}{2}\pi^{-1/2} K (2C)^{-2/3}.$$

Then the chance that

$$\left(2C \log\frac{1}{h}\right)^{2/3} h L'_{kh} \leqq y$$

and

$$(kh)^{1/2} M'_{kh} \leqq \left[2\log\frac{1}{h} - \frac{7}{3}\log\log\frac{1}{h} + x\right]^{1/2} + 2\left[2\log\frac{1}{h}\right]^{1/2}$$

converges to

$$\psi(y) \cdot \exp\{-c e^{-x/2}\}.$$

Sketch of Proof. All the action occurs near 0, and in effect we are studying

$$\max_j \left\{\alpha(hj) Z_{hj} + [2 + \beta(hj)] \left[2\log\frac{1}{h}\right]^{1/2}\right\}$$

where

$$\alpha(u) = 1 + u,$$

$$\beta(u) = -u - C|u|^{3/2},$$

$$Z_{hj} = (N_j - k p_j)/(k p_j)^{1/2}.$$

We have to estimate the chance that

$$\alpha(hj) Z_{hj} + \beta(hj) \left(2\log\frac{1}{h}\right)^{1/2} > w$$

where $w = w_h^{3/2}(x)$, the scale defined in (4.8). Let

$$\lambda = \frac{w - \beta(hj)\left(2\log\frac{1}{h}\right)^{1/2}}{\alpha(hj)}.$$

Then $\lambda \approx \left(2\log\frac{1}{h}\right)^{1/2}$. More particularly,

$$\lambda^2 \doteq w^2 - 2w^2 hj - 2w\beta(hj)\left(2\log\frac{1}{h}\right)^{1/2}$$

$$\doteq w^2 + 2(\delta j)^{3/2}$$

where

$$\delta = \left(2C\log\frac{1}{h}\right)^{2/3} \cdot h,$$

because

$$-2w^2 hj \doteq -4\left(\log\frac{1}{h}\right) hj$$

$$-2w\beta(hj)\left(2\log\frac{1}{h}\right)^{1/2} \doteq 2w\cdot hj\cdot\left(2\log\frac{1}{h}\right)^{1/2} + 2wC|hj|^{3/2}\cdot\left(2\log\frac{1}{h}\right)^{1/2}$$

$$\doteq 4\left(\log\frac{1}{h}\right) hj + 4C\left(\log\frac{1}{h}\right)|hj|^{3/2}.$$

So the chance that $Z_{hj} > \lambda$ is nearly

$$(2\pi)^{1/2}\lambda^{-1}\exp(-\tfrac{1}{2}\lambda^2) \doteq c\,\delta K^{-1}\exp\{-(\delta j)^{3/2}\}.$$

The argument is done in detail as in [4]. □

The density in (4.13) is somewhat contrived. Our last example uses a standard denisty: the beta.

(4.15) *Example.* Let $f(x) = \frac{3}{2}\sqrt{x}$ for $0 \leqq x \leqq 1$, and vanish elsewhere. This violates all our regularity conditions, for the maximum of f occurs at the endpoint 1, where f is locally linear:

(4.16) $$f(1-\varepsilon) = \tfrac{3}{2} - \tfrac{3}{4}\varepsilon + 0(\varepsilon^2) \quad \text{as } \varepsilon \to 0^+.$$

Furthermore, $f(x)^{1/2} + f'(x)$ has a maximum of infinity at $x=0$, where f vanishes. The asymptotic behavior of L'_{kh} and M'_{kh}, defined as in (3.14), is very different from what is described there. Indeed, with positive probability the max will occur near 0 and be due to bias. With the remaining positive probability the max will occur near 1 and be due to sampling error. The max cannot occur in the interior. These paradoxical assertions are true provided k goes to infinity and h goes to zero at the right rate (4.21) below. The parameter λ in that equality controls the balance between 0 and 1.

The starting point is (3.1). Consider the cell $[0, h]$. There,

(4.17) $$kh\max(f_h - f) = khf_h = kh^{3/2}.$$

Cells $[h, 2h]$, $[3h, 4h]$, etc. have strictly smaller bias terms, and do not matter. Indeed, since f is strictly concave,

(4.18) $$h^{-1}\int_x^{x+h} f(u)\,du - f(x) \quad \text{is strictly decreasing as } x \text{ increases.}$$

We are interested in $x = jh$. Near 0, the sampling errors $N_j - kp_j$ are essentially normal, with standard deviations of the order $k^{1/2}h^{3/4}$. These will not matter either.

Consider next a neighborhood of 1. There, bias will be negligible, and the sampling-error component $\max_j(N_j-kp_j)$ can be handled by (4.12). The relevant a is 1 by (4.16) and this is substituted into the scale w_h^a defined at (4.8). In more detail, let M_1 be the max of (N_j-kp_j) over cells in say [.01, 1], and let L_1 be the location of this maximum; by our convention, L_1 is a negative integer, counting cells down from $x_0=1$. Now if the growth condition $k \gg \frac{1}{h}\left(\log\frac{1}{h}\right)^3$ is satisfied, L_1 and M_1 are asymptotically independent. Furthermore, L_1 is asymptotically exponential:

$$P\left\{\tfrac{1}{2}\left(\log\frac{1}{h}\right)\cdot h\cdot(-L_1)>y\right\}\to e^{-y}.$$

And M_1 is asymptotically double-exponential, but on the scale w_h^1:

$$P\left\{M_1<(\tfrac{3}{2}kh)^{1/2}\left(2\log\frac{1}{h}-3\log\log\frac{1}{h}+x\right)^{1/2}\right\}\to\exp\{-c_1e^{-x/2}\} \tag{4.19}$$

where c_1 is defined by (4.11) with $a=1$ and $f(x_0)=3/2$.

Fix any real number λ. In particular M_1 will be of order

$$(\tfrac{3}{2}kh)^{1/2}\left[2\log\frac{1}{h}-3\log\log\frac{1}{h}+2\lambda\right]^{1/2}. \tag{4.20}$$

Choose k and h to equalize sampling-error (4.20) and bias (4.17).

$$k=3\left(\frac{1}{3}\right)^2\left[\log\frac{1}{h}-\tfrac{3}{2}\log\log\frac{1}{h}+\lambda\right]. \tag{4.21}$$

This is different from the rate defined in (3.9).

Now, let $h\to 0$ and $k\to\infty$ at the critical rate (4.21). In particular, (4.17) and (4.20) both tend to infinity. To avoid trival difficulties, suppose h is the reciprocal of an integer, and 0 falls on the boundary of a cell. The asymptotics of $\max(H-f)$ can now be described. Recall from (3.1) that $(kh)(H-f)$ and M_1 are comparable. Near 0, $(H-f)\doteq kh^{3/2}$ by (4.17). Near 1, we have (4.19). Now M_1 exceeds $kh^{3/2}$ with probability approaching $\exp\{-c_1e^{-\lambda}\}$. If M_1 exceeds $kh^{3/2}$, then with conditional probability approaching one, $\max(H-f)=M_1/kh$: so the max occurs near 1. If M_1 fails to exceed $kh^{3/2}$, then with conditional probability approaching one, $\max(H-f)$ occurs in the cell [0, h], and is essentially $h^{1/2}$.

The conditional probability fails to be one exactly due to the normal sampling error surrounding the bias $kh^{3/2}$. However, as noted above, this error is of order $k^{1/2}h^{3/4}$, and is much smaller than the the randomness in M_1, which is of order $\left(kh/\log\frac{1}{h}\right)^{1/2}$; asymptotically, this all washes out.

The interval from e.g. 100h to 0.01 requires special attention. There, sampling error and bias can be estimated separately, by (2.4) and (4.17) respectively, and added. They are of too small an order of magnitude to matter. We omit further details. □

References

1. Bickel, P.J., Rosenblatt, M.: On some global measures of the deviations of density function estimates. Ann. Statist. **1**, 1071–1095 (1973)
2. Diaconis, P., Freedman, D.: The distribution of the mode of an empirical historgram. Stanford Technical Report No. 105 (1979)
3. Freedman, D.: Another note on the Borel-Cantelli lemma and the strong law, with the Poisson approximation as a by-product. Ann. Probab. **1**, 920–925 (1973)
4. Freedman, D.: On the maximum of scaled multinomial variables. Pacific J. Math. [To appear (1981)]
5. Freedman, D., Diaconis, P.: On the maximum difference between the empirical and expected histograms for sums. Pacific J. Math. [To appear (1981a)]
6. Freedman, D., Diaconis, P.: On the maximum difference between the empirical and expected histograms for sums: Part II. Pacific J. Math. [To appear (1981b)]
7. Freedman, D., Diaconis, P.: On the histogram as a density estimator: L^2 theory. Z. Wahrscheinlichkeitstheorie verw. Gebiete [To appear (1981c)]
8. Freedman, D., Diaconis, P.: On the mode of an empirical historgram for sums. Pacific J. Math. [To appear (1981d)]
9. Hasminsky, R.: On the lower bounds for risks of nonparametric estimates of density functions in the uniform metrics. Probability Theory and its Applications **23**, 824–828 (1978)
10. Reiss, R.D.: Approximate distributions of the maximum deviation of histograms. Preprint in statistics. Univ. Köln (1976)
11. Revesz, P.: On empirical density function. Periodica Math. Hungarica **2**, 85–110 (1972)
12. Serfling, R.: Properties and applications of metrics on nonparametric density estimators. Johns Hopkins University Technical Report No. 432 (1980)
13. Smirnov, N.V.: Approximation of distribution laws of random variables by empirical data. Uspehi. Mat. Nauk. **10**, 179–206 (in Russian) (1944)
14. Tumanjan, S.H.: On the maximal deviation of the empirical density of a distribution. Nauk. Trudy. Erevansk. Univ. **48**, 3–48 (1955)
15. Woodroofe, M.: On the maximum deviation of the sample density. Ann. Math. Statist., **38**, 475–481 (1967)

Received October 6, 1980