MAXIMA OF PARTIAL SAMPLES IN GAUSSIAN SEQUENCES

Short title: MAXIMA OF PARTIAL SAMPLES

Yashaswini Mittal

TECHNICAL REPORT NO. 3
August 1, 1976PREPARED UNDER THE AUSPICES OF
SIAM INSTITUTE FOR MATHEMATICS AND SOCIETY DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

To the best of the patent
office of law is the publication or
the publication of the document(s)
in this letter.

DO NOT HAVE PATENT GROUP

12/8 76 cpl

MASTERPrepared For
THE U.S. ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION
UNDER CONTRACT NO. EY-76-S-02-2874.*000

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

MAXIMA OF PARTIAL SAMPLES IN GAUSSIAN SEQUENCES

Short Title: MAXIMA OF PARTIAL SAMPLES

by

YASHASWINI MITTAL

TECHNICAL REPORT NO. 3

AUGUST 1, 1976

STUDY ON STATISTICS AND ENVIRONMENTAL
FACTORS IN HEALTH

PREPARED UNDER SUPPORT TO SIMS FROM
ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION (ERDA)
ROCKEFELLER FOUNDATION

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

NOTICE
This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Energy Research and Development Administration, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

leg

The problem arose from some practical considerations of collecting air pollution data. Air pollution standards are generally in terms of the maximum concentrations. Various practical difficulties (cost of sampling or missing data) are at times responsible for having only a partial sample available for making inferences. The behavior of the difference of two maxima considered here can thus be used to retract information about the original maximum via the maximum of only a partial sample.

Summary

Maxima of Partial Samples in Gaussian Sequences

Let $\{X_n, n \geq 0\}$ be a Gaussian sequence with $EX_i = 0$; $V(X_i) = 1$ and $EX_i X_j = r_{ij}$. Define $M_n = \max_{0 \leq i \leq n} X_i$ and $m_{n,k} = \max(X_i; i \in G_n)$ where $G_n = (t_1, \dots, t_{n'})$ is a subset of $(0, 1, 2, \dots, n)$ and $n' = [n/k]$ for some integer $k \geq 1$, $[x]$ being the integral part of x . We show that $P\{c_n (M_n - m_{n,k}) \leq x\} \Rightarrow (1 + (k-1)e^{-x})$ as $n \rightarrow \infty$ for all $x \geq 0$ where $c_n = \sqrt{(2\ell nn)^{\frac{1}{2}}}$, if the sequence is "moderately dependent" namely if

- (i) $\sup_{i,j} |r_{ij}| < \delta < 1$
- (1)
- (ii) $|r_{ij}| \leq \rho(i-j)$ for $|i-j| > N_0$ such that $\rho_n \ell nn = o(1)$.

Somewhat surprisingly the same result holds even though the sequence is "strongly dependent" namely if

- (i) $r_{ij} = r(i-j)$; r_n convex for $n \geq 0$ and $r_n = o(1)$.
- (2)
- (ii) $(r(n)\ell nn)^{-1}$ is monotone and $o(1)$.

1. INTRODUCTION: - Let $\{X_n, n \geq 0\}$ be a Gaussian sequence with $EX_n = 0$; $V(X_n) = 1$ which will be called "standard sequence" from now on. Define $M_n = \max_{0 \leq i \leq n} X_i$ and $m_{n,k} = \max(X_i; i \in G_n)$ where $G_n = (t(1), \dots, t(n'))$ is a subset of $(0, 1, \dots, n)$ and $n' = [n/k]$ for some integer $k \geq 1$, $[x]$ being the integral part of x . We assume that

$$(1.1) \quad (i) \quad \sup_{i,j} |EX_i X_j| < \delta < 1$$

$$(ii) \quad |EX_i X_j| \leq \rho(i-j) \quad \text{for} \quad |i-j| > N_0 \quad \text{such that}$$

$$\rho_n \ell n n = o(1) .$$

By Berman (1964, Theorem 3.1), $c_n(M_n - b_n)$ converges in distribution to X

where $c_n = (2\ell n n)^{1/2}$; $b_n = c_n - \frac{\ell n(4\pi \ell n n)}{2c_n}$ and $P(X \leq x) = e^{-e^{-x}}$ for

$-\infty < x < \infty$. It is easy to see that for large n , M_n and $m_{n,k}$ will be approximately the same in size. In Section 2 we show that $c_n(M_n - m_{n,k})$ converges in distribution to a truncated logistic distribution under condition (1.1). (The word "truncated" is used rather freely here. In fact this distribution is the same as the logistic distribution on the positive axis while as the mass of the negative axis is pooled in to an atom at the origin. All references to the truncated logistic distributions should be interpreted in this manner.)

In Section 3 we deal with "strongly dependent" Gaussian sequences but require them to be stationary. Instead of (1.1) we assume there

$$(i) \quad EX_i X_j = r(i-j); \quad r_n \text{ convex for } n \geq 0 \text{ and } r_n = o(1) \\ (1.2)$$

$$(ii) \quad (r(n) \ln n)^{-1} \text{ is monotone and } o(1).$$

(These requirements easily imply 1.1(i).) It is shown via a rather long proof that under (1.2), $c_n(M_{n-k}^{m_{n,k}})$ has the same truncated logistic limit distribution. This fact seems counterintuitive at first and is much harder to prove. Beginning of Section 3 has some comments in this direction.

The problem arose from some practical considerations of collecting air pollution data. Air pollution standards are generally in terms of the maximum concentrations. Various practical difficulties (cost of sampling or missing data) are at times responsible for having only a partial sample available for making inferences. The behavior of the difference of two maxima considered here can thus be used to retract information about the original maximum via the maximum of only a partial sample.

I want to thank my colleague Paul Switzer for suggesting this problem.

2. ASYMPTOTICALLY INDEPENDENT SEQUENCES:-

The basic result is given by

Theorem 2.1:- Let $\{X_n, n \geq 0\}$ be a standard sequence with $r_{ij} = EX_i X_j$ and M_n and $m_{n,k}$ as defined in the beginning of the last section. If (1.1) holds then

$$(2.1) \quad P\left\{M_{n-k}^{m_{n,k}} \leq \frac{x}{c_n}; m_{n,k} \leq b_{n,k} + \frac{z}{c_n}\right\} \rightarrow \frac{\exp\{-e^{-z}(1+(k-1)e^{-x})\}}{1+(k-1)e^{-x}}$$

as $n \rightarrow \infty$ where $c_n = (2\ell n n)^{1/2}$; $b_{n,k} = (2\ell n(n/k))^{1/2} - \frac{\ell n(4\pi \ell n n)}{2c_n}$ and $0 \leq x < \infty$; $-\infty < z < \infty$.

Proof:- Let us denote $\eta_{0,n} = b_{n,k} - \frac{g(n)}{c_n}$ and $\eta_{z,n} = b_{n,k} + \frac{z}{c_n}$. We choose $g(n) \rightarrow \infty$ such that $\{\max(\rho_n \ell n n); \frac{\ell n \ell n n}{(\ell n n)^2}\} e^{4g(n)} = o(1)$. We know by Berman (1964) that $P\{m_{n,k} \leq \eta_{0,n}\} = o(1)$ hence the probability in (2.1) can be replaced by

$$(2.2) \quad P\{M_n - m_{n,k} \leq \frac{x}{c_n}; \eta_{0,n} \leq m_{n,k} \leq \eta_{z,n}\}.$$

Define $m_{n,k}^i = \max\{X_j; j \in G_n; j \neq t(i)\}$. Then

$$\begin{aligned} \{\eta_{0,n} \leq m_{n,k} \leq \eta_{z,n}\} &= \bigcup_{i=1}^{n'} \{m_{n,k}^i \leq X_{t(i)}; \eta_{0,n} \leq X_{t(i)} \leq \eta_{z,n}\} \\ &= \bigcup_{i=1}^{n'} A_i \quad \text{say.} \end{aligned}$$

Now $P\{\bigcup_{i=1}^{n'} A_i\} = \sum_{n=1}^{n'} P(A_i)$ since the intersection of any two of the above sets e.g. $A_i \cap A_j$ is a subset of the set $\{X_{t(i)} = X_{t(j)}\}$. But this set has probability zero unless X_{t_i} and X_{t_j} have correlation 1 and for $i \neq j$ this is impossible by assumption (1.1 ii). Thus the probability in (2.2) is equal to

$$(2.3) \quad \sum_{i=1}^{n'} P\{M_n^D \leq X_{t_i} + \frac{x}{c_n}; m_{n,k}^i \leq X_{t(i)}; \eta_{0,n} \leq X_{t(i)} \leq \eta_{z,n}\}$$

where $M_n^D = \max\{X_u, u \in S_D\}$; $S_D = \{u, 0 \leq u \leq n \text{ but } u \notin G_n\}$ is the maximum of the sample having deleted about n/k variables. Conditioning on X_{t_i} , we get the quantity in (2.3) to be equal to

$$(2.4) \quad \sum_{i=1}^{n'} \int_{\eta_{0,n}}^{\eta_{z,n}} P\{M_n^D \leq y + \frac{x}{c_n}; m_{n,k}^i \leq y | X_{t(i)} = y\} \varphi(y) dy.$$

Note that $\varphi(y) = (2\pi)^{-1/2} \exp\{-y^2/2\}$ is the standard normal density.

Let us define $Y_u^i = \frac{X_u | X_{t(i)} = y - r_{u,t(i)} y}{(1-r_{u,t(i)}^2)^{1/2}}$ $u = 0, 1, \dots, n, u \neq t_i$. We know

that Y_u^i are standard Gaussian with

$$E Y_u^i Y_v^i = \frac{r_{uv} - r_{u,t(i)} r_{v,t(i)}}{(1-r_{u,t(i)}^2)^{1/2} (1-r_{v,t(i)}^2)^{1/2}} = \gamma_{uv}^i \text{ say.}$$

The probability in (2.4) is equal to

$$(2.5) \quad P\{Y_u^i \leq y \left(\frac{1-r_{u,t(i)}}{1+r_{v,t(i)}}\right)^{1/2} + \frac{x}{c_n}, u \in S_D; Y_v^i \leq y \left(\frac{1-r_{v,t(i)}}{1+r_{u,t(i)}}\right)^{1/2}, v \in S_i\}$$

where $S_i = G_n - \{t(i)\}$. The absolute difference of (2.5) and the probability in (2.5) when the variables Y_u^i for all u such that $|u-t(i)| < n^\alpha$ ($1 > \alpha > 0$) are excluded from consideration is at most

$$(2.6) \quad 1 - P\{Y_u^i \leq y \left(\frac{1-r_{u,t(i)}}{1+r_{u,t(i)}}\right)^{1/2}; |u-t(i)| < n^\alpha\} \\ \leq 1 - P\{\max_{|u-t(i)| < n^\alpha} Y_u^i \leq y \left(\frac{1-\delta}{1+\delta}\right)^{1/2}\}.$$

The last inequality follows because looking at (2.4) we see that only positive values of y need to be considered and by (1.1 ii)

$\frac{1-r_{u,t(i)}}{1+r_{u,t(i)}} \geq \frac{1-\delta}{1+\delta}$. The R.H.S. of (2.6) is at most $2n^\alpha (1-\Phi(y(\frac{1-\delta}{1+\delta})^{1/2}))$, where $\Phi(y) = \int_{-\infty}^y \phi(u)du$. Thus if we delete X_u for $|u-t(i)| < n^\alpha$ from consideration in (2.4) the absolute difference will be bounded above by

$$\begin{aligned} n' & \cdot 2n^\alpha \int_{\eta_{0,n}}^{\eta_{z,n}} (1-\Phi(y(\frac{1-\delta}{1+\delta})^{1/2})) \phi(y) dy \\ & \leq (\text{const.}) n^{1+\alpha} (1-\Phi(\eta_{0,n}(\frac{1-\delta}{1+\delta})^{1/2})) (1-\Phi(\eta_{0,n})) \\ & \leq (\text{const.}) \frac{n^{1+\alpha}}{\ell n n} e^{-\eta_{0,n}^2/(1+\delta)} \\ & \leq (\text{const.}) \exp\{(1+\alpha - \frac{2}{1+\delta})\ell n n + 2g(n)\} \end{aligned}$$

By choosing $\alpha < \frac{1-\delta}{1+\delta}$ and noticing that $\frac{g(n)}{\ell n n} \rightarrow 0$, we see that the R.H.S. above is $o(n^{-\beta})$ for some $\beta > 0$. Finally, for $|u-t(i)| > n^\alpha$,

$|r_{u,t(i)}| \leq \rho(n^\alpha)$ and $\frac{1-r_{u,t(i)}}{1+r_{u,t(i)}} \geq \frac{1-\rho(n^\alpha)}{1+\rho(n^\alpha)}$. Thus the quantity in (2.4) is almost

$$(2.7) \quad \sum_{i=1}^{n'} \int_{\eta_{0,n}}^{\eta_{z,n}} P\{\max_{u \in S_D^O} Y_u^i \leq y(\frac{1-\rho(n^\alpha)}{1+\rho(n^\alpha)})^{1/2} + \frac{x}{c_n}; \max_{v \in S_i^O} Y_v^i \leq y(\frac{1-\rho(n^\alpha)}{1+\rho(n^\alpha)})^{1/2}\} \phi(y) dy$$

apart from error terms of $o(n^{-\beta})$. We note that $S_D^O = S_D \cap \{|u-t(i)| > n^\alpha\}$

and $S_i^O = S_i \cap \{|u-t_i| > n^\alpha\}$. Define $CM_n^D = \max_{u \in S_D^O} Y_u^i$ and $cm_{n,k}^i = \max_{v \in S_i^O} Y_v^i$.

Since y is at least $\eta_{0,n} = O((\ell n n)^{1/2})$, we can write

$$y \left(\frac{1-p(n^\alpha)}{1+p(n^\alpha)} \right)^{1/2} = y + \frac{o(1)}{c_n} = y_2 \quad \text{say}$$

and let $y + \frac{x+o(1)}{c_n} = y_1$. By Berman ((1971) 4.2 pg. 931) we can write

the expression in (2.7) as

$$(2.8) \quad \sum_{i=1}^{n'} \int_{\eta_{0,n}}^{\eta_{z,n}} P\{CM_n^D \leq y_1\} P\{Cm_{n,k}^i \leq y_2\} \phi(y) dy \\ + \sum_{i=1}^{n'} \int_{\eta_{0,n}}^{\eta_{z,n}} E_{i,y} \phi(y) dy$$

where

$$(2.9) \quad |E_{i,y}| \leq \sum_{u=0}^n \sum_{v=0}^n \frac{|\gamma_{uv}^i|}{(1-\gamma_{uv}^{i2})^{1/2}} \exp\left\{-\frac{(y_1^2 - 2\gamma_{uv}^i y_1 y_2 + y_2^2)}{2(1-\gamma_{uv}^{i2})}\right\}$$

γ_{uv}^i is defined above (2.5). Also by Berman(1964, Lemma 3.1)

$$(2.10) \quad |P\{CM_n^D \leq y_1\} - \Phi^{n_1}(y_1)| \leq \sum_{u,v \in S_D^O} \frac{|\gamma_{uv}^i|}{(1-\gamma_{uv}^{i2})^{1/2}} e^{-y_1^2/(1+\gamma_{uv}^i)}$$

and

$$(2.11) \quad |P\{Cm_{n,k}^i \leq y_2\} - \Phi^{n_2}(y_2)| \leq \sum_{u,v \in S_i^O} \frac{|\gamma_{uv}^i|}{(1-\gamma_{uv}^{i2})^{1/2}} e^{-y_2^2/(1+\gamma_{uv}^i)}$$

Here $n_1 = |S_D^O|$ is cardinality of S_D^O and $n_2 = |S_i^O|$. Also $\frac{n_1}{n} \rightarrow \frac{k-1}{k}$

and $\frac{n_2}{n} \rightarrow \frac{1}{k}$. We notice that in the R.H.S. of (2.9), (2.10) and (2.11), y_1 and y_2 could be replaced by y since $y = O(\ln n)^{1/2}$. In the following we bound the sum in the R.H.S. of (2.9) by splitting it in two parts $|u-v| \leq n^\alpha$ and $|u-v| > n^\alpha$. It's easy to see that the same bounds will work for the R.H.S. of (2.10) and (2.11). Since for each i , $|u-t(i)| > n^\alpha$, $\gamma_{uv}^i \leq \frac{r_{uv} + \rho(n^\alpha)}{1 - \rho(n^\alpha)} = \gamma_{uv}$ say and $|\gamma_{uv}| \leq \delta < 1$ for large n as well as $|\gamma_{uv}| \leq \frac{\rho(u-v) + \rho(n^\alpha)}{1 - \rho(n^\alpha)}$. Thus we rewrite (2.8) as

$$(2.12) \quad \left[\frac{n}{k} \right] \int_{\eta_{0,n}}^{\eta_{z,n}} \phi^{n_1}(y_1) \phi^{n_2}(y_2) \phi(y) dy + E$$

where

$$(2.13) \quad \begin{aligned} |E| &\leq (\text{const.}) n \cdot \int_{\eta_{0,n}}^{\eta_{z,n}} \sum_{u=0}^n \sum_{v=0}^n |\gamma_{uv}| e^{-y^2/(1+\gamma_{uv})} \phi(y) dy \\ &\leq (\text{const.}) n \cdot \sum_{u=0}^n \sum_{v=0}^n |\gamma_{uv}| e^{-\eta_{0,n}^2/(1+\gamma_{uv})} (1 - \phi(\eta_{0,n})) \\ &\leq (\text{const.}) e^{g(n)} \left\{ \sum_{|u-v| \leq n^\alpha} |\gamma_{uv}| e^{-\eta_{0,n}^2/(1+\gamma_{uv})} \right. \\ &\quad \left. + \sum_{|v-u| > n^\alpha} |\gamma_{uv}| e^{-\eta_{0,n}^2/(1+\gamma_{uv})} \right\}. \end{aligned}$$

The first term in the R.H.S. is at most $n^{1+\alpha} \delta e^{-\eta_{0,n}^2/(1+\delta)}$ which is $o(n^{-\beta})$ for some $\beta > 0$ by the same argument as before. For $|u-v| > n^\alpha$, $|\gamma_{uv}| < 3\rho(n^\alpha)$ and the second sum in the R.H.S. of (2.13) is at most

$$\begin{aligned}
& (\text{const.}) n^2 \rho(n^\alpha) e^{-\eta_{0,n}^2 / (1 + 3\rho(n^\alpha))} \\
& \leq (\text{const.}) \rho(n^\alpha) \ln n e^{-2 \ln n \left(\frac{1}{1 + 3\rho(n^\alpha)} - 1 \right) + 2g(n)} \\
& = (\text{const.}) \rho(n^\alpha) \ln n \exp\{\rho(n^\alpha) \ln n + 2g(n)\}.
\end{aligned}$$

Substituting in (2.13), $|E| \rightarrow 0$ as $n \rightarrow \infty$ via choice of $g(n)$. It remains to show that the first term in (2.12) approaches the right limit as $n \rightarrow \infty$.

$$\begin{aligned}
(2.14) \quad & \left[\frac{n}{k} \right] \int_{\eta_{0,n}}^{\eta_{z,n}} \phi_1^{n_1}(y_1) \phi_2^{n_2}(y_2) \phi(y) dy \\
& \approx \frac{n}{k} \int_{-g(n)}^z \phi_1^{n_1}\left(b_{n,k} + \frac{x+y}{c_n}\right) \phi_2^{n_2}\left(b_{n,k} + \frac{y}{c_n}\right) \frac{\phi\left(b_{n,k} + \frac{y}{c_n}\right)}{c_n} dy.
\end{aligned}$$

Using the definition of $b_{n,k}$, we see that

$$\begin{aligned}
\phi_1^{n_1}\left(b_{n,k} + \frac{x+y}{c_n}\right) &= e^{n_1 \ln \phi\left(b_{n,k} + \frac{x+y}{c_n}\right)} \\
&\approx \exp \left\{ - \frac{n_1 \phi\left(b_{n,k} + \frac{x+y}{c_n}\right)}{b_{n,k} + \frac{x+y}{c_n}} \right\}
\end{aligned}$$

Thus for $-g(n) \leq y \leq z$,

$$\begin{aligned}
\phi_1^{n_1}\left(b_{n,k} + \frac{x+y}{c_n}\right) \exp\{(k-1)e^{-(x+y)}\} &\leq \exp\{(k-1)e^{-(x+y)} [1 - e^{-\frac{g(n) \ln n \ln n}{c_n}}]\} \\
&\quad + o((\ln n)^{-1/2}) \\
&\leq \exp\{(\text{const.}) \cdot \frac{g(n) \ln n \ln n}{c_n} e^{g(n)}\}
\end{aligned}$$

The R.H.S. above tends to 1 as $n \rightarrow \infty$ by choice of $g(n)$. Similar calculations show that

$$\left(\frac{n}{k}\right)^{n_1} \left(b_{n,k} + \frac{x+y}{c_n}\right)^{n_2} \frac{\varphi(b_{n,k} + \frac{y}{c_n})}{c_n} (e^{-(k-1)e^{-(x+y)}} e^{-e^{-y}} e^{-y})^{-1}$$

tends to 1 as $n \rightarrow \infty$ uniformly in y for $-g(n) \leq y \leq z$. Also the second term above is integrable in y . Applying dominated convergence theorem we see that the L.H.S. of (2.14) tends to

$$\int_{-\infty}^z e^{-(1+(k-1)e^{-x})e^{-y}} e^{-y} dy = \frac{1}{1+(k-1)e^{-x}} \int_{(1+(k-1)e^{-x})e^{-z}}^{\infty} e^{-u} du$$

$$\frac{\exp\{-e^{-z}(1+(k-1)e^{-x})\}}{1+(k-1)e^{-x}}$$

Hence the result.

Remark:- The rate of convergence of (2.1) can be found in the above discussion. We note that via choice of the function $g(n)$ and in showing the second term in (2.13) is $o(1)$ the rate of convergence depends on the function $\rho(n)$ (and consequently on r_{ij}). If it can be assumed that $\rho(n)(\ln n)^\alpha \rightarrow 0$ for some $0 < \alpha < 1$ then the rate of convergence could be shown to be $o((\ln n)^{-\beta})$ for some $\beta > 0$. This is quite slow for practical calculation.

3. STRONGLY DEPENDENT SEQUENCES:- In this section we assume throughout that $\{X_n, n \geq 0\}$ is a standard stationary Gaussian sequence with $r_n \equiv EX_i X_{i+n}$. We assume that (1.2) holds. The aim (realized in Theorem 3.1)

will be to show that the same limiting distribution as in (2.1) holds for $M_n - m_{n,k}$ under these conditions where M_n and $m_{n,k}$ are the same as before. (However definitions of various other quantities are different in this section from the last and should be carefully noted.)

As was seen in Mittal and Ylvisaker (1975), the limit distribution of the normalized maxima will now change due to the strong interdependence of the variables. The fact that such dependence plays no part in the distribution considered in Theorem 2.1 seems odd at first but could be explained if we look at the representation of the sequence considered in [3]. We can write for $0 \leq i \leq n$

$$X_i = (1-r_n)^{1/2} \xi_i + (r_n)^{1/2} U$$

where $\{\xi_i, 0 \leq i \leq n\}$ are standard normal with covariance sequence

$r'_i = \frac{r_i - r_n}{1 - r_n}$ and U is again standard normal independent of $\{\xi_i, 0 \leq i \leq n\}$.

(The dependence of the variables ξ_i and r'_i on n is suppressed for the sake of simpler notation.) Thus

$$M_n = (1-r_n)^{1/2} \max_{0 \leq i \leq n} \xi_i + (r_n)^{1/2} U = (1-r_n)^{1/2} M'_n + (r_n)^{1/2} U \text{ say,}$$

$$m_{n,k} = (1-r_n)^{1/2} \max_{i \in G_n} \xi_i + (r_n)^{1/2} U = (1-r_n)^{1/2} m'_{n,k} + (r_n)^{1/2} U \text{ say}$$

and

$$(3.1) \quad C_n(M_n - m_{n,k}) = C_n(1-r_n)^{1/2} (M'_n - m'_{n,k}),$$

the part $(r_n)^{1/2} U$ which makes the process strongly dependent cancels out and allows us to see the same behavior as in the last section for the

difference of the two maxima. Eventhough the ideas are the same the proofs are longer and more involved due to the bulky nature of the covariance function.

The following theorem is stated somewhat differently than Theorem 2.1 because the second order terms of $m_{n,k}$ are not comparable with the difference $M_n - m_{n,k}$.

Theorem 3.1:- If (1.2) holds for $\{X_n, n \geq 0\}$ as described in the beginning of this section then

$$(3.2) \quad P\{M_n - m_{n,k} \leq \frac{x}{c_n}\} \rightarrow \frac{1}{1+(k-1)e^{-x}},$$

as $n \rightarrow \infty$ for all $0 \leq x < \infty$.

Proof:- Looking at (3.1), it is sufficient to show that $P\{M'_n - m'_{n,k} \leq \frac{x}{c_n(1-r_n)^{1/2}}\}$ converges to the R.H.S. of (3.2). Let us define $\eta_{0,n} = b_{n,k} - g(n)/c_n$ where $g(n) \rightarrow \infty$ is chosen such that $(r_n)^{1/4} e^{2g(n)} = o(1)$. (Notice the different definitions of $\eta_{0,n}$ and $g(n)$). In order to follow the proof of Section 2, we first have to show that $P\{m'_{n,k} < \eta_{0,n}\} = o(1)$. The proof of this follows that of (2.17) in [3]. Define

$$(3.3) \quad \begin{aligned} \tau(-1) &\equiv 0 \\ \tau(0) &= [n^\alpha] \quad \text{for } 0 < \alpha < \frac{1-\delta}{(1+\delta)^2} \\ \tau(i) &= n \exp\{-r_n^{1/2} i\} \quad i=1,2,\dots,q(n) \\ \tau(q(n)+1) &= n \end{aligned}$$

where $q(n)$ is chosen so that

$$(3.4) \quad r_n^{\frac{q(n)+1}{2}} > \frac{(r_n)^{1/4}}{\ell n n} \geq r_n^{\frac{q(n)+2}{2}}.$$

As illustrated in [3], (1.2)(ii) implies that

$$r_{\tau(j)}^{-r_n} \leq r_n^{\frac{j+2}{2}}.$$

By Berman's Lemma [1]

$$(3.5) \quad \begin{aligned} P\{m'_{n,k} < \eta_{0,n}\} &\leq \Phi^{n'}(\eta_{0,n}) + (\text{const.}) \sum_i \sum_j r'(i-j) \exp\left\{-\frac{\eta_{0,n}^2}{1+r'(i-j)}\right\} \\ &\leq \Phi^{n'}(\eta_{0,n}) + (\text{const.}) n \sum_{i=0}^{q(n)} \tau(i+1) r'_{\tau(i)} \exp\left\{-\frac{\eta_{0,n}^2}{1+r'_{\tau(i)}}\right\}. \end{aligned}$$

Direct computations will show that the first term of the R.H.S. of (3.5) is $o(1)$ for any $g(n) \rightarrow \infty$ and that the second term is $o(1)$ can be seen by some modification of the procedure in [3]. The same method is also used in (3.14). Hence we exclude the details at this stage. Thus it is sufficient to show that

$$(3.6) \quad P\{M'_n - m'_{n,k} \leq \frac{x}{c_n(1-r_n)^{1/2}}; m'_{n,k} > \eta_{0,n}\}$$

tends to the desired limit in (3.2) as $n \rightarrow \infty$. Conditioning on $\xi_{t_i} \in G_n$ and proceeding as in Section 2, the quantity in (3.6) is equal to

$$(3.7) \quad \sum_{i=1}^{n'} \int_{\eta_{0,n}}^{\infty} P\{Y_u^i \leq y(\frac{1-r'_{u,t(i)}}{1+r'_{u,t(i)}})^{1/2} + \frac{x}{c_n(1-r_n)^{1/2}} \mid u \in S_D\},$$

$$Y_v^i \leq y(\frac{1-r'_{v,t(i)}}{1+r'_{v,t(i)}})^{1/2} \mid v \in S_i\} \varphi(y) dy$$

where $\{Y_u^i, 0 \leq u \leq n\}$ are standard Gaussian with

$$E Y_u^i Y_v^i = \frac{r'_{u,v} r'_{u,t(i)} r'_{v,t(i)}}{(1-r'^2_{u,t(i)})^{1/2} (1-r'^2_{v,t(i)})^{1/2}} = \gamma_{uv}^i \text{ say.}$$

By Berman (1971, (4.5), pg. 932) we can write (3.7) as

$$(3.8) \quad \sum_{i=1}^{n'} \int_{\eta_{0,n}}^{\infty} \prod_{u \in S_D} \left(\Phi(y(\frac{1-r'_{u,t(i)}}{1+r'_{u,t(i)}})^{1/2} + \frac{x}{c_n(1-r_n)^{1/2}}) \prod_{v \in S_i} \left(\Phi(y(\frac{1-r'_{v,t(i)}}{1+r'_{v,t(i)}})^{1/2}) \right) \varphi(y) dy \right. \\ \left. + \sum_{i=1}^{n'} \int_{\eta_{0,n}}^{\infty} E_{i,y} \varphi(y) dy \right)$$

where the error $E_{i,y}$ is such that

$$(3.9) \quad |E_{i,y}| \leq \sum_u \sum_v \int_0^{\gamma_{uv}^i} \varphi(y_1, y_2, \lambda) d\lambda$$

and

$$\varphi(y_1, y_2, \lambda) = (2\pi)^{-1} (1-\lambda^2)^{-1/2} \exp\left\{-\frac{y_1^2 - 2\lambda y_1 y_2 + y_2^2}{2(1-\lambda^2)}\right\}.$$

We know that y_1 is either $y_u = y(\frac{1-r'_{u,t(i)}}{1+r'_{u,t(i)}})^{1/2}$ or $y_u + \frac{x}{c_n(1-r_n)^{1/2}}$

and y_2 is either $y_v = y \left(\frac{1-r'_{v,t(i)}}{1+r'_{v,t(i)}} \right)^{1/2}$ or $y_v + \frac{x}{c_n(1-r_n)^{1/2}}$. Noticing that $|\gamma_{uv}^i| \leq \delta < 1$ for large n for each i , we can see that

$$(3.10) \quad |E_{i,y}| \leq (\text{const.}) \sum_u \sum_v |\gamma_{uv}^i| \exp\left\{-\frac{y_u^2 - 2|\gamma_{uv}^i| y_u y_v + y_v^2}{2(1-\gamma_{uv}^i)^2}\right\}$$

$$= (\text{const.}) \sum_u \sum_v h(u,v) \quad \text{say.}$$

Define

$$\Delta_0 = \sum_{0 < |u-t(i)| \leq \tau(0)} \sum_v h(u,v) = \sum_u \sum_{0 < |v-t(i)| \leq \tau(0)} h(u,v).$$

The last equality follows by symmetry. We notice that the sum of $h(u,v)$ over the set $\{|u-t(i)| \leq \tau(0) \text{ or } |v-t(i)| \leq \tau(0)\}$ is at most $2\Delta_0$.

Also define

$$\Delta_j = \text{sum of } h(u,v) \text{ on set } \{\tau(j-1) < |u-t(i)| \leq \tau(j)\} \\ \cap \{\tau(j-1) < |v-t(i)| \leq \tau(j)\}$$

$j=1,2,\dots,(q(n)+1)$. Then by the same argument as above the sum in the

R.H.S. of (3.10) is at most $2 \sum_{j=0}^{q(n)+1} \Delta_j$. Next we will estimate Δ_j ,

$j=0,1,\dots,(q(n)+1)$ for $y > \eta_{0,n}$. First

$$(3.11) \quad \Delta_0 = \sum_{|u-t(i)| \leq \tau(0)} \sum_{|v-t(i)| \leq 2\tau(0)} h(u,v) + \sum_{|u-t(i)| \leq \tau(0)} \sum_{|v-t(i)| > 2\tau(0)} h(u,v).$$

To bound the exponent in $h(u,v)$,

$$\begin{aligned}
& \frac{y_u^2 - 2|\gamma_{uv}^i| y_u y_v + y_v^2}{2(1-\gamma_{uv}^2)} \\
&= y^2 \left\{ \frac{1}{2} (1-\gamma_{uv}^2)^{-1} \left[\left(\frac{1-r'_{u,t(i)}}{1+r'_{u,t(i)}} \right)^{1/2} - \left(\frac{1-r'_{v,t(i)}}{1+r'_{v,t(i)}} \right)^{1/2} \right]^2 \right. \\
&\quad \left. + (1+|\gamma_{uv}^i|)^{-1} \left(\frac{1-r'_{u,t(i)}}{1+r'_{u,t(i)}} \right)^{1/2} \left(\frac{1-r'_{v,t(i)}}{1+r'_{v,t(i)}} \right)^{1/2} \right\} \\
(3.12) \quad & \geq \frac{y^2}{1+|\gamma_{uv}^i|} \left(\frac{1-r'_{u,t(i)}}{1+r'_{u,t(i)}} \right)^{1/2} \left(\frac{1-r'_{v,t(i)}}{1+r'_{v,t(i)}} \right)^{1/2}.
\end{aligned}$$

We will use this bound for all sums Δ_j except the second term in the R.H.S. of (3.11). For this term notice that when $|u-t(i)| \leq \tau(0)$; $|v-t(i)| > 2\tau(0)$ we have $|u-v| > \tau(0)$ and $|\gamma_{uv}^i| \leq (\text{const.})r'(n^\alpha)$. Thus the L.H.S. of (3.12) is at least

$$\frac{y^2}{2} \left(\frac{1-\delta}{1+\delta} + \frac{1-r'(n^\alpha)}{1+r'(n^\alpha)} - (\text{const.})r'(n^\alpha) \right).$$

Substituting, we get

$$\begin{aligned}
\Delta_0 &\leq (\text{const.}) \left\{ \tau^2(0) \exp(-y^2 \frac{(1-\delta)}{(1+\delta)^2}) + n\tau(0) \exp(-y^2 (\frac{1}{1+\delta} - (\text{const.})r'(n^\alpha))) \right. \\
&\leq (\text{const.}) e^{2g(n)} \left\{ \exp(2\ell \ln(\alpha - \frac{1-\delta}{(1+\delta)^2})) + \exp(\ell \ln(\alpha + (\text{const.})r'(n^\alpha) - \frac{1-\delta}{1+\delta})) \right\}.
\end{aligned}$$

Since $r'(n^\alpha) = o(1)$, by choosing $\alpha < \frac{1-\delta}{(1+\delta)^2} - \beta$ for some $\beta > 0$ we see that the R.H.S. above is $o(n^{-\beta})$. Now

$$(3.13) \quad \Delta_j \leq (\text{const.}) \sum_{\tau(j-1) < |u-t(i)|; |v-t(i)|} \sum_{\tau(j)} |\gamma_{uv}^i| \exp\left\{-\frac{y^2}{1+\gamma_{uv}^i} \left(\frac{1-r'_{\tau(j-1)}}{1+r'_{\tau(j-1)}}\right)\right\}.$$

Before proceeding to find upper bounds for Δ_j let us notice that

$$\gamma_{uv}^i \leq r'_{uv} (1-r'_{u,t(i)})^{-1/2} (1-r'_{v,t(i)})^{-1/2} \leq 2(r_{u,v} - r_n)$$

for large n if $|u-t(i)| > \tau(0)$ and $|v-t(i)| > \tau(0)$. In the following discussion we will bound $2(r_{u,v} - r_n)$ above and use the same bound for $|\gamma_{uv}^i|$. This is justified because of the following. We will bound γ_{uv}^i below by say d and then show that $|d|$ is smaller than the smallest upper bound we use for $(r_{u,v} - r_n)$. Now $\gamma_{uv}^i \geq r'_{u,v} - r'_{u,t(i)} r'_{v,t(i)}$. Suppose $|u-t(i)| = a$ and $|v-t(i)| = b$ then the lowest value of $r'_{u,v}$ will be when $|u-v| = a+b$. Suppose $a \leq b$ then $b \geq \frac{a+b}{2}$. Thus

$$\begin{aligned} r'_{u,v} - r'_{u,t(i)} r'_{v,t(i)} &\geq r'_{a+b} - r_n \alpha^{\frac{r'_{a+b}}{2}} \\ &= (r_{a+b} - r_n - r_n \alpha^{\frac{r_{a+b} - r_n}{2}}) (1 - r_n)^{-1}. \end{aligned}$$

But $r_{\frac{a+b}{2}} \leq r_{a+b} (1 + \frac{\ln 2}{\ln \frac{a+b}{2}})$ because of condition (1.2)(ii). Thus the

R.H.S. above is at least

$$(1-r_n)^{-n} \left[(r_{a+b} - r_n)(1-r_n^\alpha) - \frac{r_n^\alpha r_{a+b} \ell n 2}{\ell n \frac{a+b}{2}} \right] \\ \geq - \frac{r_n^\alpha r_{a+b} \ell n 2}{(\ell n \frac{a+b}{2})(1-r_n)}.$$

We will see later that the lowest upper bound used for $r_{uv} - r_n$ is

$$\frac{q(n)+2}{r_n^2}. \text{ But}$$

$$\frac{q(n)+2}{r_n^2} - \frac{r_n^\alpha r_{a+b} \ell n 2}{(\ell n \frac{a+b}{2})(1-r_n)} = \frac{q(n)+2}{r_n^2} (1 - (\text{const.})(r_n^{\frac{q(n)-2}{2}} \ell n n)^{-1}).$$

By choice of $q(n)$ as in (3.4) we see that $\frac{q(n)-2}{r_n^2} \ell n n \rightarrow \infty$ as $n \rightarrow \infty$. Returning to (3.13), for $y \geq \eta_{0,n}$

$$(3.14) \quad (\text{const.}) \Delta_j \leq \tau(j) \sum_{|v-t(i)| > \tau(j-1)} |\gamma_{uv}^i| \exp\left\{-\frac{\eta_{0,n}^2}{1+|\gamma_{uv}^i|} (1-2r'_{\tau(j-1)})\right\} \\ \leq \tau(j) \left\{ 8 \exp\left(-\frac{\eta_{0,n}^2}{1+\delta} (1-2r'_{\tau(j-1)})\right) \right. \\ \left. + \sum_{\ell=1}^{j-1} \tau(\ell) r'_{\tau(\ell-1)} \exp\left(-\frac{\eta_{0,n}^2}{1+2r'_{\tau(\ell-1)}} (1-2r'_{\tau(\ell-1)})\right) \right. \\ \left. + n r'_{\tau(j-1)} \exp(-\eta_{0,n}^2 (1-4r'_{\tau(j-1)})) \right\}.$$

The first term above is at most

$$(\text{const.}) \ell n n \exp\{2g(n) - \ell n n \left(\frac{1-\delta}{1+\delta} - 4r' \tau(j-1) + r_n^{j/2}\right)\} = O(n^{-\beta})$$

for some $\beta > 0$. The last term is at most

$$\begin{aligned} & (\text{const.}) r' \tau(j-1) \ell n n \exp\{2g(n) - r_n^{j/2} \ell n n \left(1 - \frac{8r' \tau(j-1)}{r_n^{j/2}}\right)\} \\ & \leq (\text{const.}) r_n^{\frac{j+1}{2}} \ell n n \exp\left\{-\frac{r_n^{j/2} \ell n n}{2}\right\} \end{aligned}$$

because of the bound shown right after (3.4). Lastly, the middle term in the R.H.S. of (3.14) is at most

$$\begin{aligned} & (\text{const.}) e^{2g(n)} \sum_{\ell=1}^{j-1} r_n^{\frac{\ell+1}{2}} \ell n n \exp\{-\ell n n (r_n^{\ell/2} + r_n^{j/2} - 4r' (\ell-1)^{-8r'} (j-1))\} \\ & \leq (\text{const.}) e^{2g(n)} \sum_{\ell=1}^{j-1} r_n^{\frac{\ell+1}{2}} \ell n n \exp\left\{-\frac{r_n^{\ell/2} \ell n n}{2}\right\} \end{aligned}$$

since $r' \tau(\ell-1) \leq 2r_n^{\frac{\ell+1}{2}}$ and $\ell \leq j-1$. Substituting in (3.14) we see that by choice of $g(n)$

$$\begin{aligned} \Delta_j & \leq (\text{const.}) \{O(n^{-\beta}) + \sum_{\ell=1}^j r_n^{\frac{\ell+1}{2}} \ell n n \exp\left(-\frac{r_n^{\ell/2}}{2} \ell n n\right)\} \\ & \leq (\text{const.}) \{O(n^{-\beta}) + \sum_{\ell=1}^j \exp\left(-\frac{r_n^{\ell/2} \ell n n}{4}\right)\} \end{aligned}$$

since

$$-\frac{r_n^{\ell/2} \ell n n}{2} + \ell n (r_n^{\ell+1/2} \ell n n) \leq -\frac{r_n^{\ell/2} \ell n n}{2} + 2\ell n (r_n^{\ell/2} \ell n n)$$

and $\ell \leq j \leq q(n)$ implies $r_n^{\ell/2} \ell n n \rightarrow \infty$ as $n \rightarrow \infty$. Thus the sum in the R.H.S. above is at most

$$\sum_{\ell=1}^j \frac{4}{r_n^{\ell/2} \ell n n}.$$

Choice of $q(n)$ implies $(\sqrt{r_n})^\ell \ell n n > (\sqrt{r_n})^{-(q(n)-\ell+\frac{1}{2})}$ and the upper bound for the above will be

$$\begin{aligned} 4 \sum_{\ell=1}^j (\sqrt{r_n})^{(q(n)-\ell+\frac{1}{2})} &= 4 (\sqrt{r_n})^{q(n)-j+\frac{1}{2}} \sum_{\ell=0}^{j-1} (\sqrt{r_n})^\ell \\ &\leq 8 (\sqrt{r_n})^{q(n)-j+\frac{1}{2}}. \end{aligned}$$

Thus

$$\sum_{j=1}^{q(n)} \Delta_j \leq (\text{const.}) \sum_{j=1}^{q(n)} (\sqrt{r_n})^{q(n)-j+\frac{1}{2}} \leq (\text{const.}) (r_n)^{1/4}.$$

For the last sum $\Delta_{q(n)+1}$ we proceed as in (3.14). The only term that needs attention is the last in the R.H.S. of (3.14). Similar procedures will show that it is at most

$$(\text{const.}) r_n^{\frac{q(n)+2}{2}} \ell n n \exp\{-r_n^{\frac{q(n)+1}{2}} \ell n n\} \leq (\text{const.}) (r_n)^{1/4}$$

by choice of $q(n)$. Thus

$$\sum_{i=1}^{n'} \int_{\eta_{0,n}}^{\infty} E_{i,y} \varphi(y) dy \leq (\text{const.}) (r_n)^{1/4} \exp(2g(n))$$

which by choice of $g(n)$ is $o(1)$. It remains to be seen that the first term in (3.8) approaches the right limit as $n \rightarrow \infty$.

Let us write $y = b_{n,k} + \frac{w}{c_n}$ and look at

$$\prod_{u \in S_D} \phi(b_{n,k} (\frac{1-r'_{u,t(i)}}{1+r'_{u,t(i)}})^{1/2} + \frac{x+w+o(1)}{c_n})$$

$$\approx \exp\{- \sum_{u \in S_D} \frac{\phi(b_{n,k} (\frac{1-r'_{u,t(i)}}{1+r'_{u,t(i)}})^{1/2} + \frac{x+w+o(1)}{c_n})}{b_{n,k} (\frac{1-r'_{u,t(i)}}{1+r'_{u,t(i)}})^{1/2} + \frac{x+w+o(1)}{c_n}}\}.$$

We will split the sum in the exponent above as usual and see that it is approximately equal to

$$(3.15) \quad \sum_{j=0}^{q(n)+1} \sum_{\substack{\tau(j-1) \leq |u-t(i)| < \tau(j) \\ u \in S_D}} (\ell n n)^{-1/2} \phi(b_{n,k} (\frac{1-r'_{u,t(i)}}{1+r'_{u,t(i)}})^{1/2} + \frac{x+w}{c_n}).$$

for $-g(n) \leq w < \infty$. The above sums for $0 \leq j \leq q(n)$ tend to zero and

for $j=q(n)+1$ it tends to $(k-1)e^{-(x+w)}$. The proof of this is very

much similar to that of (3.14). In fact we can see that the first two

terms of the R.H.S. of (3.14) could be taken as the upper bound for

$q(n)$

$\sum_{j=0}^{q(n)+1}$ in (3.15). The sum for $j=q(n)+1$ is taken over all $u \in S_D$ such

that $\tau(q(n)) \leq |u-t(i)| < n$. It is easy to see that the number of terms

in the sum are asymptotically equal to $\frac{(k-1)n}{k}$. Thus the sum for

$j=q(n)+1$ is about

$$\frac{(k-1)n}{k(\ell n n)^{1/2}} \exp\{-\frac{b_{n,k}^2}{2} - (x+w) + o(1)\}$$

which tends to $(k-1)\exp(-(x+w))$. Similarly we can show that

$\prod_{v \in S_i} \phi(b_{n,k} \frac{(1-r_{v,t(i)})^{1/2}}{(1+r_{v,t(i)})} + \frac{w}{c_n})$ tends to $\exp(-e^{-w})$. We follow the

same steps as in the end of Section 2 to complete the proof of

Theorem 3.1. Notice that the additional condition required there

namely $\frac{\ell n \ell n n}{(\ell n n)^{1/2}} \exp(2g(n)) = o(1)$ is automatically satisfied here

since $\frac{\ell n \ell n n}{(\ell n n)^{1/2} (r_n)^{1/4}} = o(1)$ in view of (1.2)(ii), and the choice

of $g(n)$.

References

- [1] Berman, S.M. (1964). Limit theorems for the maximum term in stationary sequences. Ann. Math. Stat. 35 502-516.
- [2] Berman, S.M. (1971). Asymptotic independence of the numbers of high and low level crossings of stationary Gaussian processes. Ann. Math. Stat. 42 927-945.
- [3] Mittal, Yash and Ylvisaker, Donald (1975). Limit distributions for the maxima of stationary Gaussian processes. Stochastic Processes and Appl. 3 1-18.