ON THE DIFFERENCE BETWEEN THE EMPIRICAL HISTOGRAM
AND THE NORMAL CURVE, FOR SUMS: PART II

BY

DAVID FREEDMAN and PERSI DIACONIS

TECHNICAL REPORT NO. 160
AUGUST 1980

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT MCS77-16974

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ON THE DIFFERENCE BETWEEN THE EMPIRICAL HISTOGRAM
AND THE NORMAL CURVE, FOR SUMS: PART II

BY

DAVID FREEDMAN and PERSI DIACONIS

TECHNICAL REPORT NO. 160
AUGUST 1980

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT MCS77-16974

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
On the Difference Between the Empirical Histogram 
and the Normal Curve, for Sums: Part II 

by 

David Freedman\textsuperscript{1} 
University of California, Berkeley 

and 

Persi Diaconis\textsuperscript{2} 
Stanford University 

1. Introduction 
Let \( X_1, X_2, \ldots \) be independent, identically distributed random variables. Suppose the \( X_i \) are integer-valued and have g.c.d. one: 

\begin{equation}
\text{g.c.d.}\{m : P(X_i = m) > 0\} = 1.
\end{equation}

Suppose too 

\begin{equation}
E(X_i^3) < \infty.
\end{equation}

Let 

\begin{equation}
\mu = E(X_1), \quad \sigma^2 = \text{Var} X_1, \quad \mu_3 = E[(X_1 - \mu)^3].
\end{equation}

Let \( S_n = X_1 + \ldots + X_n \). Take \( k \) independent copies of \( S_n \), and let 

\textsuperscript{1}Research partially supported by NSF Grant MCS-77-01665. 
\textsuperscript{2}Research partially supported by NSF Grant MCS-77-16974.
Let \( N_{nj} \) be the number of these sums which are equal to \( j \). Up to scaling, the counts \( N_{nj} \) correspond to the empirical histogram for the \( k \) sums. Of course,
\[
E(N_{nj}) = k \, p_{nj}, \text{ where } p_{nj} = P(S_n = j).
\]

In a previous paper [2] we studied the behavior of

\[
\max_j (N_{nj} - k \, p_{nj}),
\]

(1.4)

corresponding to the maximum deviation between the empirical histogram and its expected value. In this paper we will study the maximum deviation between the empirical histogram and an approximation to the expected value, based on the normal curve.

In more detail, the probabilities \( p_{nj} \) can be well approximated by

\[
\hat{p}_{nj} = \frac{1}{\sigma \sqrt{2\pi n}} \exp\left(-\frac{1}{2} \frac{t_{nj}^2}{n}\right)
\]

(1.5)

where

\[
t_{nj} = (j - n\mu)/(\sigma \sqrt{n}).
\]

The asymptotic behavior of the maximum deviation

\[
\max_j (N_{nj} - k \, \hat{p}_{nj}),
\]

(1.6)

is the topic of this paper.
Our main results give the asymptotic distribution of the location and size of this maximum deviation. When the number of repetitions \( k \) is "small", sampling error dominates and the maximum deviation is asymptotically the same as the maximum in (1.4). When the number of repetitions \( k \) is "large", the bias term enters. In both cases the maximum deviation is taken on at a unique location with probability approaching one. The location and size of the maximum are asymptotically independent and, suitably normalized; the location has a limiting normal distribution while the size has a limiting extreme value distribution. The results are more carefully described in Section 2.

Of course, the empirical histogram could be approximated by the normal curve directly. In this case too, the asymptotic behavior of the maximum deviation can be analyzed by methods very similar to the ones presented here, but we do not pursue the details. Similar remarks apply to the frequency polygon derived from the empirical histogram, and to Edgeworth expansions for \( \hat{p}_{nj} \).

2. The Normal Approximation

Clearly,

\[
N_{nj} - k \hat{p}_{nj} = (N_{nj} - k p_{nj}) + k(p_{nj} - \hat{p}_{nj}).
\]

The first term on the right represents sampling error; the second, bias. Suppose that

\[
k/[n^{1/2} (\log n)^3] \to \infty \text{ as } n \to \infty.
\]
This condition insures that the histogram converges uniformly to the expected histogram $p_{nj}$. See [3] for further discussion. Suppose too

$$
(2.3) \quad \mu_3 = E[(X_1 - \mu)^3] \neq 0, \text{ where } \mu = E(X_1).
$$

The results of this section can be summarized as follows.

If $k \ll n^{3/2}$, bias is negligible, so $\max_j (N_{nj} - k \tilde{p}_{nj})$ shows the same asymptotic behavior as $\max_j (N_{nj} - k p_{nj})$. This maximum has been carefully analyzed in [2].

If $k >> n^{3/2} \log n$, sampling error is negligible. The maximum is analyzed in Section 3.

If $k$ is between $n^{3/2}$ and $n^{3/2} \log n$ in order of magnitude, sampling error and bias both contribute to $\max_j (N_{nj} - k \tilde{p}_{nj})$. The asymptotic behavior of $\max_j (N_{nj} - k \tilde{p}_{nj})$ will be described in this section.

If $\mu_3 = 0$, the critical rates for $k$ change: we do not pursue this. Likewise, if (2.2) fails, the asymptotics change: large deviation corrections become relevant. We do not pursue this either. Finally, if the fourth-moment condition is dropped, new behavior is possible: see Section 5 of [2] for a related discussion. We begin with case $k = 0(n^{3/2} \log n)$, and use theorem (1.24) of [4].

The following notation will be helpful, although it seems tedious indeed. In view of (1.1-1.2), we have from the Edgeworth expansion

$$
(2.4) \quad \sqrt{\frac{2}{n}} (p_{nj} - \tilde{p}_{nj}) = \frac{c}{\sqrt{n}} H_3(t_{nj}) \exp(-\frac{1}{2} t_{nj}^2) + O(\frac{1}{n})
$$
where

\[ t_{nj} = (j - \mu)/\sigma \sqrt{n} \]

\[ H_3(t) = t^3 - 3t \]

\[ c = \frac{1}{6} \mu_3/\sigma^3 \]

\[ \mu = E(X_1), \quad \sigma^2 = \text{Var } X_1, \quad \mu_3 = E[(X_1 - \mu)^3] . \]

The "0" is uniform in \( j \). Let

(2.5) \[ \xi = \sqrt{\frac{k}{\sigma \sqrt{2\pi n}}} . \]

(2.6) \[ (N_{nj} - k \tilde{p}_{nj})/\xi = \alpha_{nj} Z_{nj} + \beta_{nj} [2 \log (\sigma \sqrt{n})]^{1/2} \]

where

(2.7) \[ Z_{nj} = (N_{nj} - k \tilde{p}_{nj})/\sqrt{k \tilde{p}_{nj}} \]

(2.8) \[ \alpha_{nj} = (\sigma \sqrt{2\pi n} \tilde{p}_{nj})^{1/2} \]

(2.9) \[ \beta_{nj} = [2 \log (\sigma \sqrt{n})]^{-1/2} \cdot \xi \cdot \sigma \sqrt{2\pi n} (\tilde{p}_{nj} - \tilde{p}_{nj}) . \]

By (2.4) \( \beta_{nj} \) can be approximated as

(2.10) \[ \beta_{nj} = \gamma_n \beta(t_{nj}) + 0(\gamma_n/\sqrt{n}) \]
where

\[(2.11) \quad \gamma_n = [2 \log (\sigma \sqrt{n})]^{-1/2} \cdot n^{-3/4} \cdot k^{1/2} \]

\[(2.12) \quad \beta(t) = \sigma^{-1/2} (2\pi)^{-1/4} c_H(t) \exp(-\frac{1}{2} t^2) . \]

Let

\[(2.13) \quad \alpha(t) = \exp(-\frac{1}{4} t^2) \]

\[(2.14) \quad \omega_n(x) = (\log n - 2 \log \log n + x + \log 4\sigma^2)^{1/2} \]

\[(2.15) \quad \phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} \exp(-\frac{1}{2} u^2) du . \]

The main result of this section is (2.17), which disposes of the case \( k = 0(n^{3/2} \log n) \). We next give the precise conditions for this result to hold.

**Condition for (2.16).** Suppose (1.1-1.3) and (2.2). Do not assume (2.3). Define \( \gamma_n \) by (2.11). Suppose \( \gamma_n + \gamma \) finite as \( n \to \infty \). Note that \( \gamma = 0 \) is allowed. Suppose, as will be the case for most \( \gamma \)'s, that the function \( \alpha + \gamma \beta \) has a unique global maximum, say at \( t_\infty \); and that \( \alpha''(t_\infty) + \gamma \beta''(t_\infty) < 0 \). Abbreviate

\[(2.16) \quad \rho^2 = -[\alpha''(t_\infty) + \gamma \beta''(t_\infty)]/\alpha(t_\infty) > 0 . \]
As is seen early, for \( n \) sufficiently large, \( \alpha + \gamma_n \beta \) has a unique global maximum, say at \( t_n \); and \( t_n \to t_\infty \).

(2.17) **Proposition.** Suppose the conditions given above: in particular,

\[
\frac{1}{n} \left( \log n \right)^3 \ll k = 0(n^{3/2} \log n).
\]

With probability approaching one, \( M_n = \max_j (N^j_n - k \tilde{P}_n^j) \) is assumed at a unique index \( L_n \). Furthermore, the chance that

\[
\rho \left[ 2 \log (\sigma \sqrt{n}) \right]^{1/2} \cdot \left[ \frac{1}{\sigma \sqrt{n}} \left( L_n - n\mu \right) - t_n \right] < y
\]

and

\[
M_n / \xi < \alpha (t_n) \varphi_n (x) + \alpha_n \beta (t_n) \left[ 2 \log (\sigma \sqrt{n}) \right]^{1/2}
\]

converges to

\[
\phi(y) \exp \left\{ - \frac{1}{2^\rho} e^{-x/2} \right\}.
\]

**Proof.** Let \( I \) be a long (but finite) closed interval, which contains \( t_\infty \) as an interior point. If the \( j \) in \( \max_j (N^j_n - k \tilde{P}_n^j) \) is restricted so that \( t_n \in I \), the conclusions of the proposition follow from theorem (1.24) of [4], taking \( \varepsilon_n = 1/(\sigma \sqrt{n}) \) and \( c_n = n\mu \) and \( \alpha_n = \alpha \) and \( \beta_n = \gamma_n \beta \), so \( \beta_\infty = \gamma \beta \). Conditions (1.1-23) of [4] are satisfied by our assumptions and the Edgeworth expansion (2.4).
It only remains to show that if $I$ is long enough, the $j$'s with $t_{nj} \notin I$ make essentially no contribution to the maximum: compare (2.34) of [4]. Indeed, the maximum over $I$ has been proved to be of order

$$\rho \cdot \left[ \alpha(t_{\infty}) + \gamma \beta(t_{\infty}) \right] \cdot \sqrt{\log n}.$$ 

Now $t_{\infty}$ is the location of the global maximum of $\alpha(\cdot) + \gamma \beta(\cdot)$, which is necessarily positive, for $\alpha(0) + \gamma \beta(0) > 0$. By (4.1-3) of [2], for $I$ long enough, the maximum of $N_{nj} - k p_{nj}$ over $j$ with $t_{nj} \notin I$ is, with probability near one only a small multiple of $\lambda \sqrt{\log n}$. Likewise, by (2.4) of the present paper, the maximum of $k(p_{nj} - \tilde{p}_{nj})$ over $j$ with $t_{nj} \notin I$ is bounded above by

$$c \cdot n^{-1/2} \lambda^2 \left[ \sup_{t \notin I} H_3(t) \exp(-\frac{1}{2} t^2) \right] + o(n^{-1} \lambda^2).$$

In the remainder term, $n^{-1} \lambda^2 = o(\lambda(\log n)^{1/2})$. In the lead term, $n^{-1} \lambda^2 = 0(\lambda \cdot (\log n)^{1/2})$, and the supremum is small for $I$ long. It was at this point that the growth condition $k = O(n^{3/2} \log n)$ became critical.

(2.18) **Corollary.** Suppose (1.1-2). Suppose $k/n^{1/2} (\log n)^3 \to \infty$ as $n \to \infty$, but $k/n^{3/2} \to 0$. Then, the asymptotic joint distribution of the location $L_n$ and size $M_n$ of $\max_j (N_{nj} - k \tilde{p}_{nj})$ coincides with that of $\max_j (N_{nj} - k p_{nj})$, as determined in [2].

**Proof.** Clearly, $\alpha$ has its maximum of 1 at $t_{\infty} = 0$, where it is locally quadratic:
\[ \alpha(t) = 1 - \frac{1}{4} t^2 + o(t^4) \text{ as } t \to 0. \]

Furthermore,

\[ \beta(t) = bt + o(t^3) \text{ as } t \to 0, \]

where \( b \) depends on \( \sigma \) and \( \mu_3 \); it may vanish. Recall that \( t_n \) is the location of the maximum of \( \alpha + \gamma_n \beta \). Recall \( \gamma_n \) from (2.11) and verify that \( \gamma_n \to 0 \).

Some easy calculus shows that

\[ t_n = o(\gamma_n) = o[\log (\sigma \sqrt{n})^{-1/2}] \]

so \( t_n \) may be dropped from the normalization of \( L_n \) in (2.11). Likewise,

\[ \gamma_n \beta(t_n) = o(\gamma_n^2) = o(1/\log n) \]

so the term

\[ \gamma_n \beta(t_n)[2 \log (\sigma \sqrt{n})]^{1/2} = o[(\log n)^{-1/2}] \]

can be dropped from the normalization of \( M_n \). Finally,

\[ \alpha(t_n) = 1 - o(\gamma_n^2) \]

so \( \alpha(t_n) \) can be replaced by 1 in the normalization.

3. The Bias Term

We now consider the case \( k/[n^{3/2} \log n] \to \infty \), when the bias term in (2.1) dominates. Assumption (2.3) is back in force. Recall \( \beta \)
from (2.12). Note that $c = \mu_3/[6 \sigma^3] \neq 0$; without real loss of generality, suppose $c > 0$.

The graph of $\beta$ is sketched in Figure 1 below. This function is anti-symmetric about 0, where it vanishes. Likewise, it vanishes at $\pm \infty$. It has four critical points, at the roots of $x^4 - 6x^2 + 3 = 0$. The global max occurs at

$$
t^* = -(3 - \sqrt{6})^{1/2} \approx -0.74.
$$

The second derivative of $\beta$ does not vanish at any of the four critical points.

**Figure 1.** The graph of $\beta$

(3.2) **Proposition.** Suppose (1.1-2), and $k \gg n^{3/2} \log n$. Suppose $\mu_3 > 0$, the case $\mu_3 < 0$ being symmetric. Then

$$
\sqrt{n} \max_j (N_{nj} - k \tilde{p}_{nj})/\beta(t^*) \chi^2 + \sigma^{1/2}(2\pi)^{1/4}
$$

in probability. Furthermore, for any $\delta$ positive, with probability
approaching one, the \( \max \) is taken on only for \( j \)'s with

\[
|t_{nj} - t^*| < \delta.
\]

**Proof.** Refer back to (2.1). As theorem (1.8) of [2] demonstrates, the sampling error term in (2.1) is of order \( \ell \cdot (\log n)^{1/2} \). On the other hand, the Edgeworth expansion (2.4) shows the bias term to be of order

\[
(3.3) \quad n^{-1/2} \lambda^2 \sigma^{1/2} (2\pi)^{1/4} \beta(t_{nj}) + O(n^{-1/2} \lambda^2).
\]

If \( t_{nj} \) is close to \( t^* \), then \( \beta(t_{nj}) \) is close to \( \beta^* > 0 \), and \( n^{-1/2} \lambda^2 \) dominates the sampling-error order \( \ell \cdot (\log n)^{1/2} \). Plainly, \( n^{-1/2} \lambda^2 \) also dominates the remainder \( n^{-1} \lambda^2 \). If on the other hand \( t_{nj} \) is bounded away from \( t^* \), then \( \beta(t_{nj}) \) is bounded below by \( \beta(t^*) \), so (3.3) is too small to influence the \( \max \).

Suppose

\[
(3.4) \quad \mu_3 > 0 \text{ and } n^{3/2} \log n < k < n^{5/2}/\log n.
\]

Then the asymptotic distribution of \( \max_j \{N_{nj} - k \tilde{p}_{nj}\} \) can still be deduced from [3, (1.24)], as we show next. The case \( \mu_3 < 0 \) is symmetric, but \( \mu_3 = 0 \) is different. If \( k \) is of order \( n^{5/2}/\log n \) or larger, more terms in the Edgeworth expansion (2.4) for \( p_{nj} \) become relevant. If \( k \) is of order \( n^{7/2} \) or larger, the asymptotic distribution becomes degenerate. We do not pursue these issues here: See section 6 of [1] for a related discussion.

11
Before stating Proposition (3.19) formally, we indicate the heuristics. Proposition (3.2) suggests that
\[
\max_j (N_{nj} - k \tilde{p}_{nj}) / \ell
\]
is essentially
\[
\frac{1}{\ell} (2\pi)^{1/4} \frac{\beta(t^*)}{\ell/\sqrt{n}} = \gamma_n \beta(t^*) [2 \log (\sigma \sqrt{n})]^{1/2}.
\]
We consider the difference. The idea is to use [3, (1.24)] again, but on a new scale and with new functions \( \alpha \) and \( \beta \). The starting point is (2.6). In particular,

\[
(3.5) \quad \ell^{-1} (N_{nj} - k \tilde{p}_{nj}) - \gamma_n \beta(t^*) [2 \log (\sigma \sqrt{n})]^{1/2} = \alpha_n Z_{nj} + \gamma_n
\]
where

\[
(3.6) \quad \gamma_n = [\beta_{nj} - \gamma_n \beta(t^*)] [2 \log (\sigma \sqrt{n})]^{1/2}.
\]
Now \( \beta \) is locally quadratic at \( t^* \), and in effect we expand \( \gamma_n \) around \( t^* \). Informally, by (2.9-11),

\[
\beta_{nj} = \gamma_n \beta(t_{nj})
\]
so
\[
\gamma_n = [\beta(t_{nj}) - \beta(t^*)] \cdot \gamma_n \cdot [2 \log (\sigma \sqrt{n})]^{1/2}
\]
\[
= \frac{1}{2} \beta''(t^*) (t_{nj} - t^*)^2 \cdot k^{1/2} / n^{3/4}
\]
\[
= \frac{1}{2} \beta''(t^*) \sigma^{-2} (j - n\mu - \sigma \sqrt{n} t^*)^2 \cdot k^{1/2} / n^{7/4}.
\]
Parenthetically, this heuristic can be made rigorous if
\( k >> n^{3/2}(\log n)^5 \), but is a bit too aggressive with smaller \( k \)'s.

The center \( c_n \) called for in [3, (1.23)] is now defined as follows:

\[
(3.7) \quad c_n = n\mu + \sigma \sqrt{n} t^* .
\]

The scale factor \( \varepsilon_n \) should satisfy

\[
\varepsilon_n^2 \sqrt{2 \log \frac{1}{\varepsilon_n}} = k^{1/2}/n^{7/4} .
\]

We set

\[
(3.8) \quad m = n^{7/8}/k^{1/4} \quad \text{and} \quad \varepsilon_n = m^{-1}(2 \log m)^{-1/4} .
\]

Thus,

\[
\gamma_{nj} \doteq \frac{1}{2} \beta''(t^*) \sigma^{-2} \frac{\sigma_n}{\varepsilon_n} \sqrt{2 \log \frac{1}{\varepsilon_n}}
\]

where we write

\[
(3.9) \quad \theta_{nj} = \varepsilon_n (j - c_n) .
\]

This is to avoid confusion with the \( t_{nj} = (j - n\mu)/(\sigma \sqrt{n}) \) used ear-
lier. More formally, to make contact with [3] we view \( k \) and hence
\( m \) as functions of \( n \). We make the definitions (3.6-3.9). Let \( I \)
be a long (finite) closed interval with \( 0 \) as an interior point. Let
\[ (3.10) \quad \tilde{\beta}_{nj} = \gamma_{nj} \cdot \left[ 2 \log \frac{1}{\varepsilon_n} \right]^{-1/2}. \]

We propose to study the maximum of

\[ \bar{z}^{-1}(N_{nj} - k \tilde{p}_{nj}) - \gamma_n \beta(t^*) [2 \log (\sigma \sqrt{n})]^{1/2} \]

\[ = \alpha_{nj} Z_{nj} + \tilde{\beta}_{nj} \cdot \left[ 2 \log \frac{1}{\varepsilon_n} \right]^{1/2} \]

over \( j \)'s with \( \theta_{nj} \in I \), using [3]. For the function \( \alpha_n \) of \([3, (1.4)]\) we take

\[ (3.12) \quad \alpha_n(\theta) = \alpha(t^* + \sigma^{-1} \delta_n \theta) \]

where \( \alpha \) was defined in (2.13) and

\[ (3.13) \quad \delta_n^4 = k^{-1} n^{3/2} (2 \log m). \]

Thus, \( \alpha_n \) is defined over the whole real line. Clearly,

\[ (3.14) \quad t_{nj} = t^* + \sigma^{-1} \delta_n \theta_{nj}. \]

So \( \alpha_n(\theta_{nj}) = \alpha(t_{nj}) \). Note, however, that \( t_{nj} \) is centered at \( t^* \) while \( \theta_{nj} \) is centered at \( 0 \). Further, \( \delta_n \to 0 \) because \( k \gg n^{3/2} \log n \) by (3.4). Thus, \( t_{nj} \to 0 \) uniformly over \( j \) with \( \theta_{nj} \in I \).
For the function $\beta_n$ in [3, (1.5)] we take

$$
\beta_n(\theta) = \left(\log m/\log \frac{1}{\delta_n}\right)^{1/2} \cdot \delta_n^{-2} \cdot [\beta(t^* + \sigma^{-1} \delta_n \theta) - \beta(t^*)]
$$

where $\beta$ was defined in (2.12).

Before proceeding, it is helpful to note

$$
\frac{2}{\delta} \log n < \log m < \frac{7}{\delta} \log n \text{ and } \log \frac{1}{\delta_n} \approx \log m
$$

by the growth condition (3.4). We claim

$$
\alpha_n^j = \alpha_n(\theta_n^j) + o(1/\log m)
$$

as $n \to \infty$, uniformly in $j$ with $t_{nj}$ confined to a compact interval.

This is routine to verify, estimating $\alpha_n$ from (2.8) and the Edgeworth expansion (2.4), the $\tilde{p}_n^j$ having been defined in (1.5). More explicitly,

$$
\alpha_n^2 = \sigma \sqrt{2\pi n} \tilde{p}_n^j
$$

$$
= \alpha(t_{nj})^2 + o(1/\sqrt{n}).
$$

Now $\alpha(t_{nj})$ cannot run away to zero, and

$$
\alpha_n^j = \alpha(t_{nj}) + o(1/\sqrt{n})
$$

$$
= \alpha_n(\theta_n^j) + o(1/\log m),
$$
because $\sqrt{n} \gg \log n \geq 8/7 \log m$ by (3.16). This completes the proof of (3.17). If $\theta_{nj}$ is confined to $I$, then $t_{nj} = t^* + O(\delta_n)$, and $\delta_n \to 0$ because $k \gg n^{3/2} \log n$. So (3.17) establishes condition [3, (1.4)].

Condition [3, (1.5)] can be verified by tedious algebra: this is where the growth condition $k \ll n^{5/2}/\log n$ comes in. We need a little more:

\begin{equation}
\tilde{\beta}_{nj} = \beta_n(\theta_{nj}) + o(1/\log m),
\end{equation}

as $n \to \infty$, uniformly in $j$.

Indeed,

\begin{align*}
\tilde{\beta}_{nj} &= \gamma_{nj} \cdot \left[2 \log \frac{1}{\epsilon_n}\right]^{-1/2} \\
&= \left[\log(\sigma n)/\log \frac{1}{\epsilon_n}\right]^{1/2} \cdot \left[\beta_{nj} - \gamma_n \beta(t^*)\right] \quad \text{by (3.6)} \\
&= \gamma_n \cdot \left[\log \sigma n/\log \frac{1}{\epsilon_n}\right]^{1/2} \cdot \left[\beta(t_{nj}) - \beta(t^*) + o(1/\sqrt{n})\right] \quad \text{by (2.10)} \\
&= n^{-3/4} k^{1/2} \left[2 \log \frac{1}{\epsilon_n}\right]^{-1/2} \cdot \left[\beta(t_{nj}) - \beta(t^*) + o(1/\sqrt{n})\right] \quad \text{by (2.11)} \\
&= n^{-3/4} k^{1/2} \left[2 \log \frac{1}{\epsilon_n}\right]^{-1/2} \cdot \left[\beta(t_{nj}) - \beta(t^*)\right] + o(1/\log m) \\
\end{align*}

because $k \ll n^{5/2}/\log n$ and $\log m \sim \log n$

\begin{align*}
&= n^{-3/4} k^{1/2} \delta_n^2 \left[2 \log \frac{1}{\epsilon_n}\right]^{-1/2} \delta_n^{-2} \left[\beta(t_{nj}) - \beta(t^*)\right] + o(1/\log n) \\
\end{align*}
\[ \log m / \log \left( \frac{1}{\epsilon_n} \right)^{1/2} \delta_n^{-2} [\beta(t_{nj}) - \beta(t^*)] + o(1/\log m) \quad \text{by (2.31)} \]

\[ = \beta_n(\theta_{nj}) + o(1/\log m) \quad \text{by (2.32-2.33)} . \]

This completes the argument for (2.36), i.e., condition [3, (1.5)].

Condition [3, (1.6)] is clear. For [3, (1.7)], let \( \alpha_\infty(\theta) = \alpha(t^*) \)
for all \( \theta \), so \( \alpha_n + \alpha_\infty \) because \( \alpha \) is continuous and \( \delta_n \to 0 \). Likewise, let

\[ \beta_\infty(\theta) = \frac{1}{2} \beta''(t^*) \sigma^{-2} \theta^2 . \]

That \( \beta_n + \beta_\infty \) can be verified by expanding \( \beta \) in a Taylor series
around \( t^* \), the location of its maximum; \( \beta'(t^*) = 0 \) and \( \beta''(t^*) < 0 \).

Condition [3, (1.8)] is clear, at least for large \( n \). We write \( \theta_n \)
for the location of the maximum of \( \alpha_n + \beta_n \), and note that \( \theta_\infty = 0 \).
By calculus, \( \theta_n = O(\delta_n) \), and \( \delta_n \to 0 \). The remaining conditions for
[3, (1.24)] are all verified easily. In conformity with [3, (1.15)],
let

\[ \tilde{\rho}^2 = -[\alpha''(0) + \beta''(0)]/\alpha_\infty(0) \]

\[ = -\beta''(t^*)/[\alpha(t^*)\sigma^2] . \]

(3.19) Proposition. Assume (1.1-2) and (3.4). In the notation given
above, with probability approaching one, \( M_n = \max_j (N_{nj} - k \tilde{\rho}_{nj}) \) is
assumed at a unique index \( L_n \). Furthermore, the chance that
\begin{equation}
\sqrt{2 \log \frac{1}{\varepsilon_n} \left[ \varepsilon_n (L_n - c_n) - \theta_n \right]} < y
\end{equation}

and

\begin{equation}
\varepsilon_n^{-1} M_n - \gamma_n \beta(t^*) \left[ 2 \log (\sigma \sqrt{n}) \right]^{1/2}
\end{equation}

\begin{equation}
< \alpha_n(\theta_n) \left[ 2 \log \frac{1}{\varepsilon_n} - 2 \log \log \frac{1}{\varepsilon_n} + x \right]^{1/2} + \beta_n(\theta_n) \left[ 2 \log \frac{1}{\varepsilon_n} \right]^{1/2}
\end{equation}

converges to

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

**Proof.** If the max is taken only over \( j \) with \( \theta_{nj} \in I \), the conclusion is immediate from [3, (1.24)]. The right side of (3.21) is essentially

\[
[\alpha_n(\theta_n) + \beta_n(\theta_n)] \cdot \left[ 2 \log \frac{1}{\varepsilon_n} \right]^{1/2}
\]

and \( \alpha_n(\theta_n) + \beta_n(\theta_n) > \alpha_n(0) + \beta_n(0) = \alpha(t^*) > 0 \).

Consider the \( j \)'s with

\begin{equation}
\theta_{nj} \in I \text{ but } \left| t_{nj} - t^* \right| < \delta.
\end{equation}

We have to argue that such \( j \)'s do not matter, i.e., the max over such \( j \)'s is of smaller order than \( \alpha(t^*) \cdot \left[ 2 \log 1/\varepsilon_n \right]^{1/2} \). Apply [3, (3.1)] to the \( Z_{nj} \), but use the original scaling, i.e., take the
$\varepsilon_n$ in $[3, (3.1)]$ to be $1/(\sigma\sqrt{n})$. With overwhelming probability

$$\max_j Z_{nj} < 2[2 \log (\sigma/\varepsilon_n)]^{1/2}$$

$$< 4 \left[ 2 \log \frac{1}{\varepsilon_n} \right]^{1/2}$$

by (3.16). Hence

$$\max_j \left\{ \alpha_{nj} Z_{nj} + \tilde{\beta}_{nj} \left[ 2 \log \frac{1}{\varepsilon_n} \right]^{1/2} \right\}$$

$$\leq \left[ 2 \log \frac{1}{\varepsilon_n} \right]^{1/2} \cdot [4 \max_j \alpha_{nj} + \max_j \tilde{\beta}_{nj}]$$

$$\leq \left[ 2 \log \frac{1}{\varepsilon_n} \right]^{1/2} \cdot [4 \max_j \alpha_n(\theta_{nj}) + \max_j \beta_n(\theta_{nj}) + o(1/\log m)]$$

by (3.17-3.18), where the $\max$ is taken over the $j$'s satisfying (3.12).

Now $\alpha \leq 1$ by its definition (2.13), so the definition (2.30) of $\alpha_n$ shows

$$\max_j \alpha_n(\theta_{nj}) \leq 1.$$ 

Next, $t^*$ is the location of the global maximum of $\beta$ and $\beta$ is locally quadratic at $t^*$, so for $\delta$ small, for $0 < \delta' < \delta$,

$$\max_u \{ \beta(t^* + u) : \delta' \leq u \leq \delta \} = \beta(t^* + \delta')$$

$$\max_u \{ \beta(t^* - u) : \delta' \leq u \leq \delta \} = \beta(t^* - \delta').$$
Let \( I = [-\theta_0, \theta_0] \). Then, by (3.15),

\[
\max_j \beta_n(\theta_{nj}) \leq \max \beta_n(\theta_0), \beta_n(-\theta_0).
\]

So

\[
\limsup_{n \to \infty} \max_j \beta_n(\theta_{nj}) \leq \frac{1}{2} \beta''(t^*) \sigma^{-2} \theta_0^2.
\]

Recall \( \beta''(t^*) < 0 \); choose \( \theta_0 \) so large that

\[
\lambda = 4 + \frac{1}{2} \beta''(t^*) \sigma^{-2} \theta_0^2 < 0.
\]

With overwhelming probability, then

\[
\max \left\{ \alpha_{nj} z_{nj} + \tilde{\beta}_{nj} \left[ 2 \log \frac{1}{\varepsilon_n} \right]^{1/2} \right\} < \lambda \left[ 2 \log \frac{1}{\varepsilon_n} \right]^{1/2} + o(1)
\]

where \( \lambda < 0 \) and the \( \max \) is taken over \( j \)'s satisfying (3.22).

Such \( j \)'s do not matter. Finally, \( j \)'s with \( |t_{nj} - t^*| \geq \delta \) do not matter, by (3.2).

**Note.** \( \gamma_n \to \infty \) by its definition (2.11) and the growth condition \( k \gg n^{3/2} \log n \), and \( \beta(t^*) > 0 \), so the term subtracted from \( \lambda^{-1} M_n \) on the left side of (3.21) is of order

\[
\gamma_n \beta(t^*)(\log n)^{1/2}
\]

with \( \gamma_n \to \infty \). The terms on the right side are of order
\[ \alpha(t^*)(2 \log m)^{1/2} \]

and \( \log m \sim \log n \). Thus, the term on the left dominates, in agreement with (3.2).

(3.23) **Corollary.** Suppose (3.4), and in addition \( k \gg n^{3/2} (\log n)^3 \).

Then the scaling in (3.19) can be simplified: the chance that

\[ \tilde{\rho}(2 \log m)^{1/4} m^{-1} (L_n - n|\mu - \sigma \sqrt{n} t^*) < \gamma \]

and

\[ \xi^{-1} m_n - \gamma_n \beta(t^*)[2 \log (\sigma \sqrt{n})]^{1/2} \]

\[ < \alpha(t^*) \left[ 2 \log m - \frac{3}{2} \log \log m + \frac{1}{4} \log 2 + x \right]^{1/2} \]

converges to

\[ \phi(\gamma) \cdot \exp \left\{ \frac{1}{2 \tilde{\rho}} \frac{1}{e^{-1/2x}} \right\} . \]

**Proof.** Recall that \( \theta_n \) is the location of the maximum of \( \alpha_n + \beta_n \), so \( \theta_n = o(\delta_n) \) as defined in (3.13). With our new condition on \( k \), we have \( \delta_n = o[1/(\log m)^{1/2}] \). So \( \theta_n \) can be dropped in (3.20). Likewise, in (3.21),

\[ \alpha_n(\theta_n) = \alpha(t^*) + o[1/(\log m)^{1/2}] \]

and

\[ \beta_n(\theta_n) = o[1/(\log m)] . \]
References


