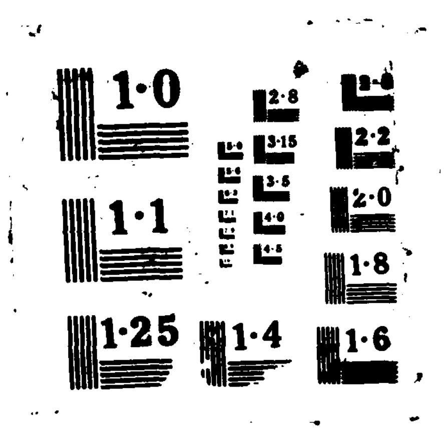
AD-A185 171 QUANTILE BASED UNIFIED DISTRIBUTION OF EXTREME VALUES 1/1
AND ORDER STATISTIC..(U) TEXAS A AND M UNIV COLLEGE
STATION DEPT OF STATISTICS E PARZEN JUN 87 TR-Q-2
UNCLASSIFIED ARO-23010.2-MA DAAL03-87-K-0003 F/G 12/3 NL

END


1·0
1·1
1·25
1·4
1·6
1·8
2·0
2·2
2·8
3·15
3·5
4·0
4·5

ARO 23010.2-MA

DTIC FILE COPY

TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843-3143

STATISTICS
Phone 409 - 845-3141

AD-A185 171

QUANTILE BASED UNIFIED DISTRIBUTION OF EXTREME VALUES AND ORDER STATISTICS

Emanuel Parzen

Department of Statistics
Texas A&M University

Technical Report No. Q-2

June 1987

Texas A&M Research Foundation
Project No. 5641

'Functional Statistical Data Analysis and Modeling'

Sponsored by the U. S. Army Research Office

Professor Emanuel Parzen, Principal Investigator

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARO 23010.2-MA	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Quantile Based Unified Distribution of Extreme Values and Order Statistics		5. TYPE OF REPORT & PERIOD COVERED Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Emanuel Parzen		8. CONTRACT OR GRANT NUMBER(s) DAAL03-87-K-0003
9. PERFORMING ORGANIZATION NAME AND ADDRESS Texas A&M University Department of Statistics College Station, TX 77843		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE June 1987
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Quantile functions, order statistics, extreme value, power-uniform distributions, power-gamma distributions, hazard quantile function, tail exponents, stationary time series.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This paper presents ideas leading to unified formulas for exact and asymptotic distributions of central and extreme order statistics of random samples and stationary time series X(t) in terms of quanitle functions Q(u), density quantile functions fQ(u), left and right hazard quantile functions hQ(u), defined to equal respectively fQ(u)/u and fQ(u)/(1-u), tail exponents, and the spectral density near zero frequency of a two-valued time series c(X(t);u), equal to 1-u or -u as $X(t) \leq Q(u)$ or $X(t) > Q(u)$.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

QUANTILE BASED UNIFIED DISTRIBUTION OF EXTREME VALUES AND ORDER STATISTICS

Emanuel Parzen[1]
Texas A&M University

Abstract. This paper presents ideas leading to unified formulas for exact and asymptotic distributions of central and extreme order statistics of random samples and stationary time series $X(t)$ in terms of quantile functions $Q(u)$, density quantile functions $fQ(u)$, left and right hazard quantile functions $hQ(u)$, defined to equal respectively $fQ(u)/u$ and $fQ(u)/(1-u)$, tail exponents, and the spectral density near zero frequency of a two-valued time series $c(X(t);u)$, equal to $1-u$ or $-u$ as $X(t) \leq Q(u)$ or $X(t) > Q(u)$.

1. Introduction.

1.0. Quantile function concepts are reviewed (Section 2). Section 3 reviews important sampling distributions. Sections 4 and 5 outline the asymptotic distribution theory of extreme and central order statistics. The unified formula is presented in Section 6; comparisons of exact and approximate distributions are made numerically. Section 8 outlines extensions to stationary time series.

1.1. A random sample $X_1, X_2, \ldots, X_n$ consists of independent random variables identically distributed as a random variable X. Its order statistics, denoted $X(1;n) \leq \cdots \leq X(k;n) \leq \cdots \leq X(n;n)$ are the values in the sample arranged in increasing order. The sample minimum and maximum are respectively equal to $X(1;n)$ and $X(n;n)$.

1.2. Extreme value theory initially seeks to identify the distributions which can arise as the distributions of $X(1;n)$ and $X(n;n)$ for large n (in the limit as $n \to \infty$). It is motivated by the hope that these distributions will provide models for variables observed in reliability studies. We present the theory of extreme value distributions in a way that broadens our understanding of the models to fit to data.

1.3. Extreme value theory does not seem to be routinely studied by statisticians. The literature on extreme value theory is extremely large and does not seem to be easy to apply. This paper proposes that the quantile function descriptions of probability laws can be used to develop a unified theory of the distribution of order statistics in large samples which is easier to apply (and even to teach in introductory courses).

1.4. In the quantile based approach a central role is played by the order statistics $U(1;n) \leq \cdots \leq U(n;n)$ of a random sample of a random variable U which is Uniform $[0,1]$. Their exact and asymptotic distributions are easily found.

1.5. A philosophical note: Our interest in order statistics is motivated in part by: (1) a conjecture that statistical inference is made possible by the fact that order statistics possess asymptotic distributions whose parameters are the quantile function and the density

[1] Research Sponsored by U. S. Army Research Office Project DAAL03-87-K-0003.

quantile function, and (2) a 'definition' of statistics: 'Arithmetic done by ranking before adding'.

How should one find the average $\bar{X}$ of $X(1), \cdots, X(n)$; the conventional answer is add and divide by n:

$$\bar{X} = \frac{1}{n} \sum_{t=1}^{n} X_t.$$

Our answer is: rank (form order statistics) and then add to form

$$\bar{X} = \frac{1}{n} \sum_{k-1}^{n} X(k;n).$$

We argue that ranking before adding to find $\bar{X}$ is better in both practice and theory. In practice the ranked values can identify the types of distribution obeyed by the sample. In theory, to study the asymptotic distribution of $\bar{X}$, one has to determine the contribution of the extreme values, and in particular if their distribution will be such as to make the distribution of $\bar{X}$ non-normal.

1.6 The goal of this paper is to state a unified formula for distribution of extreme and central order statistics. It can be summarized, using the notation in Table 1. For $0 < u_{k;n} \leq 0.5$, approximately,

$$fQ_{X(k;n)}(u) = \{\frac{n+1}{u_{k;n}(1-u_{k;n})}\}^{0.5} fQ(u_{k;n}) \phi \Phi^{-1}(u)$$
$$\times \{1 + \frac{1}{3\sqrt{m_0}} \Phi^{-1}(u) - \frac{1}{9m_0}\}^{3\alpha_0 - 2}.$$

For $0.5 \leq u_{k;n} < 1$, approximately,

$$fQ_{X(k;n)}(u) = \{\frac{n+1}{u_{k;n}(1-u_{k;n})}\}^{0.5} \quad fQ(u_{k;n}) \phi \Phi^{-1}(u)$$
$$\{1 + \frac{1}{3\sqrt{m_1}} \Phi^{-1}(1-u) - \frac{1}{9m_1}\}^{3\alpha_1 - 2}.$$

The shape of the distribution of order statistics can be determined from graphs of the approximate formulas for their density quantile functions. To apply these results in practice one has only to determine the values of α_0, α_1, m_0, m_1. It should be noted that these approximations are not valid in the extremes (as u tends to 0 or 1) because their tail exponents are not exact. The scale parameters for extreme values of these approximating distributions are the reciprocals of the hazard factors in Table 2.

Table 1: Table of Notation

Notation	Meaning
$F(x)$	Distribution Function
$f(x) = F'(x)$	Probability Density Function
$X_1, \ldots, X_n$	Random Sample
$X(k;n), X(n+k-1;n)$	Extreme order statistic if k fixed; central order statistic if $k \to \infty$
$u_{k;n} = k/(n+1)$	as $n \to \infty$ and $k/n \to p > 0$
$U(k,n)$	Order statistic from Uniform $[0,1]$
$Q(u) = F^{-1}(u), fQ(u) = f(Q(u))$	Quantile, density quantile function
$Q(\text{Beta}(k, n+1-k))$	Exact Distribution of $X(k;n)$
$(n+1)U(k;n) \to \text{Gamma}(k)$	Asymptotic Distribution of $U(k;n)$
$\text{GammaA}(m) = \frac{1}{m}\text{Gamma}(m)$ $= (1 + \frac{1}{3\sqrt{m}}N(0,1) - \frac{1}{9m})^3$	Wilson-Hilferty Approximation
$g(x;\lambda) = \frac{x^\lambda - 1}{\lambda}, \lambda < 0 \text{ or } \lambda > 0$ $= \log x, \quad \lambda = 0$	Power (Box-Cox) transformation
$g(\text{GammaA}(m);\lambda)$	PowerG(λ, m) distributed variable
$X(k;n) = Q(u_{k;n}) + W(k;n)$	$X(k;n)$ estimates $Q(u_{k;n})$
Asymptotic distribution of $X(k;n)$ $W(k;n) = \{h_0 Q(u_{k;n}\})^{-1} g(\text{GammaA}(m_0); 1-\alpha_0)$	$0 < u_{k;n} \leq 0.5$
$W(k;n) = \{h_1 Q(u_{k;n})\}^{-1} g(\text{GammaA}(m_1); 1-\alpha_1)(-1)$	$0.5 \leq u_{k;n} < 1$
$h_0 Q(u) = fQ(u)/u$	left hazard quantile function
$h_1 Q(u) = fQ(u)/(1-u)$	right hazard quantile function
$fQ(u) = u^{\alpha_0} L_0(u)$	α_0 left tail exponent
$L_0(u) \sim (\log u)^\beta$	$L_0(u)$ slowly varying, log-like
$fQ(1-u) = (1-u)^{\alpha_1} L_1(u)$	α_1 right tail exponent
$L_1(u) \sim ((\log(1-u)^{-1})^\beta$	$L_1(u)$ slowly varying, log -like
$m_0 = (n+1)u_{k;n}/(1-u_{k;n})$ $m_1 = (n+1)(1-u_{k;n})/u_{k;n}$	

Table 2: Examples of Powers of n in Hazard Factors

$1-\alpha_0, 1-\alpha_1$		$h_0Q(1/n), h_1Q(1-(1/n))$
1	Uniform, Exponential (left tail)	n
.5	Weibull (left tail) $Q(u) = (\log(1-u)^{-1})^{\beta},\ \beta = 0.5$	$n^{\beta} = n^{0.5}$
0	Weibull (right tail)	$\beta(\log n)^{0.5}$
0	Normal	$(2\log n)^{0.5}$
0	Exponential (right tail)	1
0	One-sided stable, index .5 (left tail)	$0.5(\log n)^2$
$-.5$	Infinite variance Pareto (right tail)	$n^{-0.5}$
-1	Cauchy	πn^{-1}
-2	One-sided stable, index .5 (right tail) $Q(u) = \{\Phi^{-1}(1-\frac{u}{2})\}^{-2}$	$(\pi/2)n^{-2}$

2. Quantile Functions Concepts.

2.1. The probability distribution of a random variable is usually described by the distribution function $F(x)$, $-\infty < x < \infty$, defined by

$$F(x) = \Pr[X \le x].$$

The quantile function, denoted $Q(u)$, $0 \le u \le 1$, or $F^{-1}(u)$, $0 \le u \le 1$, is defined by

$$Q(u) = F^{-1}(u) = \inf(x : F(x) \ge u).$$

2.2. **Theorem:** Inequality Inverse Functions. $F(x)$ and $Q(u)$ are inverse of each other under inequalities, in the sense that (for any x and u)

$$F(x) \ge u \text{ if and only if } x \ge Q(u).$$

2.3. Examples of Probability laws are:

1) Standard normal: $F(x) = \Phi(x), Q(u) = \Phi^{-1}(u)$, where

$$\Phi(x) = \int_{-\infty}^{x} \phi(y)\,dy, \qquad \phi(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2).$$

2) uniform [0,1]: $F(x) = x, \quad 0 < x < 1, \qquad Q(u) = u, \quad 0 < u < 1.$

3) Standard exponential: $F(x) = 1 - e^{-x}, \quad x > 0$, and

$$Q(u) = -\log(1-u), \quad 0 < u < 1.$$

4) Weibull $Q(u) = \{-\log(1-u)\}^{\beta}, \quad 0 < u < 1.$

5) Cauchy: $Q(u) = \tan \pi(u - 0.5), \quad 0 < u < 1.$

2.4 In this paper theorems on the asymptotic distribution of order statistics and extreme values assume that $F(x)$ is continuous, with probability density $f(x) = F'(x)$ and hazard function $h(x) = f(x)/(1-F(x))$. We define: $q(u) = Q'(u)$,quantile density function; $fQ(u) = f(Q(u))$,density quantile function; $h_1Q(u) = fQ(u)/(1-u)$,right hazard quantile function; $h_0Q(u) = fQ(u)/u$, left hazard quantile function.

2.5 Tail exponents of probability laws are defined in terms of representations of density quantile functions as regularly varying functions. We call α_0 the left tail exponent if

$$fQ(u) = u^{\alpha_0} L_0(u)$$

where $L_0(u)$ is slowly varying as $u \to 0$, in the sense that for every $y > 0$

$$L_0(yu)/L_0(u) \to 1 \text{ as } u \to 0.$$

Typically, $L_0(u) \sim (\log u)^{\beta}$ for some β.

We call α_1 the right tail exponent if

$$fQ(1-t) = t^{\alpha_1} L_1(t)$$

where $L_1(t)$ is slowly varying as $t \to 0$. We prefer to write this representation

$$fQ(u) = (1-u)^{\alpha_1} L_1(u)$$

Typically, $L_1(u) \sim (-\log(1-u))^{\beta}$ for some β.

Hazard quantile functions can be represented

$$h_0Q(u) = u^{\alpha_0 - 1} L_0(u)$$
$$h_1Q(u) = (1-u)^{\alpha_1 - 1} L_1(u).$$

We call slowly varying functions $L(u)$ log-like.

2.6 Left tails of a distribution are classified as short, medium, or long tail if $\alpha_0 < 1$, $\alpha_0 = 1$, $\alpha_0 > 1$ respectively. Right tails are classified as short, medium, or long tail if $\alpha_1 < 1, \alpha_1 = 1, \alpha_1 > 1$ respectively. Examples of tail classification: Normal: $\alpha_0 = \alpha_1 = 1$; Exponential: $\alpha_0 = 0, \alpha_1 = 1$; Uniform: $\alpha_0 = \alpha_1 = 0$; Cauchy: $\alpha_0 = \alpha_1 = 2$.

2.7 Power transformations (also called Box-Cox transformations) are defined to be following functions of $x > 0$:

$$g(x;\lambda) = \begin{cases} \dfrac{x^{\lambda} - 1}{\lambda}, & \lambda \neq 0 \\ \log x, & \lambda = 0 \end{cases}$$

A power transformation $g(x;\lambda)$ is increasing in x, and at $x = 1$ has value 0 and slope 1. Note that $g'(x;\lambda) = x^{\lambda-1}$.

Important classes of distributions are the power uniform and the power exponential. A random variable X is: (1) power-uniform if for some constant λ

$$X = g(U;\lambda) \text{ or } X = -g(U;\lambda)$$

where U is uniform [0,1]; (2) power-exponential if for some constant λ

$$X = g(W;\lambda) \text{ or } X = -g(W;\lambda)$$

where W is Exponential with mean 1.

The three types of extreme value distributions are actually an infinite family of power-exponential distribution. The case of $\lambda = 0$ is of great theoretical interest, but may not be useful in practical model fitting.

It should be noted that Tukey λ-family of distributions is defined to have quantile functions

$$Q(u) = g(u;\lambda) - g(1-u;\lambda)$$

which are functions of u on $0 < u < 1$ which are symmetric about $u = .5$.

Power transformations play a central role in extreme value distribution theory because of the quantile tail properties that can be established for regularly varying quantile functions with tail exponents α_0 and α_1.

2.8 Quantile Tail Value Property. Regularly varying quantile functions are power-uniform in the tail.

For any sequences $t_n \to 0$, $y_n \to y > 0$, as n tends to ∞ :

$$g_{0,n}(y_n) \to g(y;1-\alpha_0), g_{1,n}(y_n) \to -g(y;1-\alpha_1)$$

defining the tail quantile functions

$$g_{0,n}(y) = h_0Q(t_n)(Q(t_n y) - Q(t_n))$$

$$g_{1,n}(y) = h_1Q(1-t_n)(Q(1-t_n y) - Q(1-t_n))$$

One can show that $g_{0,n}(u)$, $0 < u < 1$, is the quantile function of the random variable censored at the left tail

$$Y = h_0Q(t_n)\{X - Q(t_n)\}, \quad X \le Q(t_n)$$

with distribution function, on $-\infty < x < 0$,

$$F_Y(x) = F(Q(t_n)) + \{h_0Q(t_n\}^{-1}x)/t_n,$$

To prove the quantile tail property one writes

$$h_0 Q(t_n)(Q(t_n y) - Q(t_n)) = \frac{1}{t_n}\int_{t_n}^{t_n y} \{fQ(t_n)/fQ(u)\}du$$
$$= \int_1^y \{fQ(t_n)/fQ(zt_n)\}dz$$
$$= \int_1^y z^{-\alpha_0}dz + o(t_n)$$

The rate of convergence of $L_0(t_n)/L_0(zt_n)$ to 1 determines the rate of convergence of the quantile tail property which is the most important factor affecting the rate of convergence of extreme values to their limit distributions.

DeHaan (1985) has shown that in order that extreme values from a distribution with quantile function Q(u) have limiting distributions, it is necessary and sufficient that Q(u) obey the above conditions (for suitable sequences a_n replacing $h_0 Q(t_n)$).

DeHann shows (p.33) that a distribution function F satisfies, for some choice of constants $a_n > 0$ and $b_n (n = 1, 2, \ldots)$

(1) $$F^n(a_n x + b_n) \to G(x), \text{extreme value distribution,}$$

weakly (as n tends to ∞) if and only if for a positive function $a(t)$

$$\lim_{t \to 0} \frac{Q(1-ty) - Q(1-t)}{a(t)} = -g(y;\lambda)$$

The proof consists of the following steps: (1) is equivalent to

(2) $$n\{1 - F(a_n x + b_n)\} \to -\log G(x);$$

$-\log G(x)$ has inverse function (of y) equal to

$$-Ag(y;\lambda) + B$$

for some constants $A > 0$, B and λ real since if $G(x)$ is distribution function of $-g(W;\lambda)$, W exponential with mean 1,

$$G(x) = PROB[W \geq g^{-1}(-x;\lambda)]$$
$$= \exp(-g^{-1}(-x;\lambda)),$$
$$-\log G(x) = g^{-1}(-x;\lambda);$$

the inverse of $y = n\{1 - F(a_n x + b_n)\}$ is

$$\frac{1}{a_n}\{Q(1 - \frac{y}{n}) - b_n\};$$

(2) is equivalent to convergence of the inverse functions:

$$(3) \qquad \frac{1}{a_n}\{Q(1-\frac{y}{n})-b_n\} \to -Ag(y;\lambda)+B;$$

subtracting (3) for $y = 1$, one obtains

$$\frac{1}{a_n}\{Q(1-\frac{y}{n})-Q(1-\frac{1}{n})\} \to -Ag(y;\lambda)$$

We will apply the quantile tail property with $y_n = (1/t_n)Q_{U(k;n)}(u)$ and $y = Q_G(u)$. We use the following general property of convergence in distribution (called Rubin's theorem: see Fabian and Hannan (1985), p. 142).

Theorem: If $g_n(y_n) \to g(y)$ for every sequence $y_n \to y$, and if $Y_n \to Y$ in distribution, then $g_n(Y_n) \to g(Y)$ in distribution.

2.9 A powerful property of quantile functions is how they behave for monotone functions of random variables.

Theorem: $Y = g(X)$ has $Q_Y(u) = g(Q_X(u))$, assuming $g(x)$ non-decreasing and continuous from left (which is true for a quantile function $Q(u)$). Note that when g is non-increasing and continuous,

$$Y = g(X) \text{ has } Q_Y(u) = g(Q_X(1-u))$$

2.10 The quantile representation of X in terms of U is another powerful property of quantile functions.

Theorem: Quantile Representation. $Q(U) \stackrel{D}{=} X$ where $\stackrel{D}{=}$ denotes identical distribution (equality in distribution), and U is uniform $[0,1]$.

Proof. We outline two proofs of the quantile representation. From the formula for the quantile function of a function of a random variable

$$Q_{Q(U)}(u) = Q(Q_U(u)) = Q(u) = Q_X(u).$$

Alternatively, $Q(U) \le x$ if and only if $U \le F(x)$. Consequently,

$$\Pr[Q(U) \le x] = \Pr[U \le F(x)] = F(x)$$

2.11 The foregoing results yield an important theorem relating the distribution of general order statistics to uniform order statistics.

Theorem: $X(k;n) \stackrel{D}{=} Q(U(k;n))$,

$$\begin{aligned} Q_{X(k;n)}(u) &= Q_{Q(U(k;n))}(u) \\ &= Q(Q_{U(k;n)}(u)) \end{aligned}$$

2.12 Identification Quantile function $QI(u), 0 < u < 1$, is defined by

$$QI(u) = (Q(u) - MQ)/DQ$$

where $MQ = Q(.5)$ is the median, and $DQ = 2(Q(.75) - Q(.25))$ is the quartile deviation.

The values of $QI(u)$ for u near 0 and 1 can be shown to provide quick diagnostics of the tail classification of the distribution.

2.13 Location - Scale parameter models for probability distribution are expressed in terms of quantile functions by

$$Q(u) = \mu + \sigma Q_0(u)$$

where $Q_0(u)$ is a specified 'standard quantile function.' We propose that if one always chooses $Q_0(u) = QI(u)$, then one can more easily compare μ and σ corresponding to different models.

An alternative normalization is the unitized quantile function

$$Q1(u) = (Q(u) - MQ)/q(0.5)$$

The corresponding probability density function $f_1(x)$ has the property

$$f_1(\text{median}) = 1.$$

For a normal distribution $Q1(u)$ has probability density $f_1(x) = e^{-\pi x^2}$.

3. Beta, Gamma, Power Distributions.

3.1 Beta distribution BETA(p, q) has probability density

$$f_{Beta(p,q)}(x) = \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1}, \qquad 0 < x < 1$$

where B(p, q) is the Beta function defined by

$$B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1}\,dx = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$

Note

$$\frac{1}{B(k, n+1-k)} = n\binom{n-1}{k-1}$$

3.2 Gamma distribution. Gamma(m) where $m > 0$, has probability density

$$f_{Gamma(m)}(x) = \frac{1}{\Gamma(m)} x^{m-1} e^{-x}, x > 0.$$

Note that Gamma(1) is the exponential distribution. We define

$$\text{GammaA}(m) = \frac{1}{m}\text{Gamma}(m)$$

3.3 PowerG(λ, m) distribution, where $m > 0$ and $-\infty < \lambda < \infty$, is defined by a random variable representation:

$$\text{PowerG}(\lambda, m) = g(\text{GammaA}(m); \lambda)$$

For $m = 1$, this is the power-exponential distribution.

The quantile function of the distribution is given by

$$Q_{\text{PowerG}(\lambda,m)}(u) = g(\frac{1}{m}Q_{\text{Gamma}(m)}(u); \lambda) \tag{1}$$

Note that $Q_{\text{Gamma}(m)}(u)$ does not have an analytic formula (for m not equal to 1) but is computed numerically.

3.4 NegPowerG(λ, m) distribution is defined by

$$\text{NegPowerG}(\lambda, m) = -g(\text{GammaA}(m); \lambda)$$
$$Q_{\text{NegPowerG}(\lambda,m)}(u) = -g((1/m)Q_{\text{Gamma}(m)}(1-u); \lambda)$$

3.5 Wilson Hilferty transformation of Gamma(m) to normal is an approximation formula for quantiles

$$\begin{aligned} Q_{\text{GammaA}(m)}(u) &= \frac{1}{m}Q_{\text{Gamma}(m)}(u) \\ &= \left\{1 + \frac{1}{3\sqrt{m}}\Phi^{-1}(u) - \frac{1}{9m}\right\}^3 \end{aligned} \tag{2}$$

Graphs of exact and approximate quantile functions indicate how well the approximation works, even for m=1; see Figure 1.

3.6 Quantile function of PowerG is approximated by substituting (2) into (1).

4. Exact and Asymptotic Distributions of Extreme Order Statistics.

4.1 The exact distribution of $X(k;n)$ is expressed in terms of quantile function $Q(u)$ and the quantile function $Q_{U(k;n)}$ of $U(k;n)$:

$$Q_{X(k;n)}(u) = Q(Q_{U(k;n)}(u))$$

The quantile function of $U(k;n)$ is obtained numerically from the distribution function of the Beta probability density.

4.2 Theorem: $U(k;n)$ is Beta $(k, n+1-k)$:

$$f_{U(k;n)}(x) = \frac{1}{B(k,n+1-k)} x^{k-1}(1-x)^{n-k}, \qquad 0 < x < 1.$$

We write this fact in terms of quantile functions:

$$Q_{U(k;n)}(u) = Q_{Beta(k,n+1-k)}(u)$$

Proof: To compute the probability that $U(k;n)$ is in $(x, x+dx)$, there are n ways to choose a variable in $(x, x+dx)$ and $\binom{n-1}{k-1}$ to choose $k-1$ variables in $(0,x)$. For specified variables lying in these regions, probability of event is $x^{k-1}(1-x)^{n-k}$. The product of n, $\binom{n-1}{k-1}$, and $x^{k-1}(1-x)^{n-k}$ is the value of the probability density.

Graphs of Beta probability densities should be examined for: $n = 20, k = 1, 2, ..., 20$; $n = 100$, $k = 1, ..., 10, 20, 30, 40, 50, 60, 70, 80, 90, 81, ..., 100$.

4.3 The mean of $U(k;n)$ is the expected value of the k^{th} order statistic from a uniform $[0,1]$ and is given a special notation:

$$E[U(k;n)] = \frac{k}{n+1} = u_{k;n}$$

The covariances also play a central role; one can show that, for $j < k$,

$$Cov[U(j;n), U(k;n)] = \frac{1}{n+1} u_{j;n}(1 - u_{k;n})$$

We review the proof of the formula for mean:

$$\begin{aligned} E[U(k;n)] &= n\binom{n-1}{k-1} \int_0^1 x x^{k-1}(1-x)^{n-k}\,dx \\ &= n\binom{n-1}{k-1} / (n+1)\binom{n}{k} = \frac{k}{n+1} \end{aligned}$$

4.4 The asymptotic distributions of $U(k;n)$, $U(n+1-k;n)$, expressed in terms of quantile functions, are

$$\begin{aligned} &Q_{nU(k;n)}(u) \to Q_{Gamma(k)}(u), Q_{U(k;n)/u_{k;n}}(u) \to \frac{1}{k} Q_{Gamma(k)}(u), \\ &Q_{n(1-U(n+1-k;n))}(u) \to Q_{Gamma(k)}(u). \end{aligned}$$

The form in which we apply these results in Theorem 4.5 is

$$\frac{\{U(n+1-k;n) - u_{n+1-k;n}\}}{1 - u_{n+1-k;n}} \to -\{(1/k)Gamma(k) - 1\}$$

These important relations follow (by general theorems) from the fact that the probability density functions obey similar limit theorems or from representations of uniform order statistics in terms of consecutive sums of exponential random variables. Graphical representations of this approximation are given in Figures 2A-2B.

The proof in terms of density functions:

$$
\begin{aligned}
f_{nU(k;n)}(y) &= \frac{1}{n} f_{U(k;n)}(y/n) \\
&= \binom{n-1}{k-1} (y/n)^{k-1} (1 - \frac{y}{n})^{n-k} \\
&\to \frac{1}{(k-1)!} y^{k-1} e^{-y};
\end{aligned}
$$

$$
\begin{aligned}
f_{n(1-U(n+1-k;n))}(y) &= \frac{1}{n} f_{U(n+1-k;n)}(1 - \frac{y}{n}) \\
&= \binom{n-1}{k-1} (1 - \frac{y}{n})^{n-k} (\frac{y}{n})^{k-1} \\
&\to \frac{1}{(k-1)!} y^{k-1} e^{-y}
\end{aligned}
$$

4.5 The exact distribution of $X(k;n)$ is expressed in terms of quantile functions.

$$Q_{X(k;n)}(u) = Q_{Q(U(k;n))}(u) = Q(Q_{Beta(k,n+1-k)}(u))$$

The asymptotic distribution of $X(k;n)$ is obtained by defining $u_{k;n} = k/(n+1)$, and writing

$$X(k;n) - Q(u_{k;n}) \stackrel{D}{=} Q(U(k;n)) - Q(u_{k;n}),$$

The Quantile Tail Property enables us to show that

$$
\begin{aligned}
&h_0 Q(u_{k;n})(Q(u_{k;n} \frac{(n+1)}{k} U(k;n)) - Q(u_{k;n})) \\
&\quad \to g(GammaA(k); 1 - \alpha_0)
\end{aligned}
$$

using Rubin's theorem. One can conclude the following basic theorem on the distributions of extreme values under suitable conditions of regular variation of the density quantile function.

4.5 THEOREM: ASYMPTOTIC DISTRIBUTION OF EXTREME VALUES

$$
\begin{aligned}
h_0 Q(u_{k;n})(X(k;n) - Q(u_{k;n})) &\xrightarrow{D} g(GammaA(k); 1 - \alpha_0) \\
&= PowerG(1 - \alpha_0, k) random\ variable, \\
h_1 Q(1 - u_{k;n})(X(n+1-k;n) &- Q(1 - u_{k;n})) \\
&\xrightarrow{D} -g(GammaA(k); 1 - \alpha_1) \\
&= NegPowerG(1 - \alpha_1, k) random\ variable
\end{aligned}
$$

This basic theorem is proved in "one-line" from the quantile tail property, Rubin's theorem, and the asymptotic distributions of extreme values of uniform distribution. It is more precisely stated in terms of quantile functions, and illustrated by graphs comparing exact and asymptotic quantile functions of extreme and central order statistics (to be given in the sequel).

4.6 The limit theorems that we have given for extreme values are stated in a way that will enable us to maximize the analogies between the asymptotic distribution of extreme and central order statistics. Our goal is to obtain unified formulas so that we will not have to decide in practice what is the dividing line between extreme and central order statistics. Order statistics X(k;n) are called: extreme if k is fixed; end if $k/n \to 0$ or 1 (as $n \to \infty$); and central if $k/n \to p$, $0 < p < 1$.

5. Asymptotic Normality of Central Order Statistics.

5.1 To study the distribution of central order statistics $X(k;n)$, we introduce sample quantile function

$$Q\tilde{}(u), 0 < u < 1$$

defined by: for $k = 1, ..., n$

$$Q\tilde{}(u) = X(k;n), \qquad \frac{k-1}{n} < u \le \frac{k}{n}$$

Equivalently

$$Q\tilde{}(u) = F\tilde{}^{-1}(u)$$

where $F\tilde{}(x)$, $-\infty < x < \infty$ is sample distribution function;

$$F\tilde{}(x) = \frac{k}{n} \text{ for } X(k;n) \le x < X(k+1;n)$$

5.2 When F is continuous, a basic role is played by Bahadur representation: For $0 < u < 1$

$$Q\tilde{}(u) - Q(u) = -(F\tilde{}(Q(u)) - u)/fQ(u) + R_n$$

$$n^{.5} R_n \xrightarrow{\mathcal{D}} 0 \text{ as } n \to \infty$$

5.3 The asymptotic distribution of central order statistics $X(k;n)$ is obtained from the asymptotic distribution of $Q\tilde{}(u)$ since $Q\tilde{}(u) = X(k;n)$ for $(k-1)/n < u \le k/n$. A value of u in this interval is denoted by $u_{k;n}$. Typical formulas for $u_{k;n}$ are

$$u_{k;n} = k/(n+1) \quad \text{ or } u_{k;n} = (k-0.5)/n$$

In this paper we use the first definition.

The asymptotic distribution of the sample distribution function can be written: for $0 < p < 1$

$$(n+1)^{0.5}\{F\tilde{}(Q(p) - p)) \stackrel{\mathcal{D}}{=} N(0, p(1-p))$$

From Bahadur's representation one obtains the asymptotic distribution of the sample quantile function which can be written, for $0 < p < 1$,

$$(n+1)^{0.5} fQ(p)\{Q\tilde{}(p) - Q(p)\} \stackrel{D}{=} N(0, p(1-p))$$

5.4 The asymptotic distribution of $Q(p)$ can be expressed in terms of hazard quantile functions.

$$(n+1)^{0.5} h_0 Q(p)\{Q\tilde{}(p) - Q(p)\} \stackrel{D}{=} N(0, (1-p)/p), 0 \leq p \leq 0.5;$$

$$(n+1)^{0.5} h_1 Q(p)\{Q\tilde{}(p) - Q(p)\} \stackrel{D}{=} N(0, p/(1-p)), 0.5 \leq p \leq 1$$

To unify the theory of central and order statistics we propose an alternate approximation which is equivalent in the limit to the above.

6. Unified formula for asymptotic distribution of extreme or central order statistics $X(k;n)$. .

Define $u_{k;n} = k/(n+1)$,

$$m_0 = (n+1)u_{k;n}/(1-u_{k;n}), \quad m_1 = (n+1)(1-u_{k;n})/u_{k;n}$$

For $0 < u_{k;n} \leq 0.5$,

$$h_0 Q(u_{k;n})\{Q\tilde{}(u_{k,n}) - Q(u_{k;n})\} \stackrel{D}{=} g(GammaA(m_0); 1-\alpha_0)$$

For $0.5 \leq u_{k;n} < 1$

$$h_1 Q(u_{k;n})\{Q\tilde{}(u_{k;n}) - Q(u_{k;n})\} \stackrel{D}{=} -g(GammaA(m_1); 1-\alpha_1)$$

This formula is stated in terms of random variables. For numerical calculations it is easiest to write it in terms of quantile functions: For $0 < u_{k;n} = k/(n+1) \leq 0.5$

$$Q_{X(k;n)}(u) = Q_{Q\tilde{}(u_{k;n})}(u) = Q(u_{k;n}) + Q_0(u),$$

$$Q_0(u) = \frac{1}{h_0 Q(u_{k;n})} Q_{g(GammaA(m_0);1-\alpha_0)}(u)$$

For $0.5 \leq u_{k;n} = k/(n+1) < 1$

$$Q_{X(k;n)}(u) = Q_{Q\tilde{}(u_{k;n})}(u) = Q(u_{k;n}) + Q_1(u)$$

$$Q_1(u) = \{h_1 Q(u_{k;n})\}^{-1}\{-Q_{g(GammaA(m_1);1-\alpha_1)}(1-u)\}$$

When applied to extreme order statistics (say $X(k;n)$ for k fixed) the unified formula uses as the Gamma degrees of freedom

$$m_0 = k/(1 - \frac{k}{n+1})$$

rather than k. When applied to central order statistics the unified formula uses a PowerGamma distribution rather than the normal distribution.

To empirically demonstrate the practical use of the unified formula, one compares the quantile function of the exact distribution of $X(k;n)$,

$$Q_{X(k;n)}(u) = Q(Beta(k, n+1-k))$$

with the approximate distribution given by the unified formula. The examples we have considered are $n = 20, k = 1, ..., 10$ and $n = 100$, $k = 1, ..., 10, 20, 30, 40, 50$ for

Table 3: Distributions Used to Compare Exact and Approximate Density Quantile Functions of Order Statistics

Uniform, $\alpha_0 = 0$,	$Q(u) = u$ $h_0 Q(u) = u^{-1}$
Weibull, $\alpha_0 = 1 - \beta$	$Q(u) = \{\log(1-u)^{-1}\}^{\beta}$
$\beta = 1/6, 2/6, 3/6, 4/6, 5/6$	$h_0 Q(u) = \frac{1}{\beta} u^{-1}(1-u)\{\log(1-u)^{-1}\}^{1-\beta}$
Cauchy $\alpha_0 = 2$	$Q(u) = \tan \pi(u - .5)$ $= -\cos \pi u / \sin \pi u$ $h_0 Q(u) = \{\sin(\pi u)\}^2 / \pi u$

1) Cauchy distribution $Q(u) = \tan \pi(u - 0.5)$, see Figures 4A-4D;
2) exponential-type $Q(u) = \log u$,
3) $Q(u) = (\log u)^{0.5}$

7. Approximate density quantile function of order statistics.

The density quantile function $fQ_{X(k;n)}(u)$ of the probability distribution of $X(k;n)$ can be obtained from the quantile function $Q_{X(k;n)}(u)$ by the formulas

$$q_{X(k;n)}(u) = Q'_{X(k;n)}(u), \quad fQ_{X(k;n)}(u) = q_{X(k;n)}(u)^{-1}.$$

One obtains in this way the following approximate formulas. Let $u_{k;n} = k/(n+1), m_0 = (n+1)u_{k;n}/(1-u_{k;n}), m_1 = (n+1)(1-u_{k;n})/u_{k;n}$. For $0 < u_{k;n} \leq 0.5$,

$$fQ_{X(kn)}(u) = \sqrt{m_0}\, h_0 Q(u_{k;n}) \phi \Phi^{-1}(u)$$
$$(1 + \frac{1}{3\sqrt{m_0}} \Phi^{-1}(u) - \frac{1}{9m_0})^{3\alpha_0 - 2}.$$

For $0.5 \leq u_{k;n} < 1$

$$fQ_{X(k;n)}(u) = \sqrt{m_1}\, h_1 Q(u_{k;n}) \phi \Phi^{-1}(u)$$
$$(1 + \frac{1}{3\sqrt{m_1}} \Phi^{-1}(1-u) - \frac{1}{9m_1})^{3\alpha_1 - 2}.$$

We graph the unitized density quantile functions $fQ(u)/fQ(.5)$, and record the value of $fQ(.5)$, for the cases listed in Table 3. The accuracy of the approximation can be judged by comparing (see Figures 3A-3B) the unitized approximate density $fQ_{X(1,n)}(u)$ of the sample minimum with the unitized density quantile of the Weibull distribution (which is the usual formula for the asymptotic distribution):

$$fQ(u) = \frac{(1-u)}{\beta} \{\log(1-u)^{-1}\}^{1-\beta}$$

8. Stationary Time Series.

When the sample $X(1), \ldots, X(n)$ is from a stationary time series we can represent

$$F^{\sim}(Q(p)) - p = \frac{1}{n}\sum_{t=1}^{n} c(X(t);p)$$

defining

$$\begin{aligned} c(x;p) &= 1-p \quad \text{if } x \leq Q(p) \\ &= -p \quad \text{if } x > Q(p) \end{aligned}$$

Its asymptotic distribution can be written

$$(n+1)^{0.5}(F^{\sim}(Q(p)) - p) \stackrel{D}{=} N(0, p(1-p)f_c(\frac{1}{n};p))$$

defining $f_c(\omega;p)$ to be the spectral density function of the two valued time series $c(X(t);p)$. The time series concepts are illustrated in Figures 5A-5C. The spectral density $f(\omega)$, $0 \leq \omega \leq 1$, of a stationary time series $Y(t)$ with correlation function $\rho(v) = \text{Corr}[Y(t), Y(t+v)]$ is defined by

$$f(\omega) = \sum_{v=-\infty}^{\infty} exp(-2\pi i v\omega)\rho(v).$$

The formulas for asymptotic distribution of order statistics continue to hold if we define

$$m_0 = \{f_c(\frac{1}{n}; u_{k;n})\}^{-1} \quad (n+1)u_{k;n}/(1-u_{k;n}),$$

$$m_1 = \{f_c(\frac{1}{n}; u_{k;n})\}^{-1} \quad (n+1)(1-u_{k;n})/u_{k;n}$$

We conjecture that these formulas continue to hold also for extreme value statistics. If the conjecture is true, extreme value statistics of stationary time series have unimodal distributions. We need to investigate how to compare this result with the asymptotic distribution of extreme values (in terms of waiting times of compound Poisson processes) found by Leadbetter and Hsing.

REFERENCES

DE HANN, L. (1984) *Slow Variation and Characterization of Domains of Attraction*, in Tiago de Oliveira (1984), pp. 31-40.

FABIAN, V. AND HANNAN, J. (1985) *Introduction to Probability and Mathematical Statistics*. Wiley: New York.

TIAGO DE OLIVEIRA, J. (1984) *Statistical Extremes and Applications*. Reidel Publishing: Dordrecht, Holland.

FIGURE CAPTIONS

Figure 1. Exact (solid curve), and Wilson-Hilferty Approximations (dotted curve) to, quantile functions of Gamma distributions with 1, 2, and 3 degrees of freedom.

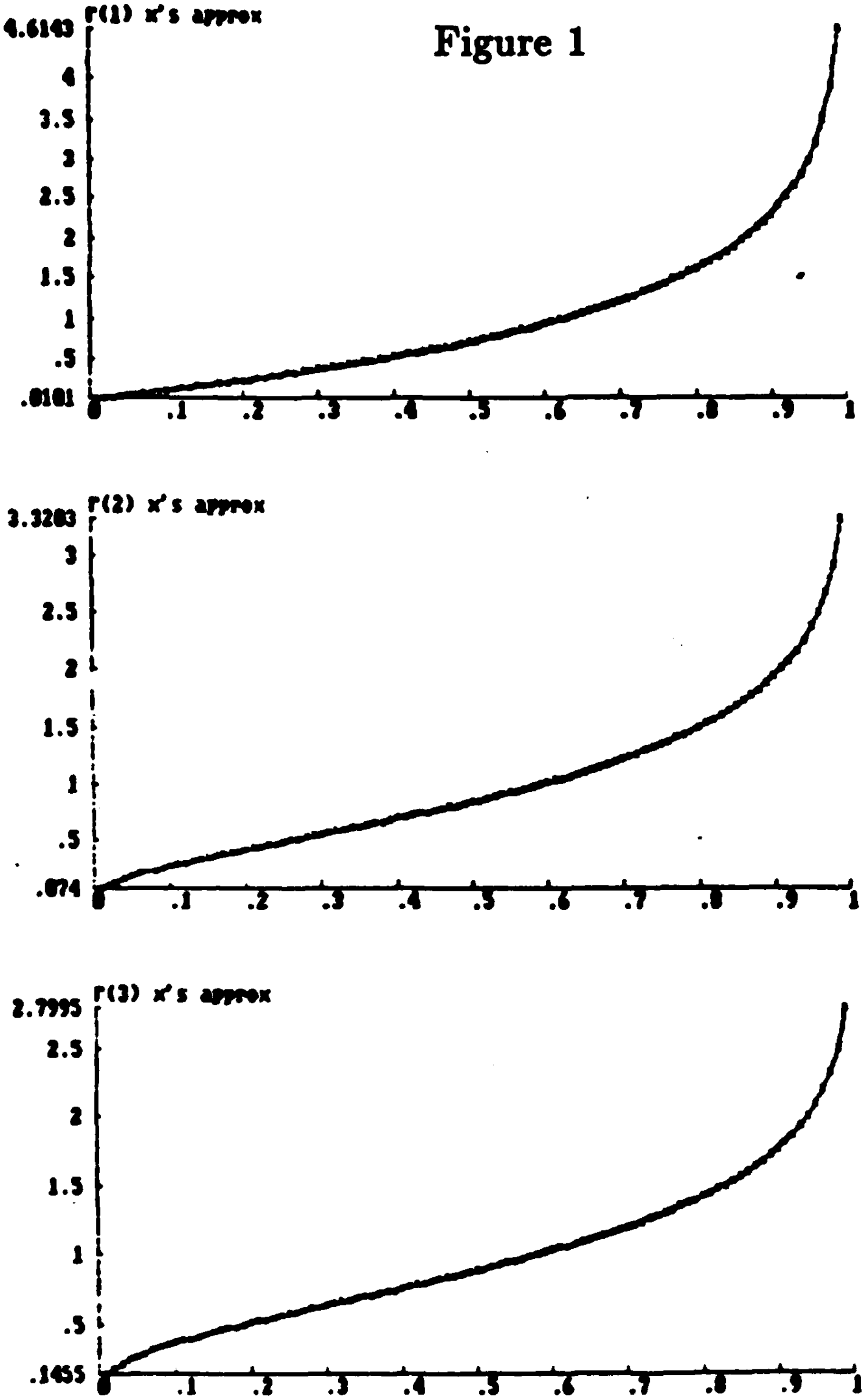
Figures 2A, 2B. Exact beta and approximating gamma distributions of $X(1;20)$ and $X(2;20)$ from uniform distribution. Graphs of density functions, identification quantile functions, and unitized density quantile function.

Figures 3A, 3B. Exact and approximating unitized density quantile functions of $X(1;100)$, $X(2;100)$, $X(3;100)$ from Weibull $Q(u) = \{-\log(1-u)\}^{\beta}$, $\beta = .333$ and $\beta = .167$.

Figures 4A, 4B, 4C, 4D. Exact and approximating quantile functions and unitized density quantile functions for $X(1;100)$, $X(2;100)$, $X(3;100)$, $X(4;100)$, $X(20;100)$, $X(30;100)$, $X(40;100)$, $X(50;100)$ from Cauchy distribution.

Figures 5A, 5B, 5C. Time series analysis of samples of length $n = 100$ from $AR(1)$ model $Y(t) - .95 \quad Y(t-1) = e(t)$ clipped at sample $Q(u)$, $u = .5$ and $u = .975$.

Figure 1



Γ(1) Exponential

Γ(2)

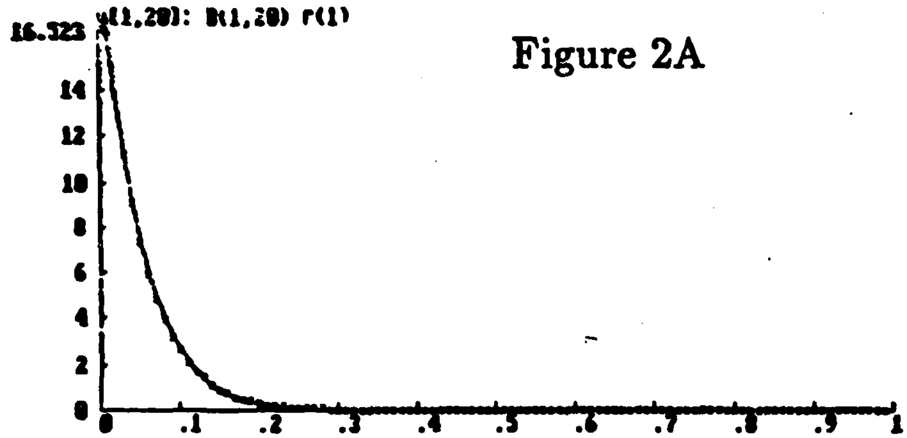
Γ(3)

WILSON-HILFERTY

APPROXIMATION

$$G_{Gamma(m)}(u) = \frac{1}{m} Q_{Gamma(m)}(u) = \{1 + \frac{1}{3\sqrt{m}} \Phi^{-1}(u) - \frac{1}{9m}\}^3$$

Figure 2A

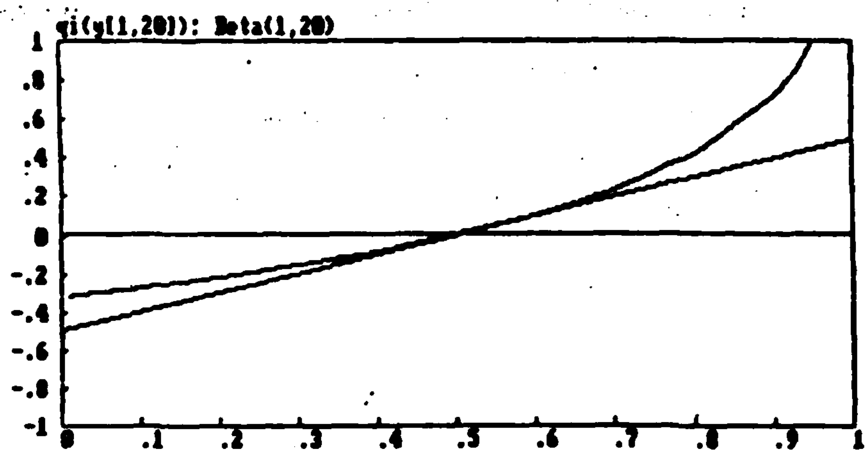


Uniform $n = 20$

$X(1;20)$

$Beta(1;20)$

Approx. $\frac{1}{n} Gamma(1)$



$QI(u)$

$Beta(1,20)$

Similar to exponential.

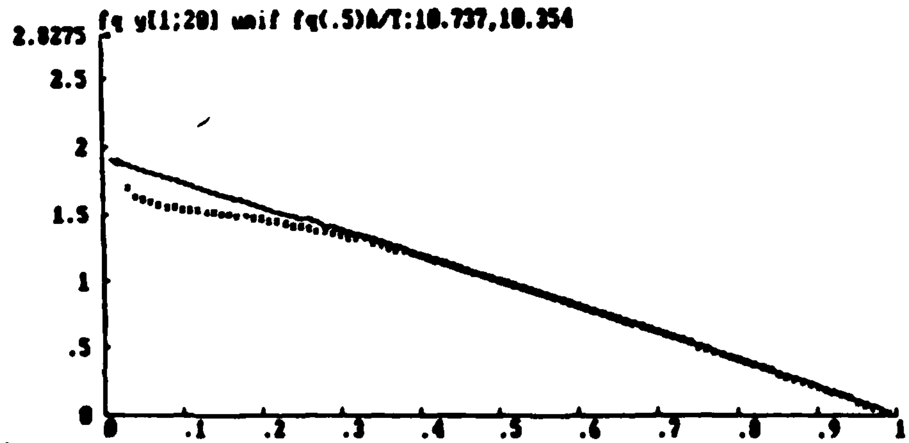


Figure 2B

Uniform $n = 20$

$X(2;20)$

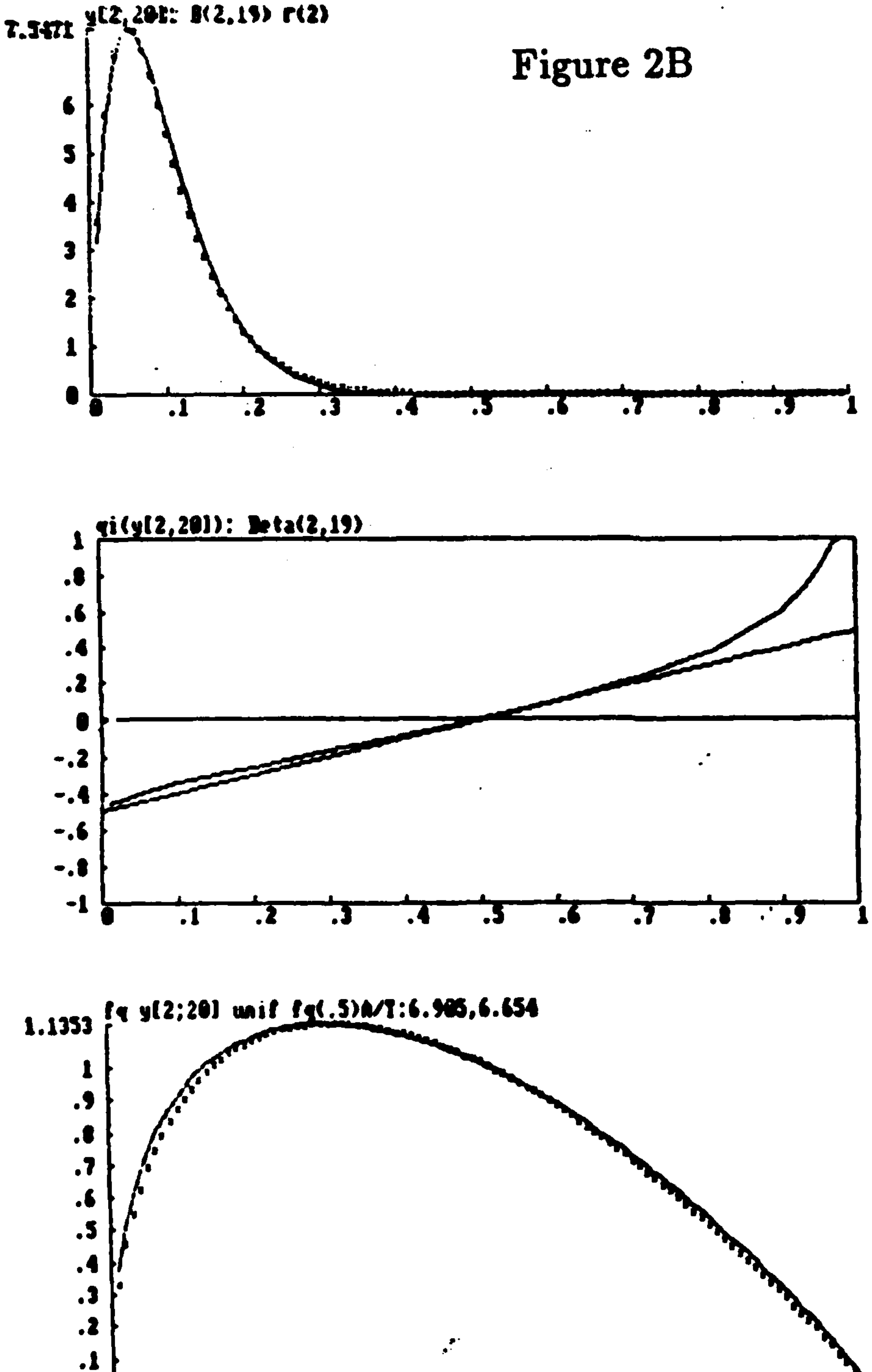
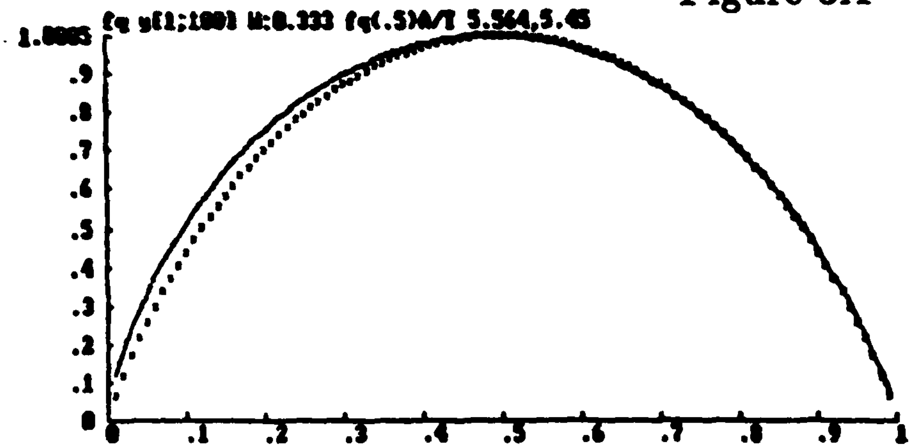
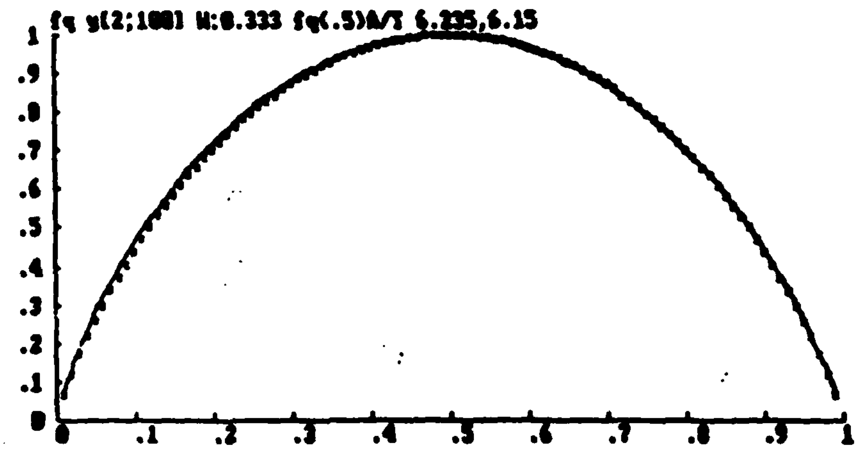


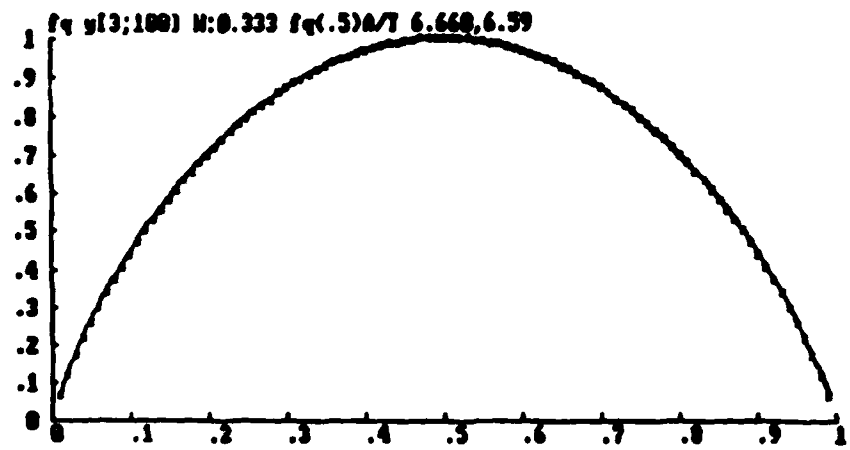
Figure 3A



$X(1;100)$



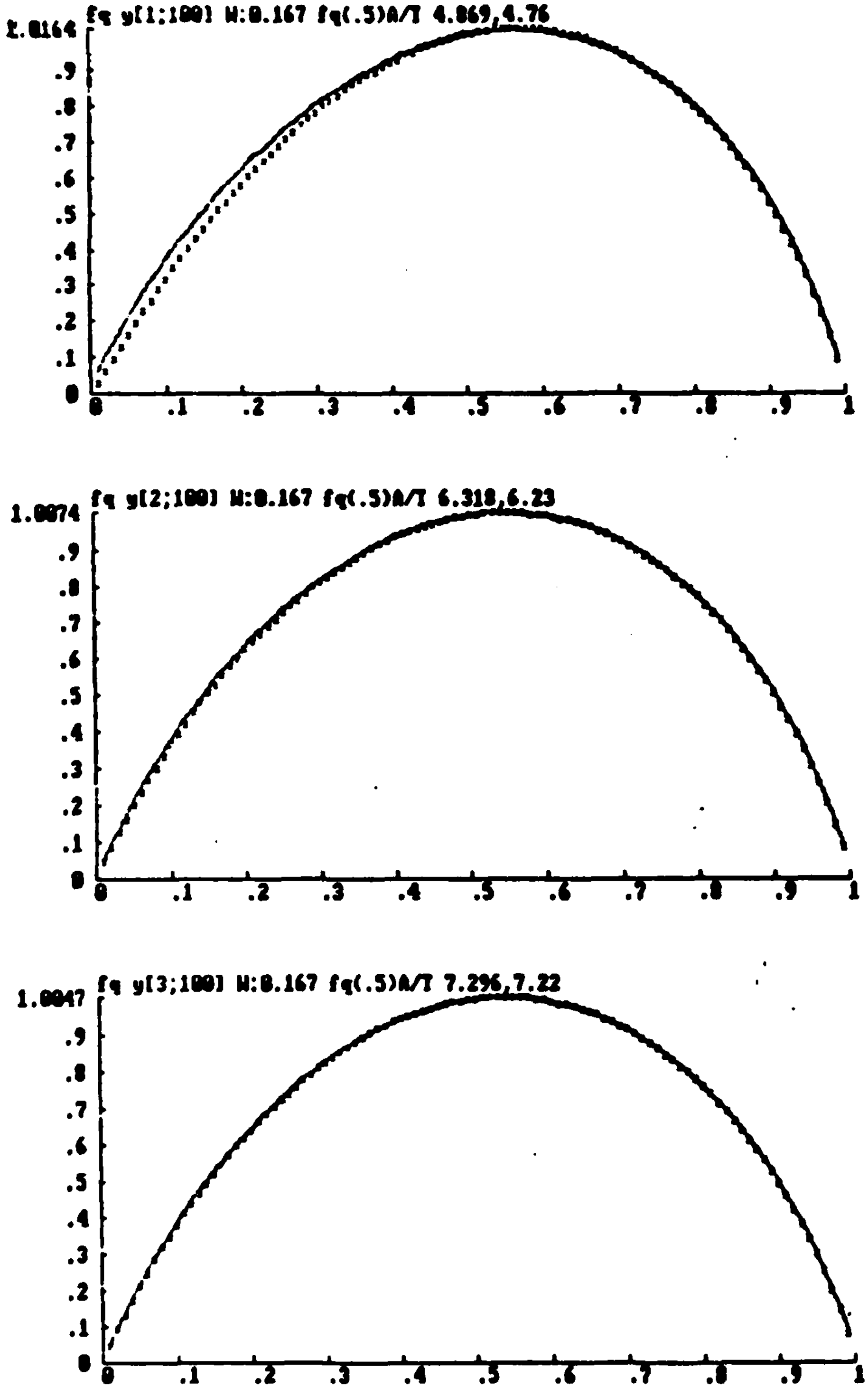
$X(2;100)$



$X(3;100)$

WEIBULL $\beta = .333$ $\qquad Q(u) = \{\log(1-u)^{-1}\}^{\beta}$

Figure 3B



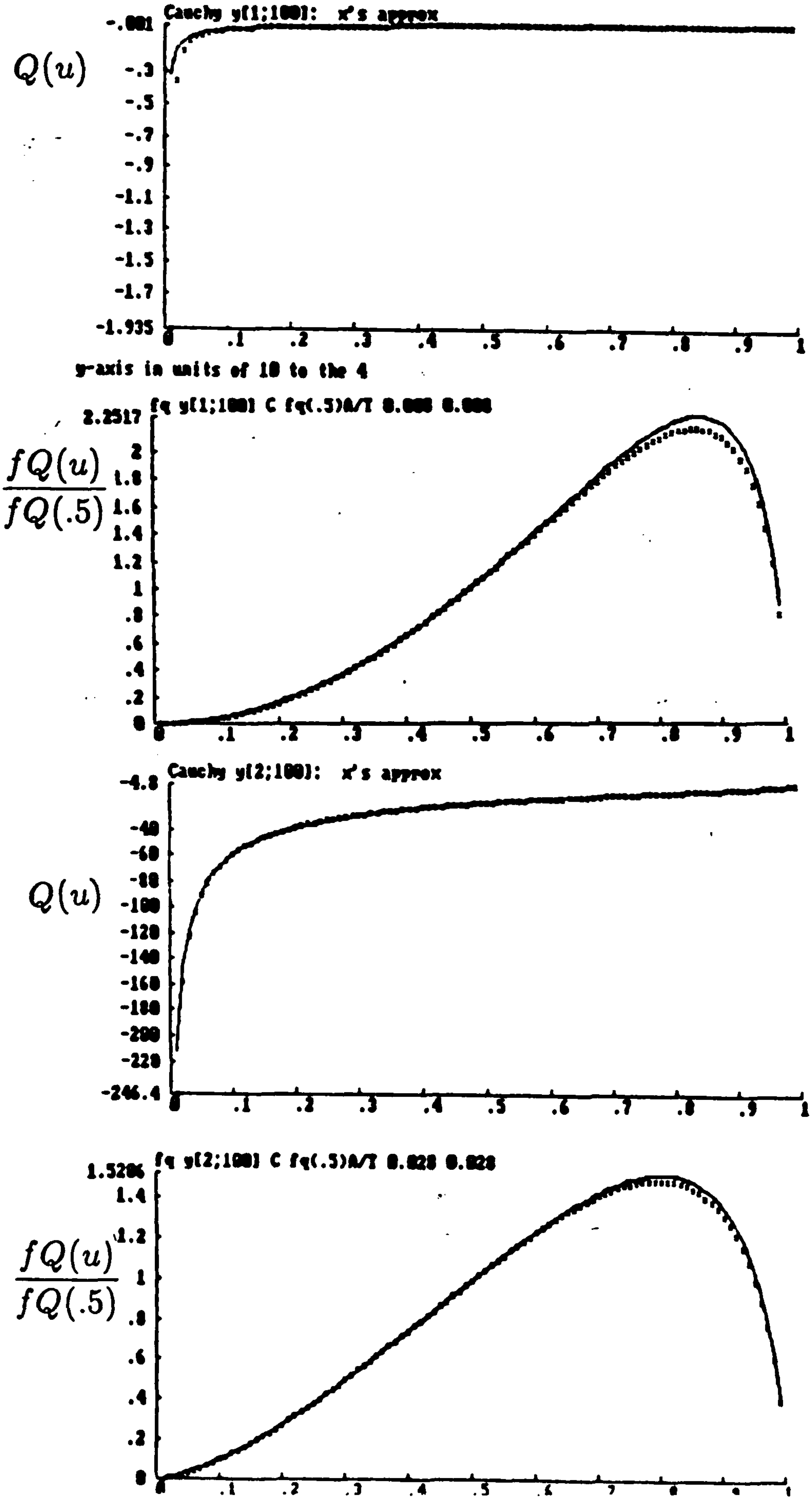
$X(1;100)$

$X(2;100)$

$X(3;100)$

WEIBULL $\beta = .167$ $\qquad Q(u) = \{\log(1-u)^{-1}\}^{\beta}$

Figure 4A



CAUCHY $X(1;100)$

Mean $= \infty$

CAUCHY $X(2;100)$

Figure 4B

CAUCHY $X(3;100)$

CAUCHY $X(4;100)$

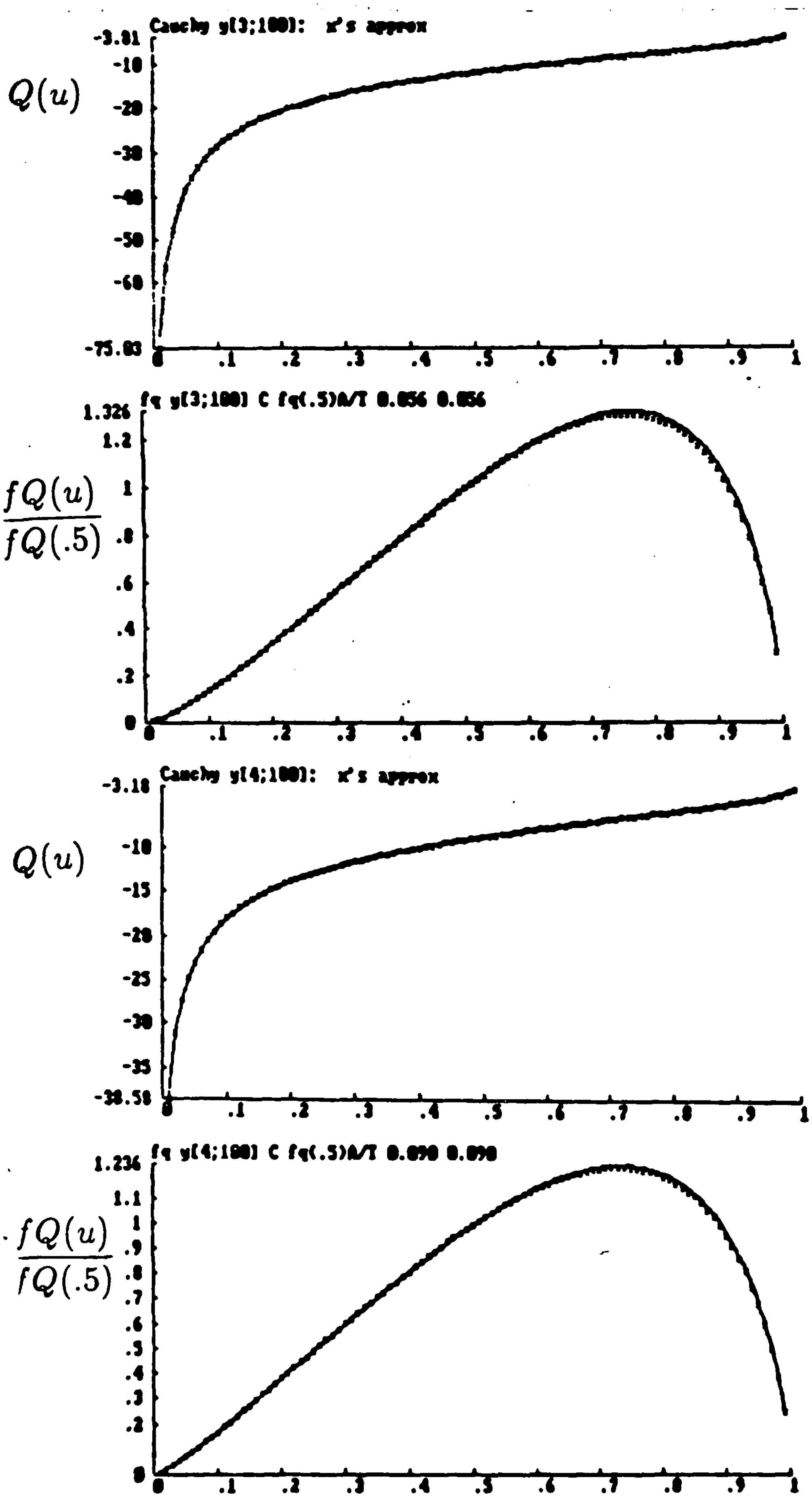


Figure 4C

CAUCHY $X(20;100)$

CAUCHY $X(30;100)$

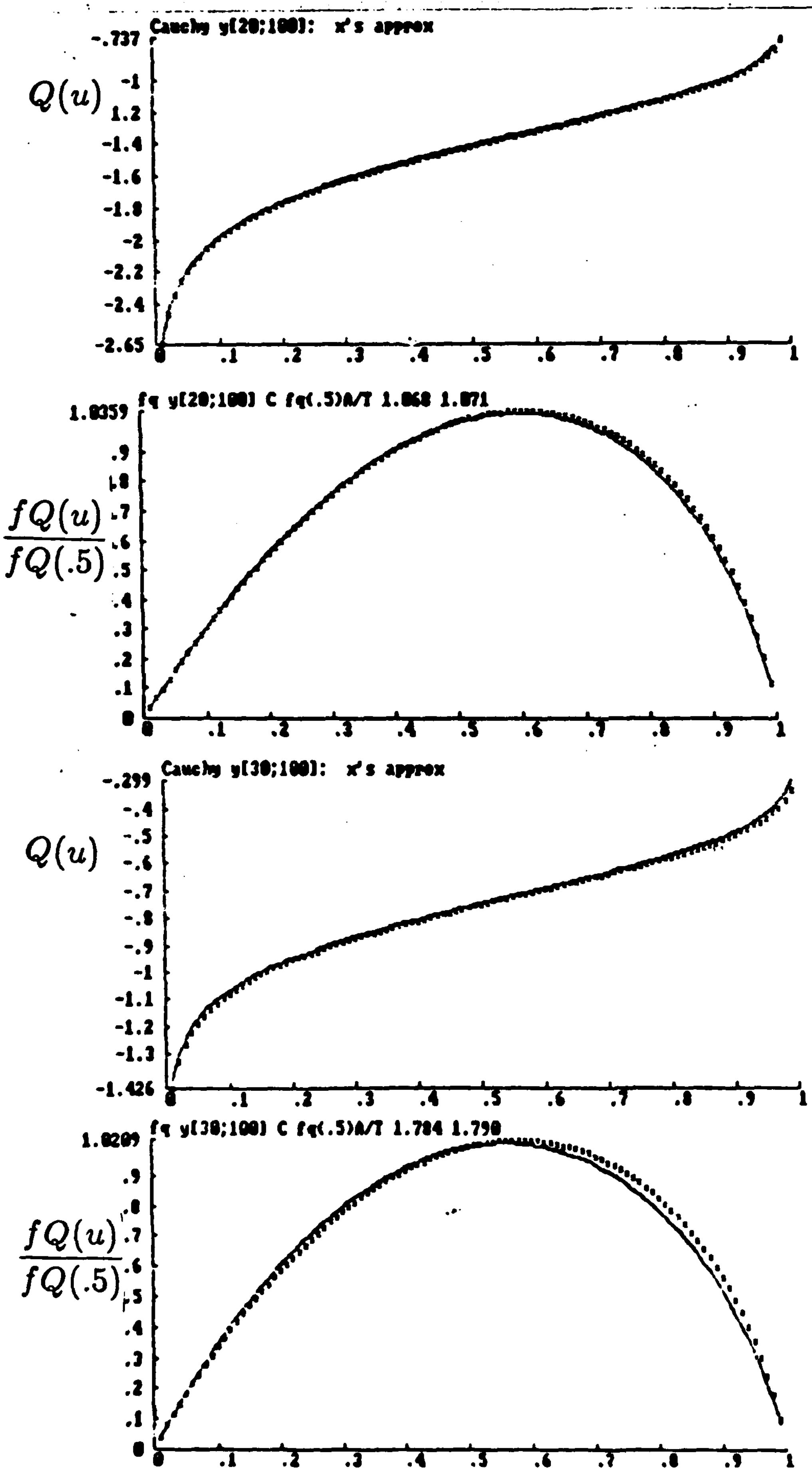
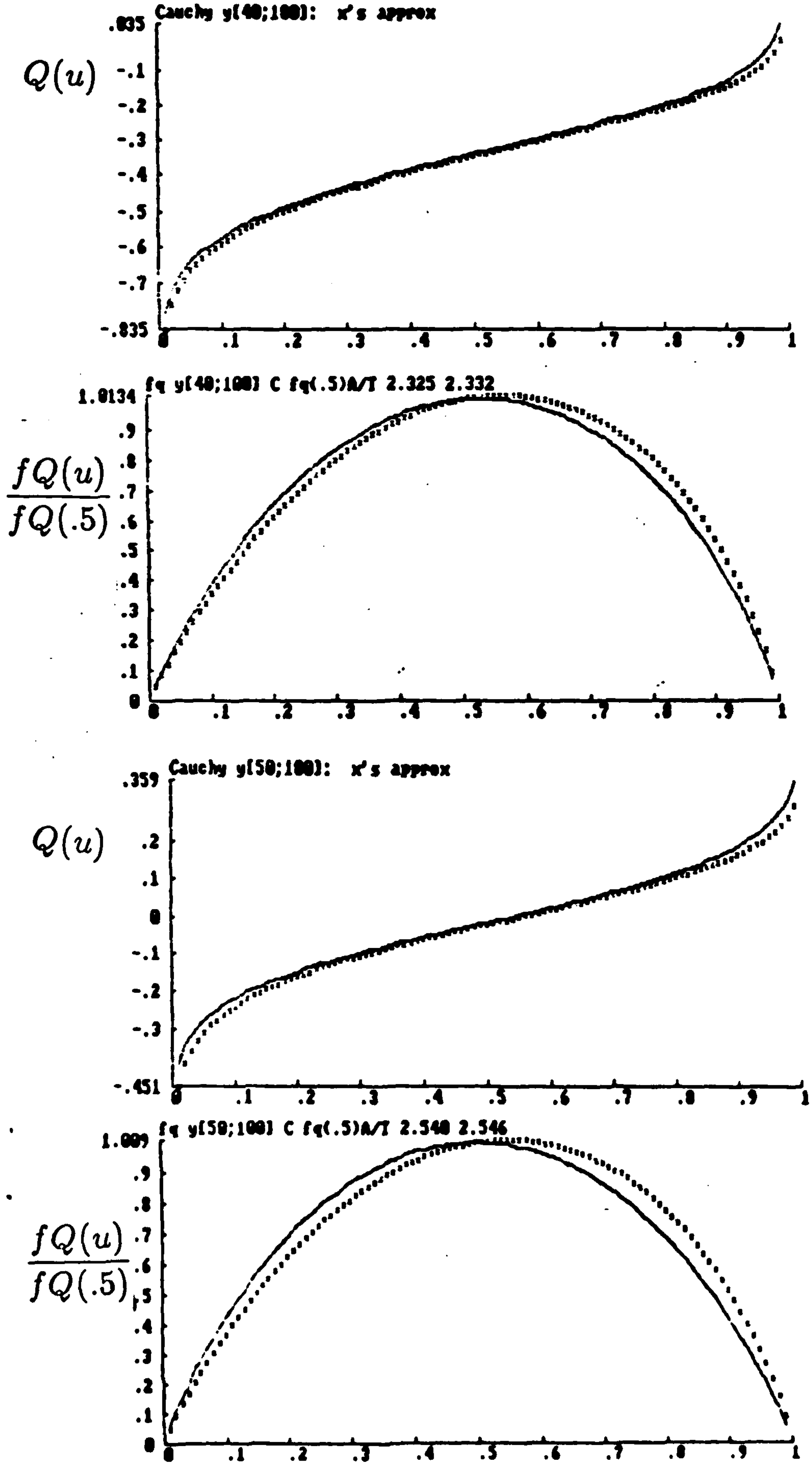


Figure 4D



CAUCHY $X(40;100)$

CAUCHY $X(50;100)$

Figure 5A

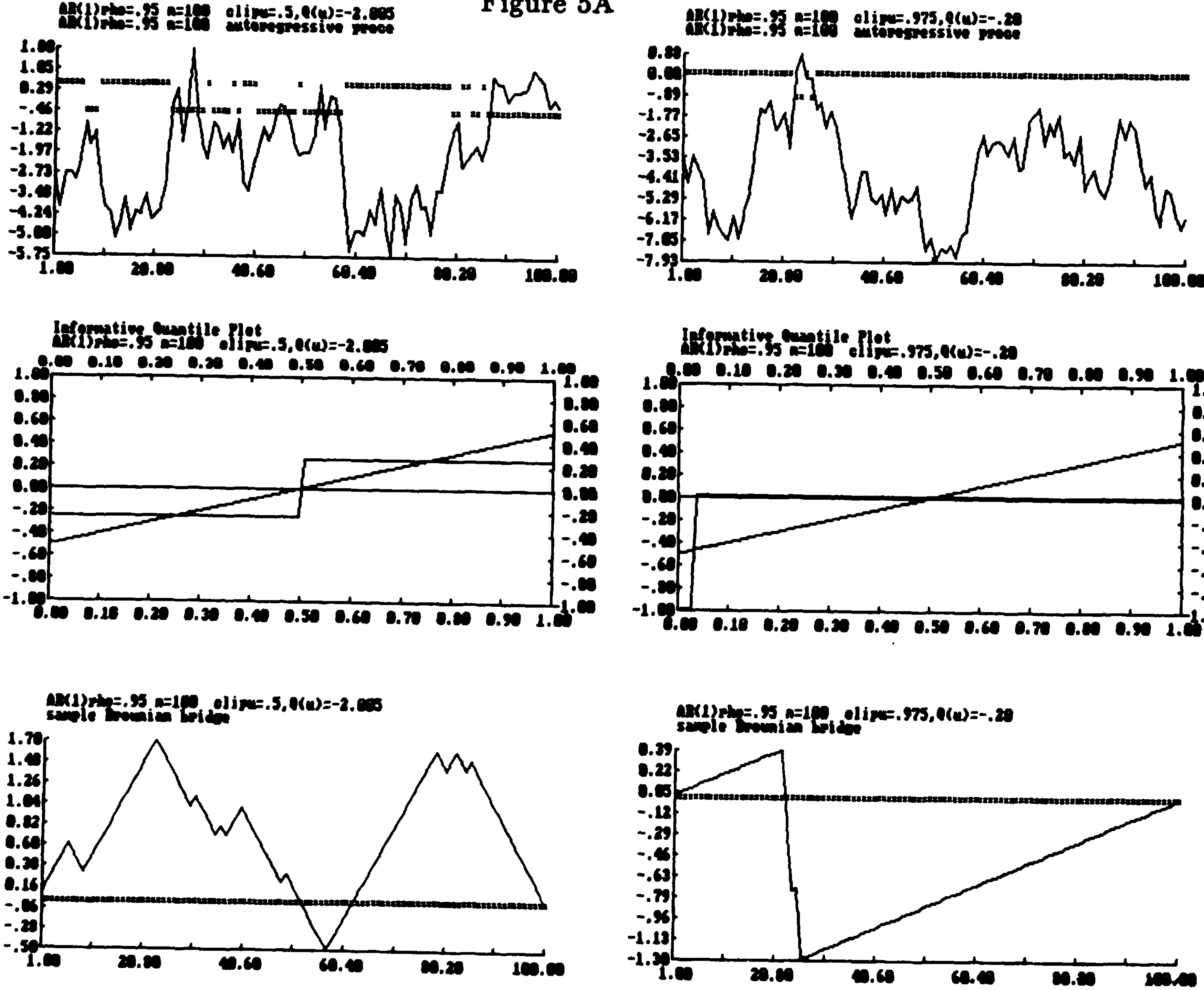
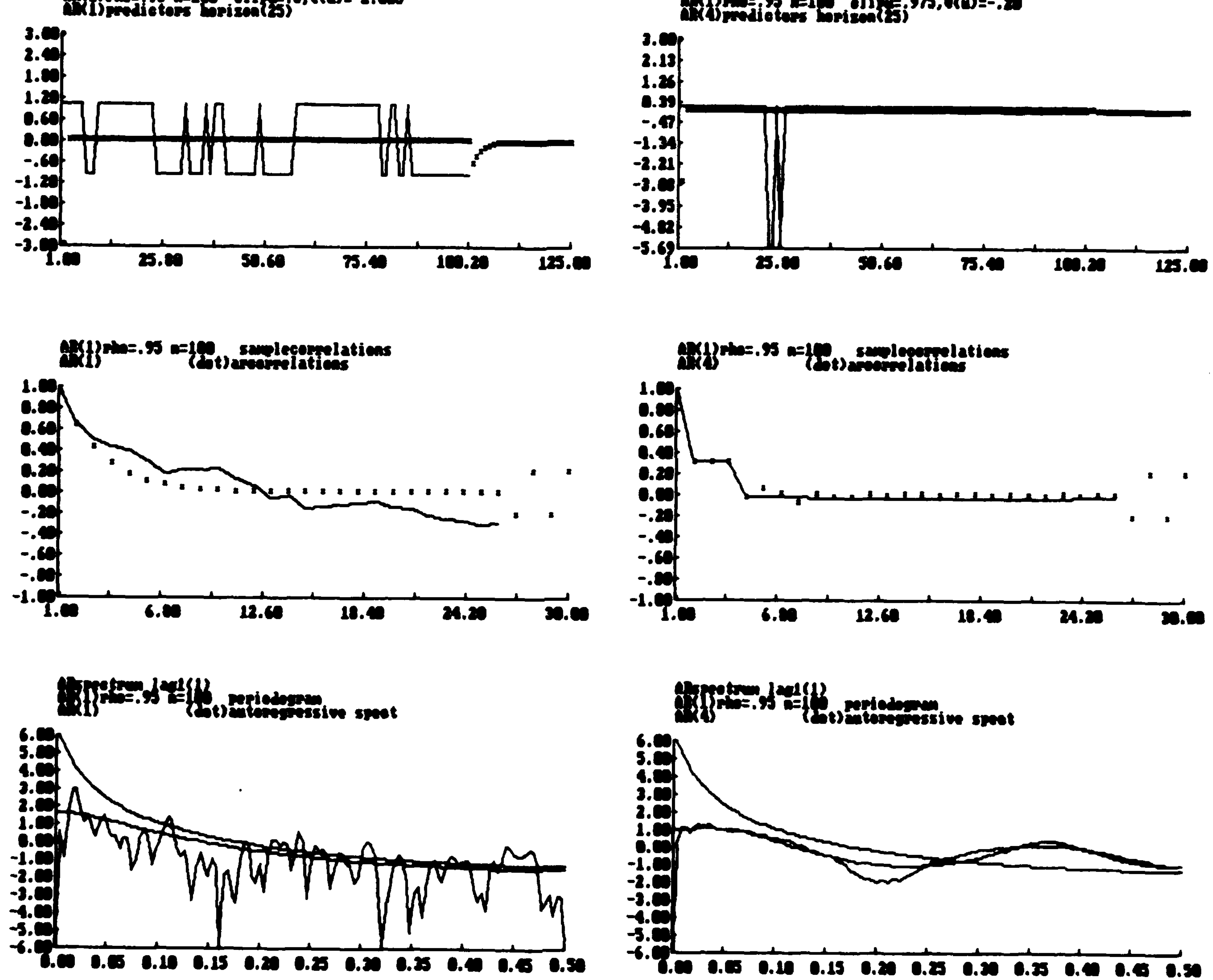
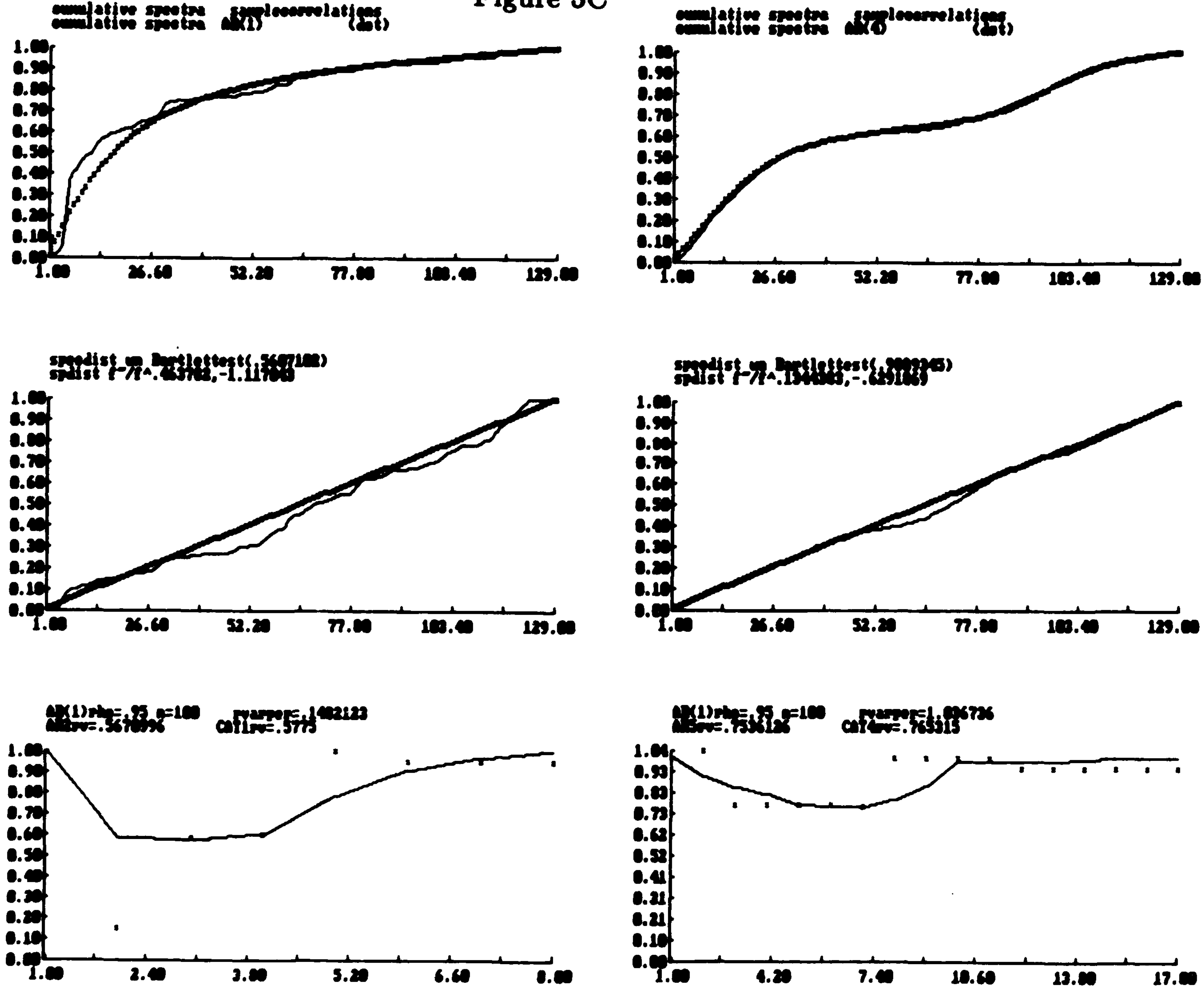

AR(1)rho=.95 n=100 clipu=.5,Q(u)=-2.085
AR(1)rho=.95 n=100 autoregressive proce
AR(1)rho=.95 n=100 clipu=.975,Q(u)=-.20
AR(1)rho=.95 n=100 autoregressive proce
Informative Quantile Plot
AR(1)rho=.95 n=100 clipu=.5,Q(u)=-2.085
Informative Quantile Plot
AR(1)rho=.95 n=100 clipu=.975,Q(u)=-.20
AR(1)rho=.95 n=100 clipu=.5,Q(u)=-2.085
sample Brownian bridge
AR(1)rho=.95 n=100 clipu=.975,Q(u)=-.20
sample Brownian bridge

Figure 5B



AR(1)rho=.95 n=100 clip=.5,0(u)=-2.005
AR(1)predictors horizon(25)
AR(1)rho=.95 n=100 clip=.975,0(u)=-.20
AR(4)predictors horizon(25)
AR(1)rho=.95 n=100 samplecorrelations
AR(1) (dot)arcorrelations
AR(1)rho=.95 n=100 samplecorrelations
AR(4) (dot)arcorrelations
ARspectrum lag1(1)
AR(1)rho=.95 n=100 periodogram
AR(1) (dot)autoregressive spect
ARspectrum lag1(1)
AR(1)rho=.95 n=100 periodogram
AR(4) (dot)autoregressive spect

Figure 5C


cumulative spectra samplecorrelations
cumulative spectra AR(1) (dnt)
1.00
26.60
52.20
77.80
103.40
129.00
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00
2.40
3.80
5.20
6.60
8.00
4.20
7.40
10.60
13.80
17.00

END
11-87
DTIC