stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE STATISTIEK SN 11/83 MEI
(DEPARTMENT OF MATHEMATICAL STATISTICS)

A.J. VAN ES & C. VAN PUTTEN

PROBABILITY PLOTS

STATAL REPORT 3

kruislaan 413 1098 SJ amsterdam

Printed at the Mathematical Centre, Kruislaan 413, Amsterdam, The Netherlands.

The Mathematical Centre, founded 11 February 1946, is a non-profit institution for the promotion of pure and applied mathematics and computer science. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

1980 Mathematics subject classification: Primary: 62-07
Secondary: 69G20, 65S05

Probability plots

by

A.J. van Es & C. van Putten

ABSTRACT

Probability plotting is a technique for examining the underlying distribution function of a sample by means of a graphical analysis of the empirical distribution function. This report gives an outline of the theory of probability plots together with a description of the computer procedure PLOTDIST, designed for making these plots, and numerous examples. PLOTDIST is implemented in the statistical package STATAL.

KEY WORDS & PHRASES: *Probability plot, empirical distribution function*

CONTENTS

1. INTRODUCTION

Generally the choice between using a parametric technique and a non-parametric one is one of the first decisions when analysing experimental data. Having chosen a parametric technique the second decision might be the choice of the underlying distributions of the parametric model which has to be postulated in order to perform an appropriate statistical analysis such as hypothesis testing or estimation. By postulating a model we mean that we assume the data to originate from a specified distribution or family of distributions.

In practical situations the considerations relevant to the choice of the model will be based on experience, on the physical circumstances of the experiment and on the statistical techniques available for the purpose of the experiment. In this setting the choice of the model does not depend on the data to be analysed, as is proper to ensure correct statistical results of the technique chosen. On the other hand a technique based on a badly fitting model, i.e. a model that does not describe the statistical properties of the experiment with sufficient accuracy, may render results that do not have the statistical characteristics guaranteed by this technique in case of a proper model. For instance a test may not have the guaranteed probability of type I error or an estimator may not be unbiased. Evidently one has to check model assumptions and to find the nature of the departures, if any.

Other circumstances in which techniques for checking model assumptions are relevant are when we are examining old data or data which are not to be used as a base for further statistical analysis. For instance when we mentioned experience as a consideration in the choice of a parametric model for a set of data, we should be aware that this experience might include analyses of data of which we have reason to believe that they are statistically similar to the data to be analysed. In this kind of explorative analysis model checking techniques are a possible tool to construct a model.

There is a numerical approach to the problem of checking model assumptions regarding distributional properties and a graphical approach. The numerical approach consists of the computation of statistics such as the Kolmogorov-Smirnov statistic, skewness and kurtosis (cf. BICKEL & DOKSUM [3], section (9.6)), while the graphical approach consists of a graphical analysis of empirical distribution functions (see (1.1)). This graphical technique, by means of so called probability plots, is the subject of this report.

Let us start by introducing some notation and definitions concerning empirical distribution functions, and let us state some of their properties without proving them.

In the sequel $\underline{x}_1,\ldots,\underline{x}_n$ *) denotes a sample from a distribution F, i.e. $\underline{x}_1,\ldots,\underline{x}_n$ are independent and identically distributed with common distribution function F. Our problem of checking model assumptions can now be described as checking whether F belongs to $\mathcal{P}$, a specified family of distribution functions.

Define the empirical distribution function by

(1.1) $\quad \hat{\underline{F}}_n(x) = [\text{number of } i \text{ such that } \underline{x}_i \leq x]/n.$

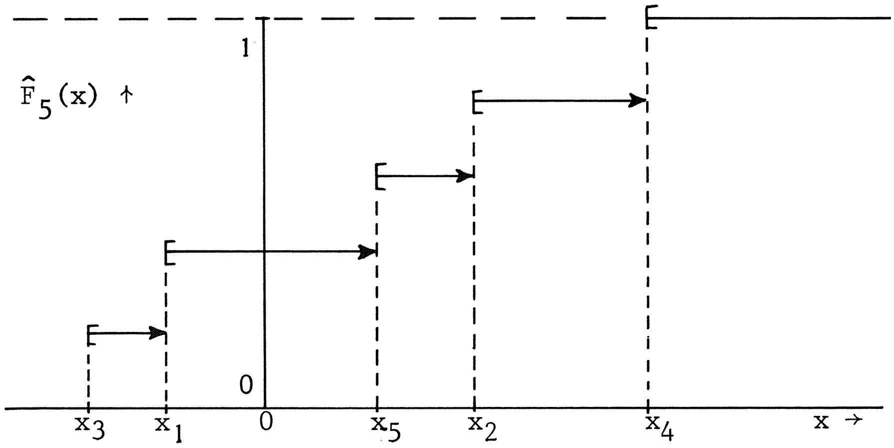


Figure 1. Illustration of $\hat{F}_5(x)$ for a realization $x_1,x_2,\ldots,x_5$ of $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_5$

To stress the difference between F and $\hat{\underline{F}}_n$, F will henceforth be called the theoretical distribution function if there might be any reason for confusion.

$n\hat{\underline{F}}_n(x)$ has a Binomial (n,F(x)) distribution and hence the mathematical expectation and variance are

*) Random variables are underlined; realizations are usually indicated by the same symbol without underlining.

$$(1.2) \qquad E\, \hat{\underline{F}}_n(x) = F(x)$$

and

$$(1.3) \qquad \operatorname{var} \hat{\underline{F}}_n(x) = \frac{1}{n} F(x)(1-F(x)),$$

respectively.

(1.2) shows that for each real x $\hat{\underline{F}}_n(x)$ is an unbiased estimator of F(x) and since (1.3) implies $\lim_{n\to\infty} \operatorname{var} \hat{\underline{F}}_n(x) = 0$ it is consistent.

Apart from considering $\hat{\underline{F}}_n(x)$ as a random variable for a fixed real number x we may consider $\hat{\underline{F}}_n(x)$ as a random function on the real line, since for each realization of $\underline{x}_1,\dots,\underline{x}_n$ $\hat{\underline{F}}_n(x)$ is a function of x. The Glivenko-Cantelli theorem (cf. LOÈVE [9]) states

$$(1.4) \qquad P\{\lim_{n\to\infty} \sup_{-\infty<x<\infty} |\hat{\underline{F}}_n(x)-F(x)| = 0\} = 1,$$

which implies that for large sample size the empirical distribution function is similar to the theoretical distribution function in the sense that with probability one the maximal absolute difference between $\hat{\underline{F}}_n(x)$ and F(x) is small for large n. Hence the empirical distribution function provides us with a means to check whether the theoretical distribution function belongs to $\mathcal{P}$.

A straightforward way of tackling this problem graphically is drawing the graph of $\hat{F}_n$, like in Figure 1, and comparing it with graphs of distributions belonging to $\mathcal{P}$. However the graphs of distributions belonging to $\mathcal{P}$ often are not easily recognizable and comparing them with another graph may not be easy, which makes this method difficult to apply. For some families $\mathcal{P}$, however, by a suitable transformation of the vertical axis (and for some families of the horizontal axis too) the distributions belonging to $\mathcal{P}$ are transformed into straight lines, which evidently makes the method easily applicable since comparing the graph of $\hat{F}_n$ with graphs of distributions belonging to $\mathcal{P}$ reduces to comparing the transformed graph of $\hat{F}_n$ with straight lines. Such a transformed graph of $\hat{F}_n$ is called a probability plot. Probability plots can be constructed for instance for the following families of distributions: the family of uniform distributions, normal distributions, exponential distributions, Gumbel distributions, Laplace distributions, Cauchy distributions and a subset of Weibull distributions, (see (3.7.1)), i.e. for every family of continuous distributions indexed by at most two parameters, one being a location parameter and the other a scale parameter.

$\hat{\underline{F}}_n$ is an estimator of F and therefore, even if F is a member of $\mathcal{P}$, a transformed graph of $\hat{\underline{F}}_n$ can only approximately be straight. By formula (1.3) we see that var $\hat{\underline{F}}_n(x)$ depends on n, x and F. If F belongs to $\mathcal{P}$ we might expect the transformed graph of $\hat{\underline{F}}_n$ to be more straight for large n then for small n, since var $\hat{\underline{F}}_n$ tends to zero as n tends to infinity. The question arises which deviations from a straight line are due to a stochastic sampling error of $\hat{\underline{F}}_n$ as estimator of a theoretical distribution function belonging to $\mathcal{P}$ and which deviations are due to the fact that F is not a member of $\mathcal{P}$ (if this is the case). This problem is closely related to hypothesis testing. Testing the hypothesis H: F = G, with given G, against the alternative K: F $\neq$ G at level α a confidence band with confidence level $1-\alpha$ can be constructed. This band consists of the area between two curves around $\hat{\underline{F}}_n(x)$, one, the upper confidence bound, lies above $\hat{\underline{F}}_n(x)$ and the other, the lower confidence bound, lies below $\hat{\underline{F}}_n(x)$. With probability $1-\alpha$ the band contains the graph of F. Of course, as we did with $\hat{\underline{F}}_n$, after transformation the confidence band can be drawn in a probability plot, providing a means for testing H: F $\in \mathcal{P}$ against K: F $\notin \mathcal{P}$ by rejecting H if the band contains no straight line. Apart from being used for testing F $\in \mathcal{P}$ the confidence band gives an impression as for which values of x we might expect small deviations of $\hat{\underline{F}}_n$ from F (the values of x where the band is narrow) and for which values of x we might expect large deviations (where the band is wide), these deviations being solely caused by the sampling error in the estimate $\hat{F}_n(x)$. Used in this way the confidence band provides an additional tool for a better judgement on whether F belongs to $\mathcal{P}$ or not.

A disadvantage of the technique of probability plots is that it is rather subjective. Even when used in combination with a confidence band the decision whether or not a model fits is left to the taste of the experimenter, with this restriction that one should not accept a distribution, belonging to a family $\mathcal{P}$, as underlying distribution of a model if the confidence band in the probability plot of type $\mathcal{P}$ contains no straight line. One should always realize that the choice between two models which both produce fairly straight lines in their corresponding probability plots may be a very difficult one. It is therefore sensible to be careful when using probability plots to construct a model. An advantage of probability plots compared to goodness of fit test statistics is that such plots contain all the rank properties of a sample, with the effect that for instance outliers can be detected more easily.

Apart from the kind of probability plots described in this report there exists another graphical model checking technique using so-called full normal plots focussing on the examination of the tails of a normal distribution (cf. ANSCOMBE & TUKEY [1] and TUKEY [14]). The problem of comparing two samples can also be treated graphically be means of so-called paircharts (cf. QUADE [12]), P-P plots and Q-Q plots (cf. WILK & GNANADESIKAN [15]).

The purpose of this report is to give a short outline of the theory of probability plots and some related subjects together with a detailed description of the computer procedure PLOTDIST for drawing these plots. PLOTDIST is contained in the statistical package STATAL of the Mathematical Centre.

In chapter 2 we shall give a description of the theory of probability plots and confidence bands. For proofs the reader is referred to the relevant literature. For each type of probability plot a description of the specific properties is given in chapter 3. Chapter 4 contains a description of the procedure PLOTDIST while chapter 5 contains guidelines for the interpretation of probability plots and numerous examples of plots made by PLOTDIST.

Some general references on probability plotting are BURY [4], DOKSUM [5], HEMELRIJK & KRIENS [7] (in Dutch) and WILK & GNANADESIKAN [15].

2. PROBABILITY PLOTTING

In the introduction we presented a rough description of probability plots and claimed that these plots can be constructed for certain families of distributions. In this chapter we shall make this more specific by defining what we mean by a "probability plot" and by defining families of distributions for which these plots can be constructed. These families include many important families of distributions. The third part of this chapter is devoted to confidence bands.

2.1. Probability plots

We shall restrict our attention to continuous distribution functions of the following type:

There exist numbers a and b with $-\infty \le a < b \le \infty$ such that

$$
(2.1.1) \qquad
\begin{array}{lll}
\text{i)} & x \le a & \Rightarrow F(x) = 0 \\
\text{ii)} & a < x < b & \Rightarrow 0 < F(x) < 1 \\
\text{iii)} & x \ge b & \Rightarrow F(x) = 1 \\
\text{iv)} & F \text{ is invertible on } (a,b). &
\end{array}
$$

The interval from a to b, including a and b in case they are finite, is called the support of F.

Let $F^{-1}: (0,1) \to (a,b)$ denote the inverse of F as a function defined on (a,b). We extend F^{-1} to [0,1] by $F^{-1}(0) = a$ and $F^{-1}(1) = b$. With this extended definition of F^{-1} we have

$$
(2.1.2) \qquad F^{-1}(F(x)) = \begin{cases} a & \text{if } x \le a \\ x & \text{if } a < x < b \\ b & \text{if } x \ge b\ . \end{cases}
$$

Now suppose F_0 satisfies the conditions of (2.1.1) with constants a and b and consider the family of distributions

$$
(2.1.3) \qquad \mathcal{P} = \{F \mid F(x) = F_0(\theta_1+\theta_2 x),\ (\theta_1,\theta_2) \in \Theta\},
$$

where Θ is a subset of $\mathbb{R} \times (0,\infty)$. (Note that the definition of $\mathcal{P}$ depends on F_0 and Θ only).

By (2.1.2) we have for all $F \in \mathcal{P}$

$$(2.1.4) \qquad F_0^{-1}(F(x)) = \begin{cases} a & \text{if } x \le (a-\theta_1)/\theta_2 \\ \theta_1 + \theta_2 x & \text{if } (a-\theta_1)/\theta_2 < x < (b-\theta_1)/\theta_2 \\ b & \text{if } x \ge (b-\theta_1)/\theta_2. \end{cases}$$

for some $(\theta_1,\theta_2) \in \Theta$.

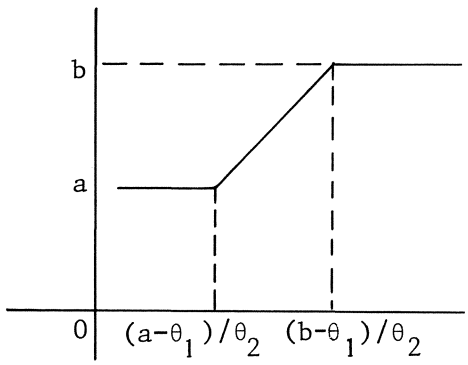


Figure 2. Graph of $F_0^{-1}(F(x))$ for $F \in \mathcal{P}$ corresponding to $(\theta_1,\theta_2) \in \Theta$

Conversely, if F is a distribution function satisfying (2.1.4) for some $(\theta_1,\theta_2) \in \Theta$ then it is easily seen that F belongs to $\mathcal{P}$. Hence the family $\mathcal{P}$ is equal to the family of distributions satisfying (2.1.4) for some $(\theta_1,\theta_2) \in \Theta$.

As mentioned in the introduction our original problem of checking model assumptions is equivalent to checking whether F, the theoretical distribution function of a sample $\underline{x}_1,\ldots,\underline{x}_n$, belongs to a specified family $\mathcal{P}$ of distributions. If this family can be defined by (2.1.3) for suitable F_0 and Θ then by the previous remarks this is equivalent to checking whether $F_0^{-1}F$ satisfies (2.1.4) for some $(\theta_1,\theta_2) \in \Theta$. As proposed in the introduction we now replace F by $\hat{F}_n$, a realization of the empirical distribution function of the sample $\underline{x}_1,\ldots,\underline{x}_n$, and it remains to check whether $F_0^{-1}\hat{F}_n$, being an estimate of $F_0^{-1}F$, approximately satisfies (2.1.4) for some $(\theta_1,\theta_2) \in \Theta$. But if $F = F_0(\theta_1+\theta_2 x) \in \mathcal{P}$, $(\theta_1,\theta_2) \in \Theta$ then $(a-\theta_1)/\theta_2 \le \min(\underline{x}_1,\ldots,\underline{x}_n) \le \max(\underline{x}_1,\ldots,\underline{x}_n) \le (b-\theta_1)/\theta_2$ with probability 1. Since $\hat{F}_n$ is constant outside the interval $[\min(x_1,\ldots,x_n), \max(x_1,\ldots,x_n)]$ it follows that checking whether $F_0^{-1}\hat{F}_n$ approximately satisfies (2.1.4) for some $(\theta_1,\theta_2) \in \Theta$ is equivalent to checking whether $F_0^{-1}\hat{F}_n$ is approximately equal to a straight line $\theta_1+\theta_2 x$ on the interval $[\min(x_1,\ldots,x_n), \max(x_1,\ldots,x_n)]$ for some $(\theta_1,\theta_2) \in \Theta$. To be able to do this we make a plot of $F_0^{-1}\hat{F}_n$. Such a plot is called a probability plot.

DEFINITION.

Let $x_1,\ldots,x_n$ be a realization of a sample $\underline{x}_1,\ldots,\underline{x}_n$. $\hat{F}_n$ is the corresponding realization of the empirical distribution function $\underline{\hat{F}}_n$.

(2.1.5) P is a family of distributions corresponding to a distribution function F_0 by (2.1.3).

A probability plot of type P *is a plot of* $(x, F_0^{-1}\hat{F}_n(x))$ *for values of* x *satisfying* $\min(x_1,\ldots,x_n) \le x \le \max(x_1,\ldots,x_n)$.

It follows that by judging the linearity of the probability plot of type P of $x_1,\ldots,x_n$ we are able to judge whether the theoretical distribution function F of $\underline{x}_1,\ldots,\underline{x}_n$ belongs to P. While doing this we should be aware that since it depends on a realization of a sample a probability plot is subject to random fluctuations. Therefore stochastic arguments should be included in our judgement. These arguments are related to the concept of confidence bands for a distribution function which is to be discussed in section (2.3). Guidelines for the interpretation of probability plots are given in section 5.1.

REMARKS.

1. The families of distributions defined by (2.1.3) include many important families, such as the uniform distributions, the normal distributions and the exponential distributions. In such cases we speak of uniform probability plots, normal probability plots and exponential probability plots etc..
2. Since the presented construction of probability plots depends on the specific type (2.1.3) of the families P we might ask whether probability plots can be constructed for other families. The answer is affirmative. Suppose we want to check whether the theoretical distribution function of a sample $\underline{x}_1,\ldots,\underline{x}_n$ belongs to a family P satisfying the condition:

(2.1.6) There exists a transformation ϕ such that the family of distributions $P' = \{F \mid F$ is the distribution function of $\phi(\underline{x})$ where $\underline{x}$ is a random variable having a distribution function belonging to $P\}$ is of type (2.1.3).

When dealing with families of this type we apply the previous theory to the sample $\phi(\underline{x}_1),\ldots,\phi(\underline{x}_n)$ and check whether the theoretical distribution function of this sample belongs to $\mathcal{P}'$. Actually this is equivalent to transforming the horizontal axis in the probability plot of type $\mathcal{P}'$ of $x_1,\ldots,x_n$. Examples of families satisfying (2.1.6) are the lognormal distributions and families of Weibull distributions with fixed location parameter. (See (3.7.2).)

3. Probability plots can also be used to examine parts of the theoretical distribution function. For instance we might be interested in the tails only. If a probability plot of type $\mathcal{P}$ is approximately linear on an interval then this indicates a resemblance of the theoretical distribution function on this interval to an element of $\mathcal{P}$.

4. Instead of the empirical distribution function defined by (1.1) we use a slightly different one, i.e.

$$(2.1.7)\qquad \hat{\underline{F}}_n(x) = \frac{\underline{m} - 0.3}{n + 0.4},$$

where $\underline{m}$ denotes the number of i such that $\underline{x}_i \leq x$. We prefer this definition because then for each jumppoint x of $\hat{\underline{F}}_n(x)$ the probability of overestimating F(x) is approximately equal to the probability of underestimating F(x) which is not true for the empirical distribution function defined by (1.1) (cf. BENARD & BOS-LEVENBACH [2]).

The different types of probability plot which can be made by the procedure PLOTDIST will be discussed in chapter 3. In this chapter, as an example, we proceed with the construction of probability plots for the family of exponential distributions.

2.2. The exponential probability plot

The construction of probability plots in the previous section depends on the specific distribution F_0 chosen and the set of parameters Θ. In this section we shall discuss this construction for a specific F_0 and two sets Θ_1 and Θ_2, resulting in a so called exponential probability plot. For a discussion on normal probability plots the reader is referred to HEMELRIJK &

KRIENS [7] (in Dutch).

Let F_0 be the standard exponential distribution function, i.e.

$$(2.2.1) \qquad F_0(x) \doteq \begin{cases} 0 & \text{if } x \le 0 \\ 1 - \exp(-x) & \text{if } x > 0 \,. \end{cases}$$

F_0 is of the type (2.1.1) with $a = 0$ and $b = \infty$.

Let Θ_1 and Θ_2 be defined as

$$(2.2.2) \qquad \Theta_1 = \{0\} \times (0,\infty), \quad \Theta_2 = \mathbb{R} \times (0,\infty).$$

As in the previous section we shall define two families of distributions P_1 and P_2 corresponding to the parameter sets Θ_1 and Θ_2. Since for $(\theta_1,\theta_2) \in \Theta_i$, $i = 1,2$

$$(2.2.3) \qquad F_0(\theta_1+\theta_2 x) = \begin{cases} 0 & \text{if } x \le -\theta_1/\theta_2 \\ 1-\exp(-\theta_1-\theta_2 x) & \text{if } x > -\theta_1/\theta_2 \end{cases}$$

the families P_1 and P_2, defined by (2.1.3) are equal to

$$(2.2.4) \qquad P_1 = \{F \mid F(x) = (1-\exp(-\theta_2 x))I_{(0,\infty)}(x),\ \theta_2 \in (0,\infty)\}$$

and

$$(2.2.5) \qquad P_2 = \{F \mid F(x) = (1-\exp(-\theta_1-\theta_2 x))I_{(-\theta_1/\theta_2,\infty)}(x),\ \theta_1 \in \mathbb{R},\ \theta_2 \in (0,\infty)\},$$

where I_A is a function which is 1 on A and 0 outside A.

It follows that P_1 is the family of exponential distribution functions with threshold parameter equal to 0 and that P_2 is the family of exponential distributions with arbitrary threshold parameter.

With

$$(2.2.6) \qquad F_0^{-1}(y) = \begin{cases} -\ln(1-y) & \text{if } 0 \le y < 1 \\ \infty & \text{if } y = 1 \end{cases}$$

we have by (2.1.4) for $F = F_0(\theta_1+\theta_2 x)$, $(\theta_1,\theta_2) \in \Theta_i$, $i = 1,2$

$$(2.2.7) \qquad F_0^{-1}F(x) = \begin{cases} 0 & \text{if } x \le -\theta_1/\theta_2 \\ \theta_1+\theta_2 x & \text{if } x > -\theta_1/\theta_2 . \end{cases}$$

By (2.1.5) an exponential probability plot is a plot of $-\ln(1-\hat{F}_n(x))$ for values of x satisfying $\min(x_1,\ldots,x_n) \le x \le \max(x_1,\ldots,x_n)$ where $x_1,\ldots,x_n$ is a realization of a sample $\underline{x}_1,\ldots,\underline{x}_n$ and $\hat{F}_n$ the corresponding realization of the empirical distribution function $\hat{\underline{F}}_n$.

2.3. Confidence bands

This section contains a discussion on confidence bands for a distribution function and their relation to hypothesis testing. Most of the remarks are from DOKSUM [5].

The concept of confidence bands and the problem of testing the hypothesis H: $F = G$ against alternatives K: $F \neq G$, in case $\underline{x}_1,\ldots,\underline{x}_n$ is a sample from a distribution F, are closely related. Using a test statistic for testing H against K we are able to derive a confidence band for F, in a way which will be presented in this section for two specific test statistics, and conversely starting with a confidence band a testing procedure for H against K can be derived.

For instance consider the Kolmogorov-Smirnov test statistic

$$(2.3.1) \qquad \underline{d}_n = \max_x \operatorname{abs}(\hat{\underline{F}}_n(x) - F(x)).$$

Under the hypothesis that $\underline{x}_1,\ldots,\underline{x}_n$ is a sample from the distribution F $\underline{d}_n$ has the same distribution for all continuous distributions F (cf. BICKEL & DOKSUM [3], p. 379) and hence for given $\alpha \in (0,1)$ there exists a critical constant $k(n,\alpha)$, independent of F, such that

$$(2.3.2) \qquad P\{\underline{d}_n \le k(n,\alpha)\} = 1 - \alpha.$$

This probability is equal to

$$(2.3.3) \qquad P\{\hat{\underline{F}}_n(x) - k(n,\alpha) \le F(x) \le \hat{\underline{F}}_n(x) + k(n,\alpha), \text{ for all real } x\},$$

implying that with probability $1 - \alpha$ the theoretical distribution function F is contained in the area between $\underline{H}^-$ and $\underline{H}^+$ where

$$(2.3.4) \qquad \underline{H}^{\pm}(x) = \hat{\underline{F}}_n(x) \pm k(n,\alpha).$$

Next, let us consider the test statistic

$$(2.3.5) \qquad \underline{w}_n = \max_{x \in S} \operatorname{abs}(\hat{\underline{F}}_n(x) - F(x))/(F(x)(1-F(x))^{\frac{1}{2}},$$

S being the support of F, defined in (2.1.1). Under the hypothesis that $\underline{x}_1,\ldots,\underline{x}_n$ is a sample from a distribution F the distribution of $\underline{w}_n$ is equal for all continuous F and a critical constant $k(n,\alpha)$ satisfying

$$(2.3.6) \qquad P\{\underline{w}_n \le k(n,\alpha)\} = 1 - \alpha$$

can be obtained (see the references in DOKSUM [5]). This probability is equal to

$$(2.3.7) \qquad P\{\underline{H}^-(x) \le F(x) \le \underline{H}^+(x), \text{ for all } x \in S\} = 1 - \alpha,$$

where $\underline{H}^-$ and $\underline{H}^+$ are defined by

$$(2.3.8) \qquad \underline{H}^{\pm}(x) = [\hat{\underline{F}}_n(x) + \tfrac{1}{2}\, k(n,\alpha)^2 \pm k(n,\alpha)\{\hat{\underline{F}}_n(x)(1-\hat{\underline{F}}_n(x)) + \tfrac{1}{4}\, k(n,\alpha)^2\}^{\frac{1}{2}}]/ (1+k(n,\alpha)^2).$$

As before it follows that with probability $1 - \alpha$ the theoretical distribution F is contained in the area between $\underline{H}^-$ and $\underline{H}^+$.

In both cases ((2.3.4) and (2.3.8)) the area between $\underline{H}^-$ and $\underline{H}^+$ is called a level $1-\alpha$ confidence band, $\underline{H}^-$ and $\underline{H}^+$ are called the lower and upper confidence bound.

Roughly speaking the difference between $\underline{d}_n$ and $\underline{w}_n$ as test statistics for the hypothesis H: F = G is that $\underline{d}_n$ gives equal weight to the differences $\hat{\underline{F}}_n(x) - F(x)$ for each real x, irrespective of the variance of these differences (see (1.2) and (1.3)), while $\underline{w}_n$ measures the differences in units equal to the standard deviation of $\hat{\underline{F}}_n(x) - F(x)$. This reflects in the properties of the corresponding confidence bands. The band (2.3.8) based on $\underline{w}_n$ is narrower in the tails than the band (2.3.4) based on $\underline{d}_n$ and is preferable unless one is particularly interested in the center of the distribution.

A confidence band can be drawn in a probability plot after transformation by F_0^{-1}, i.e. $F_0^{-1}H^-$ and $F_0^{-1}H^+$ can be drawn in the plot. The band (2.3.8) is the one used by the procedure PLOTDIST.

Now, conversely, we shall discuss a test procedure for the hypothesis H: $F \in \mathcal{P}$ against K: $F \notin \mathcal{P}$, $\mathcal{P}$ being a class of distributions suitable for probability plotting, using a confidence band in a probability plot of type $\mathcal{P}$. Consider the test procedure

(2.3.9) *Reject* H: $F \in \mathcal{P}$ *if the confidence band in the probability plot of type* $\mathcal{P}$ *of the sample* $\underline{x}_1,\ldots,\underline{x}_n$ *contains no straight line.*

From (2.3.7) it follows that under the hypothesis H: $F \in \mathcal{P}$ the probability of rejection satisfies

$$
\begin{aligned}
&P\{\text{the band contains no straight line } | \text{H is true}\} \\
&\le P\{\text{the band does not contain the line corresponding to F} | \text{H is true}\} \\
&= 1 - P\{\text{the band contains the line corresponding to F} | \text{H is true}\} \\
&= 1 - (1-\alpha) = \alpha .
\end{aligned} \tag{2.3.10}
$$

Hence the probability of type I error, i.e. rejection of H when H is true, is at most α. The procedure is conservative and not very powerful, as can be seen from the plots in subsections (5.2.3.2) and (5.2.3.3). It may therefore be advisable to use other, more specific, tests, if available.

A confidence band can also be used as an aid for the interpretation of probability plots. This and other guidelines will be discussed in section 5.1.

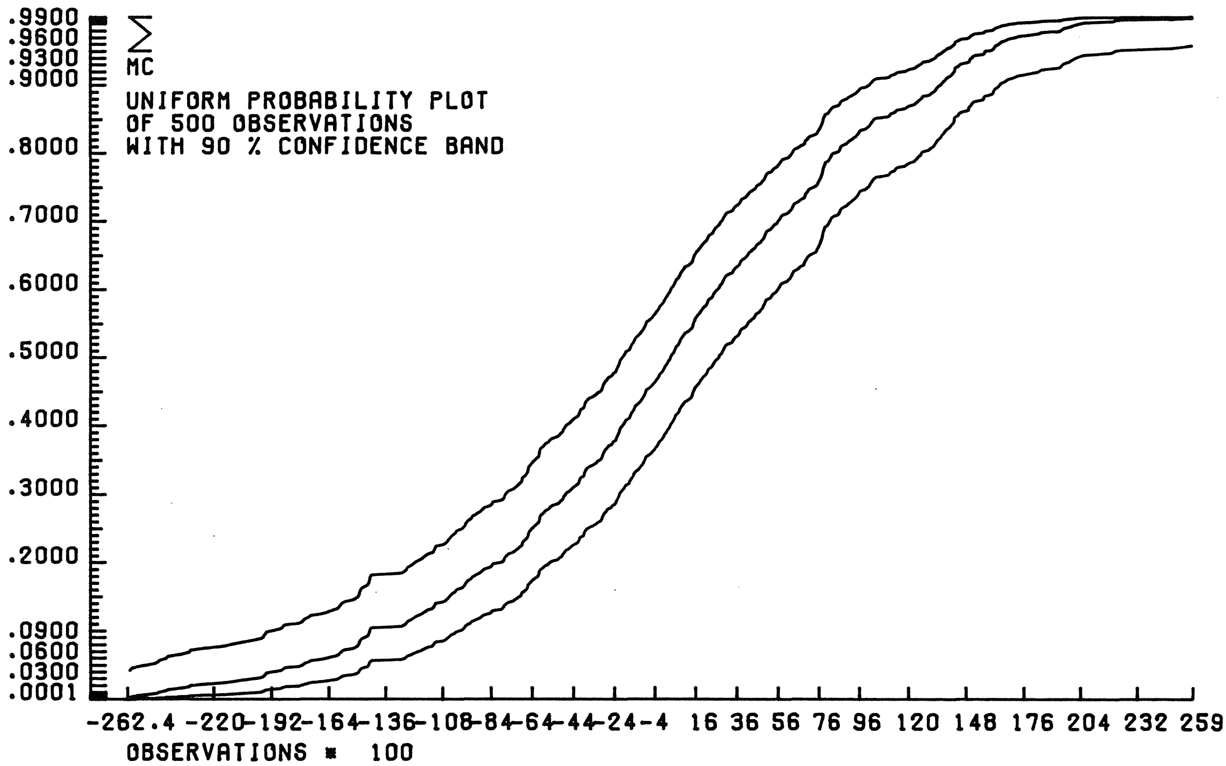


Figure 3. Sample from a standard normal distribution

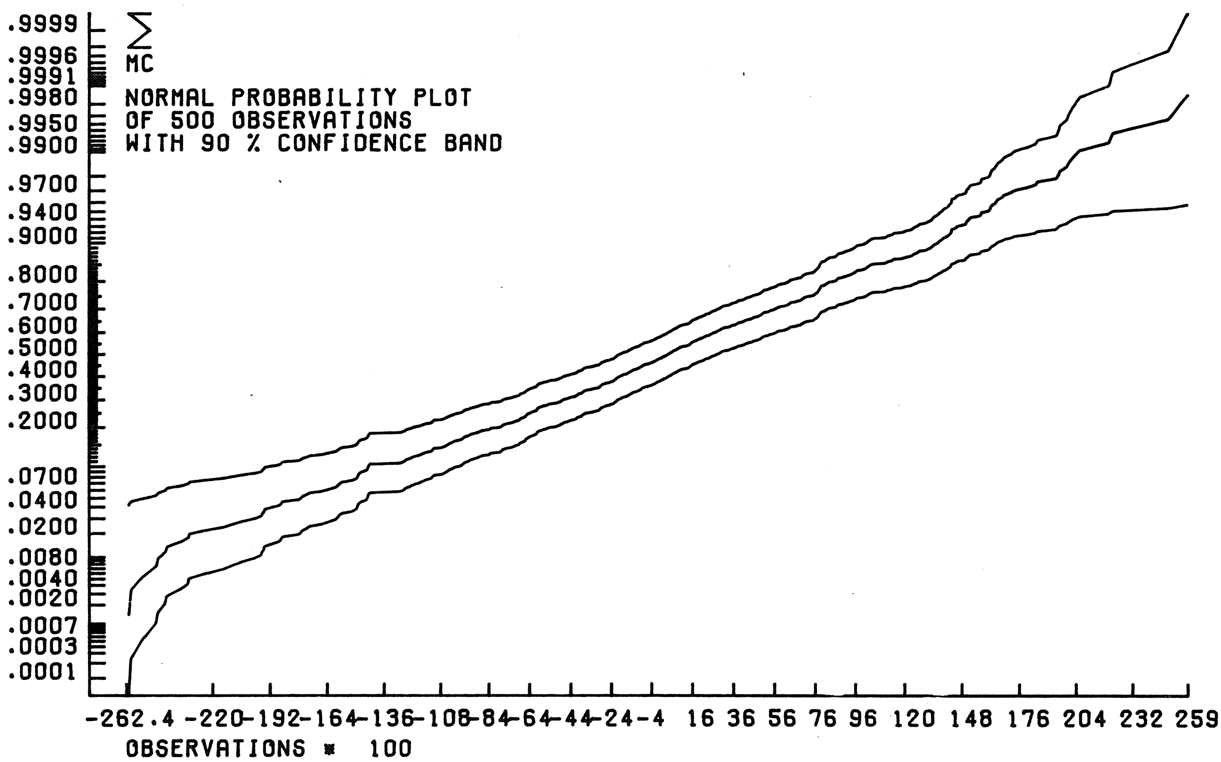


Figure 4. Sample from a standard normal distribution

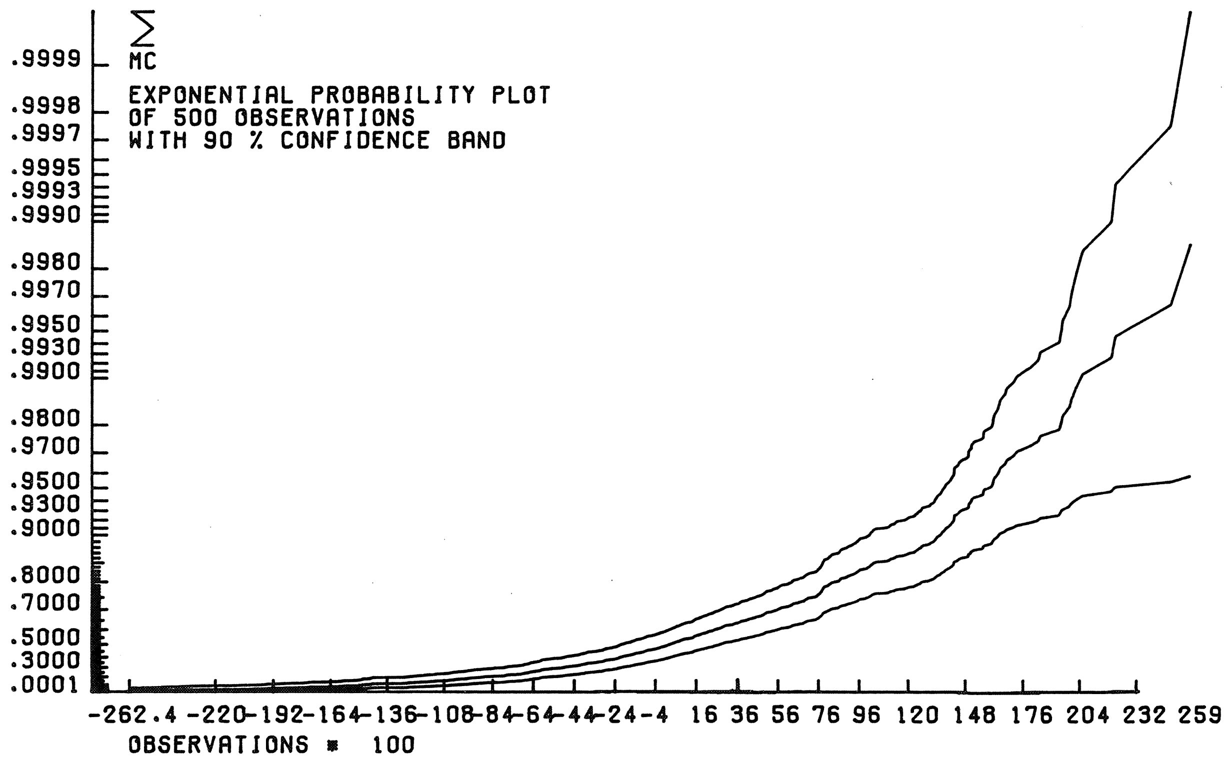


Figure 5. Sample from a standard normal distribution

Figures 3,4 and 5 contain three probability plots of one sample of size 500 from a standard normal distribution. In the plots a 90% confidence band, i.e. a confidence band for α equal to 10%, is drawn. The inner curve in the plots is the transformed empirical distribution function, while the two outer curves are the transformed lower and upper confidence bound. The jumppoints of the transformed empirical distribution function and of the transformed confidence bounds are connected by straight lines.

Figure 3 shows a uniform probability plot of the sample, which is equal to a plot of the untransformed empirical distribution function and untransformed confidence band since for the family of uniform distributions F_0^{-1} equals the identical mapping of [0,1] into the real line (see section (3.1)). Figure 4 contains a normal probability plot of the sample and shows the expected linearity, while in Figure 5, an exponential probability plot of the sample, there is no such linearity. Remark that in Figures 3 and 5 no straight line can be drawn in the confidence band and that therefore the test procedure (2.3.9) correctly rejects both a uniform and an exponential distribution as the theoretical distribution of the sample.

3. DIFFERENT TYPES

The computer procedure PLOTDIST is designed to construct probability plots of seven different types. For each of these types several options are available, which are described in chapter 4 of this report and in the STATAL manual [15]. These options consist of a confidence band, a straight line for reference and layout options. Clearly there are characteristics of the procedure which depend on the requested type of plot, such as the transformation F_0^{-1} used in the construction of the plot (see section (2.1)) and the procedure for drawing the reference line. This chapter is devoted to the description of such features of PLOTDIST, in contrast to the preceding chapter which was devoted to the general theory.

Before describing the type-dependent features of PLOTDIST let us first discuss the reference line. As is seen in section (2.1) an empirical distribution function plotted in a probability plot of appropriate type should resemble a straight line, since the underlying theoretical distribution function corresponds with a straight line in this plot. An estimate of this line thus yields an estimate of the underlying distribution, which in its turn can be used to estimate the parameters of this distribution. This is not the aim of the straight line drawn by PLOTDIST (on request), which is meant merely as reference line. It enables the user to compare the transformed empirical distribution function with a straight line without drawing one himself. Bearing this in mind we have chosen for quick, and therefore sometimes rough, procedures for drawing the line which do not pretend to yield the "best" line possible.

The type-dependent features of PLOTDIST will be discussed in the following seven sections each devoted to one type. For each type we mention

- the distribution function F_0 used in (2.1.3)
- the inverse of this function, F_0^{-1}
- the parameter space Θ used in (2.1.3)*)
- the procedure for the reference line.

Since the family of Weibull distributions is indexed by three parameters it will be treated more extensively.

In the sequel we shall use the following notation

$$x_{min} = \min_{i=1,\ldots,n} x_i \quad \text{, the sample minimum}$$

$$x_{max} = \max_{i=1,\ldots,n} x_i \quad \text{, the sample maximum}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{, the sample mean}$$

$$s = \{(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2)/(n-1)\}^{\frac{1}{2}} \quad \text{, the sample standard deviation.}$$

3.1. Uniform The general form of a Uniform (a,b) ($-\infty<a<b<\infty$) distribution function F is

$$F(x) = \begin{cases} 0 & \text{if } x < a \\ (x-a)/(b-a) & \text{if } a \le x < b \\ 1 & \text{if } x \ge b. \end{cases}$$

F_0 is the Uniform (0,1) distribution function

*) Note that the parameter $(\theta_1,\theta_2) \in \Theta$ are not the usual parameters of the distributions in this section.

$$F_0(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \le x < 1 \\ 1 & \text{if } x \ge 1. \end{cases}$$

Its inverse F_0^{-1} is

$$F_0^{-1}(y) = y \qquad y \in [0,1].$$

$$\Theta = \mathbb{R} \times (0,\infty).$$

The reference line is drawn through the points $(\hat{c} - \hat{\ell}, 0)$ and $(\hat{c} + \hat{\ell}, 1)$, where

$$\hat{c} = \tfrac{1}{2}(x_{min} + x_{max})$$

$$\hat{\ell} = \tfrac{1}{2}(x_{max} - x_{min})(n+1)/(n-1).$$

$\underline{\hat{c}}$ and $\underline{\hat{\ell}}$ are the best linear unbiased estimators of (a+b)/2, the center of the interval (a,b) and b−a, the length of (a,b) (cf. JOHNSON & KOTZ [8], 2, p.60).

3.2. Normal

The general form of the Normal (μ,σ) $(\mu \in \mathbb{R},\ \sigma > 0)$ distribution function F is

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\tfrac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right)dt, \qquad -\infty < x < \infty.$$

F_0 is the Normal (0,1) distribution function

$$F_0(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\tfrac{1}{2}t^2)dt, \qquad -\infty < x < \infty.$$

There is no explicit formula for the inverse of F_0, which therefore has to be approximated numerically.

$$\Theta = \mathbb{R} \times (0,\infty).$$

The reference line is drawn through the points $(\bar{x},0)$ and $(\bar{x}+s,1)$. $\underline{\bar{x}}$ and $\underline{s}$ are the usual estimators of μ and σ (cf. JOHNSON & KOTZ [8], 1, p.59).

3.3. Exp 1 and Exp 2

The general form of the Exponential (μ,λ) $(\mu \in \mathbb{R}, \lambda > 0)$ distribution function F is

$$F(x) = \begin{cases} 0 & \text{if } x < \mu \\ 1 - \exp(-\lambda(x-\mu)) & \text{if } x \geq \mu. \end{cases}$$

The exponential probability plot is discussed in section (2.2).

In case of Exp 1 the reference line is drawn through the points (0,0) and $(\bar{x},1)$, while in case of Exp 2 it is drawn through $(\hat{\mu},0)$ and $(\bar{x},1)$, where

$$\hat{\mu} = (nx_{min} - \bar{x})/(n-1).$$

$\underline{x}_{min}$ and $\underline{\bar{x}} - \underline{x}_{min}$ are the maximum likelihood estimators of μ and $1/\lambda$.

From E $\underline{x}_{min} = \mu + 1/n\lambda$ and E $\underline{\bar{x}} = \mu + 1/\lambda$ it follows that

$$\begin{aligned} E\,\underline{\hat{\mu}} &= E\,(n\underline{x}_{min} - \underline{\bar{x}})/(n-1) \\ &= \{n(\mu + 1/n\lambda) - (\mu + 1/\lambda)\}/(n-1) \\ &= \mu. \end{aligned}$$

Hence $\underline{\hat{\mu}}$ is an unbiased estimator of μ (cf. JOHNSON & KOTZ [8], 1, p.211).

3.4. Gumbel

The general form of the Gumbel (μ,σ) $(\mu \in \mathbb{R}, \sigma > 0)$ distribution function F is

$$F(x) = \exp(-\exp(-(x-\mu)/\sigma)) \quad , -\infty < x < \infty .$$

F_0 is the Gumbel (0,1) distribution function

$$F_0(x) = \exp(-\exp(-x)), \qquad -\infty < x < \infty.$$

Its inverse F_0^{-1} is

$$F_0^{-1}(y) = \begin{cases} -\infty & \text{if } y = 0 \\ -\ln(-\ln(y)) & \text{if } 0 < y < 1 \\ +\infty & \text{if } y = 1. \end{cases}$$

$\Theta = \mathbb{R}x(0,\infty)$

The reference line is drawn through the points $(\hat{\mu},0)$ and $(\hat{\mu}+\hat{\sigma},1)$ where

$$\hat{\mu} = x - \gamma\sqrt{6}\ s/\pi$$
$$\hat{\sigma} = \sqrt{6}\ s/\pi$$

and γ is the Euler constant ($\gamma \approx 0.5772156649$). $\hat{\underline{\mu}}$ and $\hat{\underline{\sigma}}$ are estimators of μ and σ (cf. JOHNSON & KOTZ [8], 1, p.285).

3.5. Laplace

The general form of the Laplace (μ,σ) $(\mu \in \mathbb{R}, \sigma > 0)$ distribution function F is

$$F(x) = \begin{cases} \frac{1}{2}\exp((x-\mu)/\sigma) & \text{if } x \le \mu \\ 1 - \frac{1}{2}\exp(-(x-\mu)/\sigma) & \text{if } x > \mu. \end{cases}$$

Its inverse F_0^{-1} is

$$F_0^{-1}(y) = \begin{cases} -\infty & \text{if } y = 0 \\ \ln(2y) & \text{if } 0 < y \le \frac{1}{2} \\ -\ln(2-2y) & \text{if } \frac{1}{2} < y < 1 \\ +\infty & \text{if } y = 1. \end{cases}$$

$\Theta = \mathbb{R}\,x\ (0,\infty)$.

The reference line is drawn through the points $(\bar{x},0)$ and $(\bar{x} + s/\sqrt{2},1)$.

The expected value and standard deviation of the Laplace (μ,σ) distribution are μ and $\sqrt{2}\,\sigma$. $\bar{\underline{x}}$ and $\underline{s}/\sqrt{2}$ are used as estimators of μ and σ (cf. JOHNSON & KOTZ [8], 2, p.23).

3.6. Cauchy

The general form of the Cauchy (μ,σ) $(\mu \in \mathbb{R}, \sigma > 0)$ distribution function F is

$$F(x) = \tfrac{1}{2} + \frac{1}{\pi}\tan^{-1}((x-\mu)/\sigma), \qquad -\infty < x < \infty.$$

F_0 is the Cauchy (0,1) distribution function

$$F_0(x) = \tfrac{1}{2} + \frac{1}{\pi}\tan^{-1}(x), \qquad -\infty < x < \infty.$$

Its inverse F_0^{-1} is

$$F_0^{-1}(y) = \begin{cases} -\infty & \text{if } y = 0 \\ \tan((y-\tfrac{1}{2})\pi) & \text{if } 0 < y < 1 \\ +\infty & \text{if } y = 1\,. \end{cases}$$

$$\Theta = \mathbb{R} \times (0,\infty).$$

The reference line is the least squares line determined by nine points of the transformed empirical distribution function described in subsection (3.6.1).

3.6.1. The least squares line

$(x_1^*,f_1^*),\ldots,(x_9^*,f_9^*)$ are defined as

x_m^* = the smallest observation x_i such that $\hat{F}_n(x_i) \geq m/10$

$f_m^* = F_0^{-1}(\hat{F}_n(x_m^*))$

$(m = 1,2,\ldots,9)$.

The line drawn by PLOTDIST is the least squares line determined by these nine points of the transformed empirical distribution function.

3.7. Weibull2 and Weibull3

The general form of the Weibull (μ,σ,c) $(\mu \in \mathbb{R},\ \sigma > 0,\ c > 0)$ distribution F is

$$F(x) = \begin{cases} 0 & \text{if } x \le \mu \\ 1 - \exp(-\{(x-\mu)/\sigma\}^c) & \text{if } x > \mu \end{cases}$$

Since the technique of probability plots is not designed for a three parameter family, such as the family of the Weibull distributions, it cannot be applied directly without further knowledge about the parameters. However if the value of c is known a Weibull probability plot (dependent on c) can be constructed using the method of chapter 2. If the value of μ is known the data can be transformed in such a way that they become Gumbel distributed after which they can be plotted in a Gumbel probability plot. If neither μ nor c is known μ can be estimated and then the same procedure as with known μ can be applied.

3.7.1. The value of c is known

We apply the theory of chapter 2 with F_0 equal to the Weibull $(0,1,c)$ distribution function.

$$F_0(x) = \begin{cases} 0 & \text{if } x \le 0 \\ 1 - \exp(-x^c) & \text{if } x > 0. \end{cases}$$

Its inverse F_0^{-1} is

$$F_0^{-1}(y) = \begin{cases} 0 & \text{if } y = 0 \\ \{-\ln(1-y)\}^{1/c} & \text{if } 0 < y < 1 \\ \infty & \text{if } y = 1. \end{cases}$$

The reference line is the least squares line of section (3.6.1). Remark that the transformation F_0^{-1} depends on c.

3.7.2. The value of μ is known

If the value of μ is known and if $\underline{x}_1,\ldots,\underline{x}_n$ is a sample from a Weibull (μ,σ,c) distribution then $-\ln(\underline{x}_1-\mu),\ldots,-\ln(\underline{x}_n-\mu)$ is a sample from a Gumbel $(-\ln\sigma,\ 1/c)$ distribution which can be plotted in a Gumbel probability plot (see section 3.4)).

3.7.3. Neither μ nor c is known

If both μ and c are unknown, μ can be estimated first, after which the data can be transformed and treated as in (3.7.2) where μ is assumed to be known.

μ can be estimated using the method of moments.
If $\underline{x}_1,\ldots,\underline{x}_n$ is a sample from a Weibull (μ,σ,c) distribution then

$$\begin{aligned} E\underline{\bar{x}} &= \mu + \sigma\Gamma(1+1/c) \\ E\underline{x}_{min} &= \mu + \sigma n^{-1/c}\Gamma(1+1/c) \\ q_\alpha &= \mu + \sigma(-\ln(1-\alpha))^{1/c}, \end{aligned}$$

where Γ denotes the gamma function and q_α is defined by $F(q_\alpha) = \alpha$, F being the Weibull (μ,σ,c) distribution function (cf. JOHNSON (KOTZ [8], 1, pp. 250-271).
Estimates $\hat{\mu},\hat{\sigma}$ and $\hat{c}$ are obtained from a solution of the following three equations

$$\begin{aligned} \bar{x} &= \hat{\mu} + \hat{\sigma}\Gamma(1+1/\hat{c}) \\ x_{min} &= \hat{\mu} + \hat{\sigma}\, n^{-1/2}\Gamma(1+1/\hat{c}) \\ x_{(i)} &= \hat{\mu} + \hat{\sigma}(-\ln(1 - i/(n+1))^{1/\hat{c}}\ , \end{aligned}$$

where $x_{(i)}$ denotes a realization of the ith order statistic and i is chosen in such a way that a numerical solution of the previous three equations is likely to be found.

To be more precise, i = entier (0.4n) unless $\mathrm{abs}(x_{(i)}-\bar{x})/(\bar{x}-x_{min}) < 0.1$ in which case i = entier (0.9n). Here, entier(a) denotes the greatest integer less than or equal to a and abs(a) denotes the absolute value of a. ($n \geq 3$ is an obvious condition for this to make any sense.)

After elimination of $\hat{\mu}$ and $\hat{\sigma}$ from the equations, $\hat{\theta} = 1/\hat{c}$ is seen to be a zero of the function Φ, defined by

$$\Phi(\theta) = \{x_{min}-x_{(i)}+(x_{(i)}-\bar{x})/n^{\theta}\}\Gamma(1+\theta)+(\bar{x}-x_{min})q^{\theta}_{(i)},$$

where $q_{(i)} = -\ln(1-i/(n+1))$.

Some properties of this function are the following

1. $\lim\limits_{\theta\downarrow 0} \Phi(\theta) = 0;$

2. $\lim\limits_{\theta\to\infty} \Phi(\theta) = -\infty.$

The procedure PLOTDIST tries to compute one or two zeros, according to whether $\Phi(0.001)$ is positive or not. If two zeros are found, the one for which the estimate of the median $\hat{\mu} + \hat{\sigma}\,(\ln(2))^{\hat{\theta}}$ is closest to the sample median is chosen.

4. DOCUMENTATION ON THE PROCEDURE PLOTDIST

This chapter contains the description of the procedure PLOTDIST, as it is to be found in the STATAL manual [13].

TITLE: PLOTDIST

AUTHORS: A.J. VAN ES, C. VAN PUTTEN, I. VAN DER TWEEL

INSTITUTE: MATHEMATICAL CENTRE

RECEIVED: 830301

BRIEF DESCRIPTION:

A PROBABILITY PLOT OF REQUESTED TYPE AND SIZE IS PLOTTED VIA A PLOTTER. ON REQUEST THE PLOT CONTAINS A CONFIDENCE BAND OR AN ESTIMATED STRAIGHT LINE FOR REFERENCE. ENLARGEMENTS OF PARTS OF THE PLOT ARE POSSIBLE.

KEYWORDS:

EMPIRICAL DISTRIBUTION FUNCTION, PROBABILITY PLOT, CONFIDENCE BAND

CALLING SEQUENCE:

HEADING:
```
"PROCEDURE" PLOTDIST(GRFILE,V, LV, UV, TYPE, LB, UB, MODE,
    PART,SIZE,OPTION, BETA, SORTED, PAR, IDENT);
"VALUE" V, LV, UV, LB, UB, MODE, PART, SIZE, OPTION, BETA,
    SORTED, PAR;
"INTEGER" LV, UV, MODE, PART, SIZE, OPTION;
"REAL" LB, UB, BETA, PAR;
"BOOLEAN" SORTED;
"ARRAY" V;
"STRING" GRFILE, TYPE, IDENT;
"CODE" 47005;
```

FORMAL PARAMETERS:

GRFILE: <STRING>,NAME OF THE FILE ON WHICH THE PLOT MUST BE WRITTEN AS A MAINGRAPH. IF THE STRING GRFILE IS EMPTY THEN THE NAME OF THIS FILE IS GRFILE. IN SUBSEQUENT CALLS OF PLOTDIST THE SAME VALUE MAY BE CHOSEN FOR GRFILE;

V: <ARRAY IDENTIFIER>, V[LV],..., V[UV] IS A VECTOR CONTAINING THE SAMPLE;

LV: <INTEGER ARITHMETIC EXPRESSION>, SMALLEST INDEX OF THE SAMPLE ARRAY;

UV: <INTEGER ARITHMETIC EXPRESSION>, GREATEST INDEX OF THE SAMPLE ARRAY;

TYPE: <STRING>, TYPE OF PROBABILITY PLOT. TYPE SHOULD CONTAIN ONE OF THE FOLLOWING IDENTIFIERS: UNIFORM, NORMAL, EXP1, EXP2, LAPLACE, GUMBEL, CAUCHY, WEIBULL2, WEIBULL3;

LB, UB: <REAL ARITHMETIC EXPRESSION>, LOWER AND UPPER BOUND, RESPECTIVELY, FOR THE POSSIBLE ENLARGEMENT. THE PLOT CONTAINS THE EMPIRICAL DISTRIBUTION FUNCTION AND CONFIDENCE BAND (WHEN REQUESTED) IN THOSE ARGUMENTS FOR WHICH THE EMPIRICAL DISTRIBUTION FUNCTION IS GREATER THAN OR EQUAL TO LB AND LESS THAN OR EQUAL TO UB. IF NO ENLARGEMENT IS DESIRED LB SHOULD BE TAKEN EQUAL TO 0 AND UB EQUAL TO 1;

MODE: <INTEGER ARITHMETIC EXPRESSION>, IF MODE = -1 THE JUMPS OF THE EMPIRICAL DISTRIBUTION FUNCTION WILL BE CONNECTED BY STRAIGHT LINES. IF MODE > 0 SYMBOLS WITH INTEGER REPRESENTATION MODE (SEE TABLE 1) WILL BE PLOTTED AT THE JUMPS;

PART: <INTEGER ARITHMETIC EXPRESSION>, IF PART = 0 ALL JUMPS OF THE EMPIRICAL DISTRIBUTION FUNCTION WILL BE PLOTTED. IF PART > 0 THE EMPIRICAL DISTRIBUTION FUNCTION WILL BE PLOTTED IN PART + 1 EQUIDISTANT ARGUMENTS;

SIZE: <INTEGER ARITHMETIC EXPRESSION>, SIZE OF THE PLOT. THERE ARE TWO POSSIBILITIES FOR THE PARAMETER SIZE. IF SIZE EQUALS 0, 1, 2, 3, 4 OR 5 THE FORMAT OF THE PLOT WILL BE REPORT FORMAT (I.E. THE PLOT WILL FIT ON A PAGE OF A MATHEMATICAL CENTRE REPORT), A1, A2, A3, A4 OR A5, RESPECTIVELY. THE OTHER POSSIBILITY IS THAT SIZE EQUALS A NONNEGATIVE INTEGER OF SIX DIGITS, OF WHICH THE FIRST THREE INDICATE THE WIDTH OF THE PLOT, WHILE THE LAST THREE INDICATE THE HEIGHT (IN MM). THE MINIMUM SIZE ALLOWED IS 100100. THE HEIGHT OF THE PLOT SHOULD BE LESS THAN OR EQUAL TO 725 MM;

OPTION: <INTEGER ARITHMETIC EXPRESSION>, INDICATING WHETHER A CONFIDENCE BAND OR A STRAIGHT LINE FOR REFERENCE HAS TO BE PLOTTED;
OPTION = 11: BOTH THE BAND AND THE LINE ARE PLOTTED
OPTION = 10: ONLY THE BAND IS PLOTTED
OPTION = 01: ONLY THE LINE IS PLOTTED
OPTION = 00: NEITHER ONE IS PLOTTED;

BETA: <REAL ARITHMETIC EXPRESSION>, CONFIDENCE LEVEL OF THE CONFIDENCE BAND. BETA SHOULD EQUAL 0.9, 0.95 OR 0.99;

SORTED: <BOOLEAN EXPRESSION>, INDICATING WHETHER THE SAMPLE IS SORTED (IN A NONDECREASING OR NONINCREASING ORDER). IN CASE OF A SORTED SAMPLE SORTED SHOULD BE "TRUE", OTHERWISE SORTED SHOULD BE "FALSE";

PAR: <REAL ARITHMETIC EXPRESSION>, CONTAINING INFORMATION ABOUT THE THEORETICAL DISTRIBUTION FUNCTION IN CASE TYPE EQUALS WEIBULL2 OR WEIBULL3;

IDENT: <STRING>, IDENTIFYING TEXT TO APPEAR BELOW THE PLOT. THE MAXIMUM NUMBER OF CHARACTERS ALLOWED DEPENDS ON THE SIZE OF THE PLOT, AS INDICATED BY SIZE. IF W AND H DENOTE THE WIDTH AND HEIGHT OF THE PLOT (IN MM), RESPECTIVELY, THEN IT IS 0.5 * W IF H <= 210 AND 0.5 * W * 210 / H OTHERWISE.

TABLE 1 - INTEGER REPRESENTATION OF SOME SYMBOLS (FOR THE COMPLETE TABLE SEE [2], CALCOMP, P.8)

2	Δ
3	+
4	X
11	*

DATA AND RESULTS:

THE EMPIRICAL DISTRIBUTION FUNCTION OF THE SAMPLE V[LV],..., V[UV] IS COMPUTED AND PLOTTED IN A PROBABILITY PLOT, THE TYPE OF WHICH IS INDICATED BY THE PARAMETER TYPE. A DETAILED DESCRIPTION OF THE METHOD AND PERFORMANCE OF PLOTDIST FOR THE VARIOUS TYPES CAN BE FOUND IN [1]. IT SUFFICES TO GIVE A DESCRIPTION OF THE TYPES EXP1, EXP2, WEIBULL2 AND WEIBULL3, SINCE ALL OTHER VALUES OF TYPE REFER TO DISTRIBUTIONS OF THE GIVEN TYPE WITH THE USUAL PARAMETERS.

EXP1 AND EXP2:

THE PLOTS OF TYPE EXP1 AND EXP2 ARE BOTH EXPONENTIAL PROBABILITY PLOTS WHICH DIFFER ONLY IN THE WAY THE STRAIGHT LINE (IF REQUESTED) IS COMPUTED. IN A PLOT OF TYPE EXP1 THE THRESHOLD PARAMETER IS ASSUMED ZERO AND THE LINE RUNS THROUGH THE ORIGIN, WHILE FOR PLOTS OF TYPE EXP2 THE THRESHOLD PARAMETER IS ESTIMATED.

WEIBULL2 AND WEIBULL3:

THE CUMULATIVE DISTRIBUTION FUNCTION OF THE WEIBULL DISTRIBUTION IS F(X) = 1 - EXP(-((X - LOC)/SCALE) ** C).

WEIBULL2:

WHEN LOC IS KNOWN TO BE EQUAL TO 0 PAR SHOULD HAVE THE VALUE 0. THEN THE EMPIRICAL DISTRIBUTION FUNCTION OF THE SAMPLE -LN(V[LV]),..., -LN(V[UV]) WILL BE PLOTTED IN A GUMBEL PROBABILITY PLOT.

WHEN THE VALUE OF C IS KNOWN PAR SHOULD BE EQUAL TO C (AND HENCE PAR > 0). THEN THE EMPIRICAL DISTRIBUTION FUNCTION OF V[LV],..., V[UV] WILL BE PLOTTED IN A WEIBULL (FIXED C) PROBABILITY PLOT.

WEIBULL3:

WHEN THE VALUE OF LOC IS KNOWN PAR SHOULD BE MADE EQUAL TO LOC. THEN THE EMPIRICAL DISTRIBUTION FUNCTION OF -LN(V[LV] - LOC),..., -LN(V[UV] - LOC) WILL BE PLOTTED IN A GUMBEL PROBABILITY PLOT.

WHEN PAR EQUALS 0, ALL THREE PARAMETERS ARE ASSUMED UNKNOWN. IN THIS CASE THE LOCATION PARAMETER LOC IS ESTIMATED AND THEN THE PROCEDURE IS THE SAME AS ABOVE.

ERRORMESSAGES WILL BE WRITTEN ON FILE OUTPUT VIA CHANNEL 61.

PROCEDURES USED:

AXIS	(CALCOMP)
NUMBER	(CALCOMP)
PLOT	(CALCOMP)
PLOTS	(CALCOMP)
SCALE	(CALCOMP)
SYMBOL	(CALCOMP)

```
VECINDQSORT         (STATAL3 11021)
LOGGAMMA            (STATAL3 40400)
PHI                 (STATAL3 41500)
PHIINV              (STATAL3 41501)
ZEROIN              (NUMAL3  34150)
```

LANGUAGE: ALGOL 60

REFERENCES:

[1] ES, A.J. VAN & PUTTEN, C. VAN
PROBABILITY PLOTS
STATAL REPORT 3
MATHEMATICAL CENTRE, AMSTERDAM, 1983

[2] GRAPHICS
SARA PUBLICATIE 11
STICHTING ACADEMISCH REKENCENTRUM AMSTERDAM, 1980

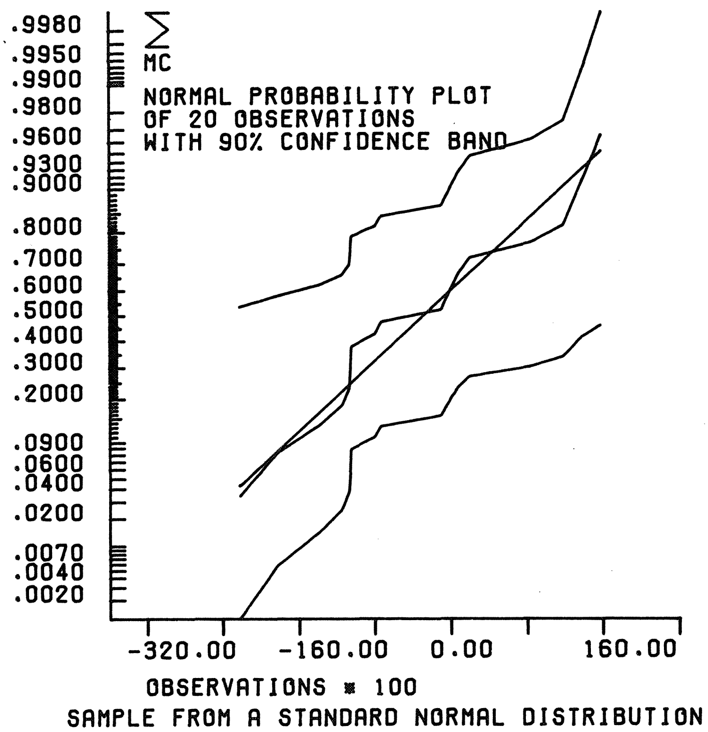
EXAMPLE OF USE:

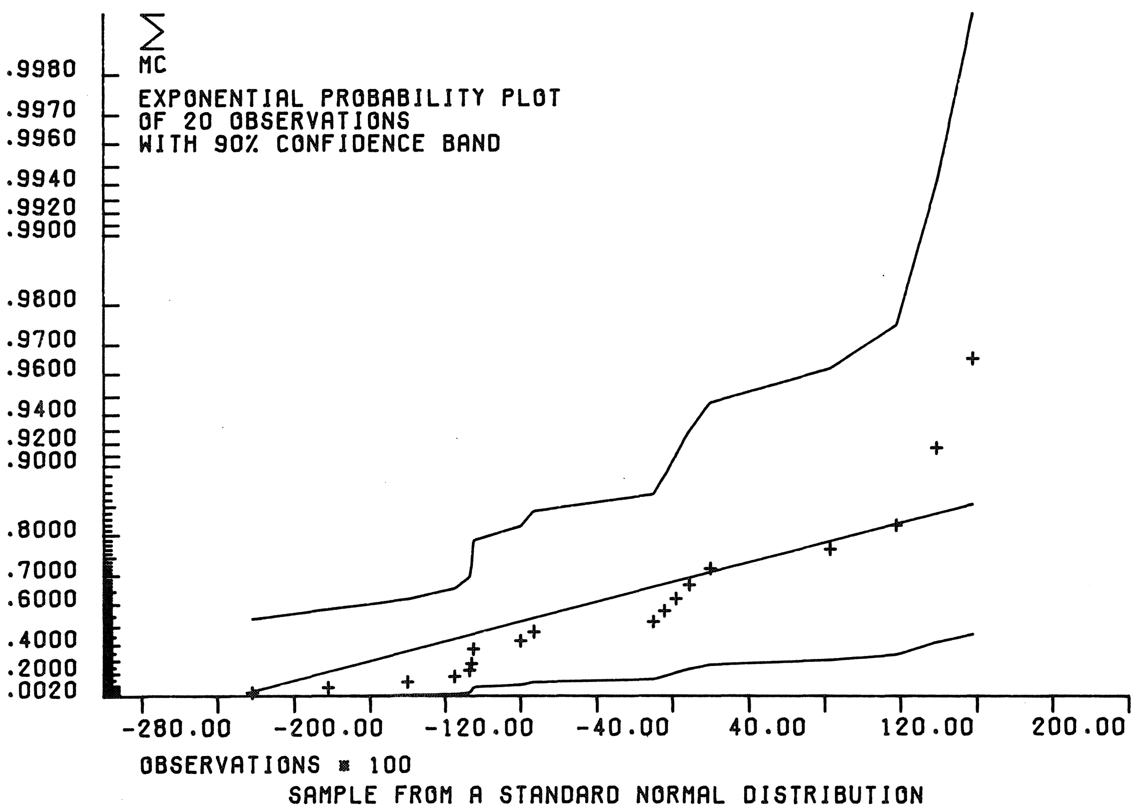
```
PROGRAM:
"BEGIN"
    "ARRAY" V[1:20];
    "PROCEDURE" PLOTDIST(GRFILE,V,LV,UV,TYPE,LB,UB,MODE,PART,
                         SIZE,OPTION,BETA,SORTED,PAR,IDENT);
    "VALUE" V,LV,UV,LB,UB,MODE,PART,SIZE,OPTION,BETA,SORTED,
            PAR;
    "INTEGER" LV,UV,MODE,PART,SIZE,OPTION;
    "REAL" LB,UB,BETA,PAR;
    "BOOLEAN" SORTED;
    "ARRAY" V; "STRING" GRFILE, TYPE, IDENT;
    "CODE" 47005;

    INARRAY(60,V);
    PLOTDIST("("")", V, 1, 20, "("NORMAL")", 0, 1,-1, 0,
             100100, 11, 0.90 , "FALSE", 0,
             "("SAMPLE FROM A STANDARD NORMAL DISTRIBUTION")");
    PLOTDIST("("")", V, 1, 20,  "("EXP2")",  0, 1, 3, 0,
             160110, 11, 0.90 , "FALSE", 0,
             "("SAMPLE FROM A STANDARD NORMAL DISTRIBUTION")")
"END"

INPUT:
-0.80   1.58   0.02   0.83
-1.05   0.20  -1.07   0.09
 1.39   1.18  -0.73  -0.04
-0.10  -1.40  -2.22  -1.05
```

RESULTING PLOTS, OBTAINED BY SENDING THE GRAPHFILE GRFILE TO A PLOTTER (SEE [2]) :





5. INTERPRETING PROBABILITY PLOTS AND EXAMPLES

5.1. Guidelines for interpreting probability plots

This section contains a brief outline of information that may be obtained by looking at probability plots. The material presented here is covered in more detail by BICKEL & DOKSUM [3], BURY [4] and HEMELRIJK & KRIENS [7] (in Dutch).

To fix ideas, assume that a probability plot of some type, A say, is given, including an estimated straight line (see chapter 3) and a confidence band with coefficient $\beta = 1-\alpha$ (see section (2.3)). The straight line does not pretend to be a very good estimate of the underlying distribution function, it is merely for reference. The width of the confidence band at some argument x, i.e. the distance between upper and lower bound of the band at x, may provide a relative weight of the departure of the empirical distribution function from the straight line at x. By this we mean that a departure is more serious as the width is smaller.

First we consider the case where the observations have been transformed by a linear transformation only.

a. When there are but minor deviations from a straight line, one has to be aware that this does not prove that the observations come from a type A distribution; some other distribution might fit equally well.
b. When there are major deviations, one might want to reject the hypothesis of a type A distribution, if no straight line fits into the confidence band. This is a conservative procedure (BURY [4], p.383, see also section (2.3)), by which we mean that the size of this test is smaller than $1-\beta$. Note that the confidence band covers the graph of the underlying, unknown distribution function with probability β. As an alternative procedure one could use a general goodness of fit test, like the Kolmogorov test (cf. BICKEL & DOKSUM [3], section (9.6)), or a test designed for a particular family of distributions, such as the Shapiro-Wilk test for normal distributions.

The nature of the deviations from a straight line may provide additional information.

1. When the right part of the empirical distribution is concave, this

indicates that the right tail of the underlying distribution might be heavier than the one corresponding to a straight line. In the same way convexity indicates a relatively lighter right tail.

2. Similar statements about the left tail can be obtained by replacing "concave" by "convex" and "convexity" by "concavity" in 1.
3. When the graph of the empirical distribution function looks like one of those in Figure 6, one might suspect a mixture of type A distributions. To be more precise, one might suspect the observations to be a non-decreasing, piecewise linear transformation of type A distributed underlying data. Note that, generally, mixtures with distribution functions of the type pF + qG, where F and G are distribution functions and p and q are weights, do not yield linear probability plots.

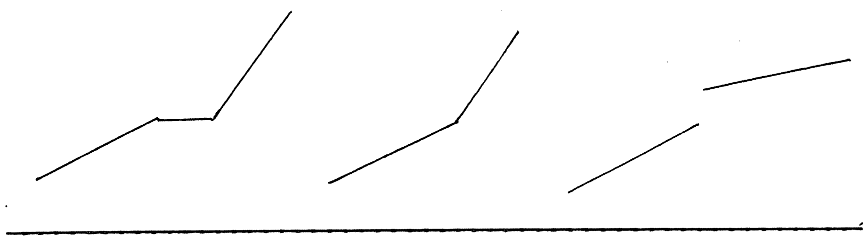


Figure 6.

Secondly, we consider the presence of a location parameter in case of a logarithmic transformation of the observations. To be more explicit, the case is considered where the empirical distribution function based on $-\ln(x-\mathrm{loc})$, instead of observations x, is plotted. Here loc is an estimate of the location parameter. Convexity or concavity of the graph in this situation may also be due to a bad value of loc. If one is convinced that the transformed observations are a sample from a type A distribution, the convexity of the graph may indicate that the location estimate should be increased. Of course, concavity indicates that the estimate should be decreased. As an illustration of this see Figures 62, 63, 64 and 65, which present three Weibull plots of the same data with different values of loc.

REMARK. When plotting heavy tailed distributions or plotting on heavy tailed probability paper, we strongly recommend to make enlargements (too). We suggest the following values of the parameters LB and UB.

TYPE	EXP	CAUCHY	WEIBULL
LB	0	.1	0
UB	.9	.9	.9

For Weibull plots with very heavy or light tails, sometimes it is favourable to delete the plotting of a confidence band and to increase the vertical size of the plot.

5.2. Examples of probability plots

This section contains numerous examples of probability plots. The first subsection shows the effect of the layout parameters of PLOTDIST (see chapter 4) on the actual plot. Changes of probability plots due to variation of the sample size, are presented in the second subsection. The third subsection consists of three parts, the first one containing plots of samples on the correct type of probability paper, the second one showing what samples from different distribution plotted on uniform, normal and exponential probability paper look like, the third one presenting the converse, i.e. uniformly, normally and exponentially distributed samples plotted on different types of probability paper.

All samples used here have been generated by means of the pseudo-random number generator ASELECT from the library STATAL [13] (cf. VAN ES & VAN PUTTEN [6]) in addition to transformation procedures as given in VAN PUTTEN & VAN DER TWEEL [11]. All procedures used are implemented in STATAL. They will be described in the final version of [13].

Relevant information about the values of the parameters is displayed at the top and the bottom of the plots. To interpret the parameter PAR correctly in Weibull2 and Weibull3 plots, the reader is referred to chapter 4.

Since the program had to be adapted to a new ALGOL compiler the plots made by the current version of PLOTDIST are slightly different from the plots presented in this section. The main difference is in the horizontal axis which now gives equidistant numbers.

5.2.1. Layout parameters of PLOTDIST

This subsection contains nine plots of one sample from a standard normal distribution. They illustrate how the layout of the plot can be controlled when using the STATAL procedure PLOTDIST. Relevant information about the parameters of this procedure is given as a bottom line.

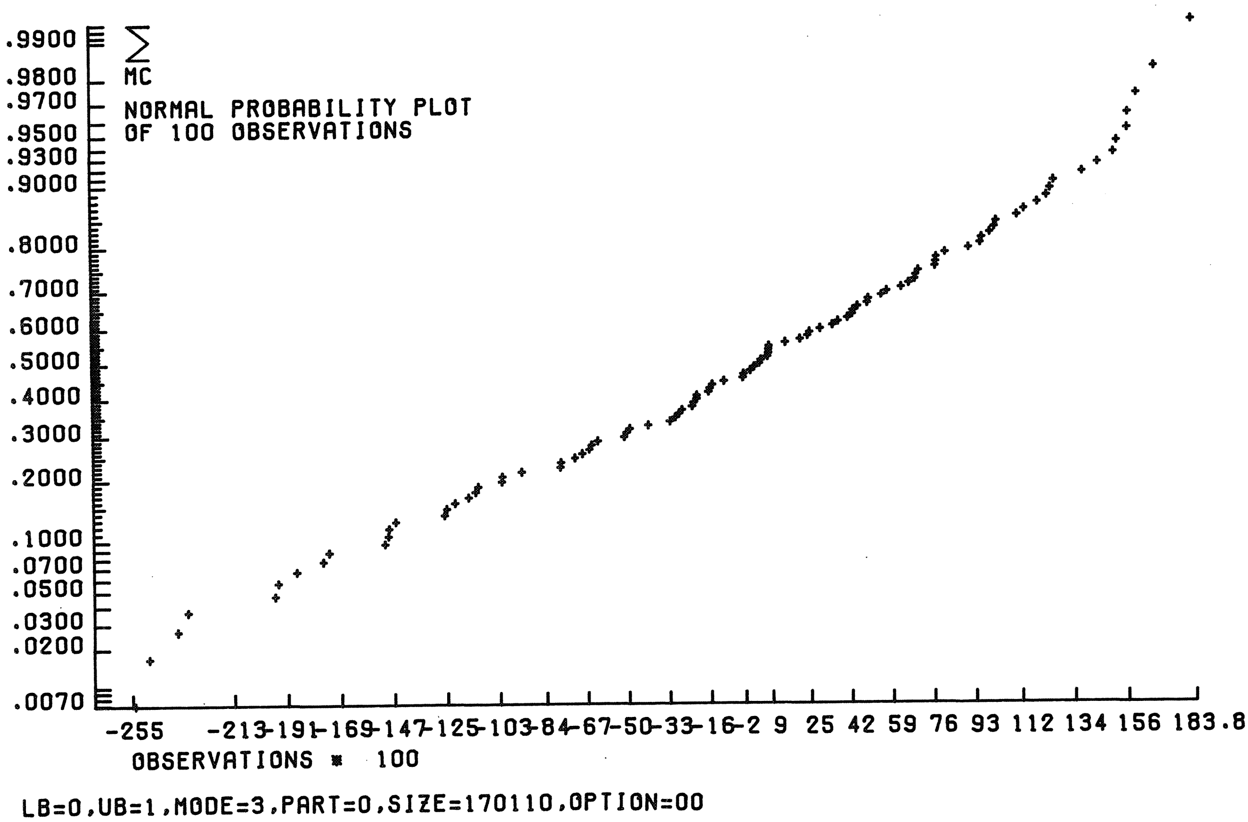
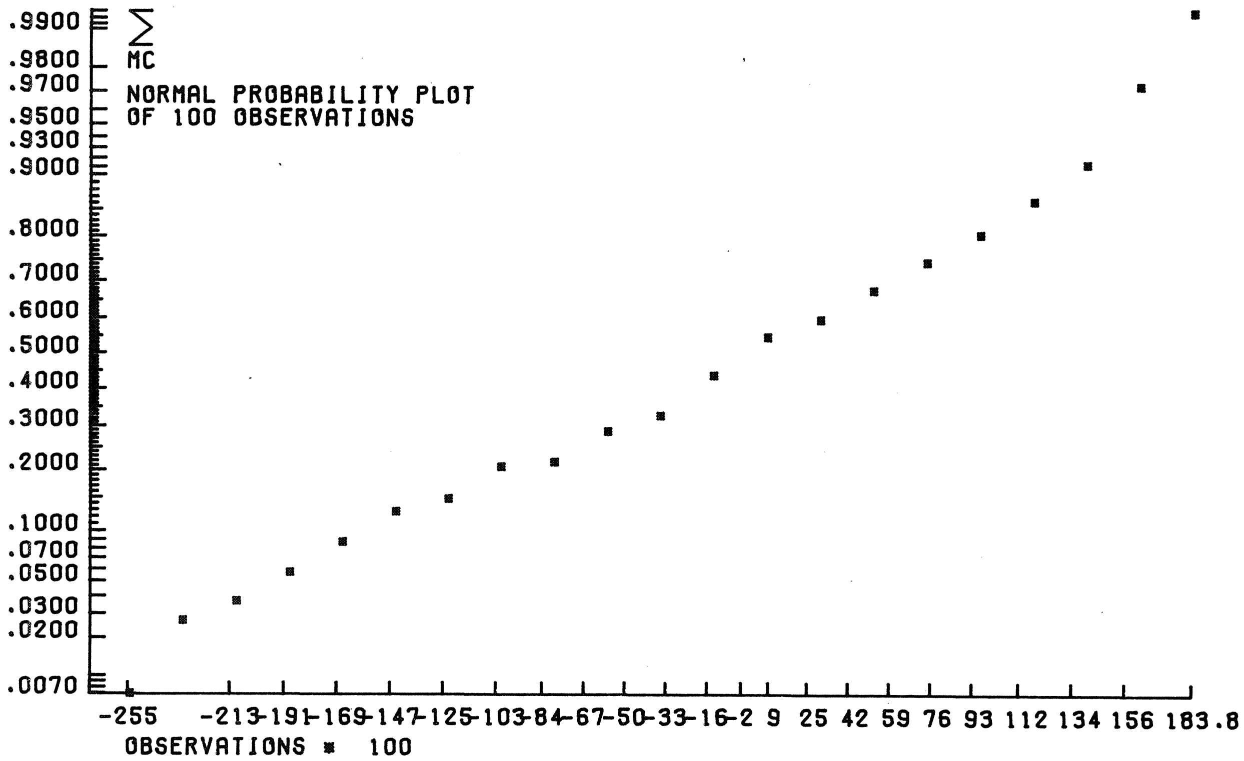
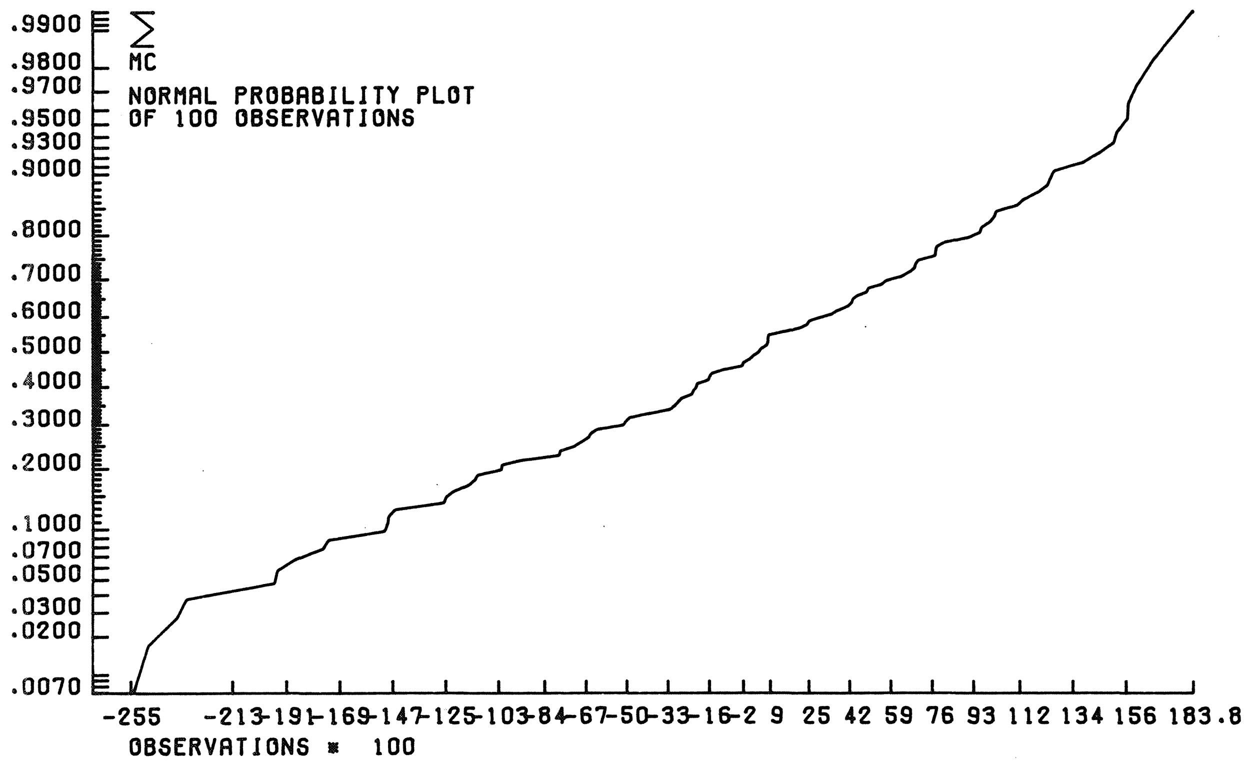


Figure 7.



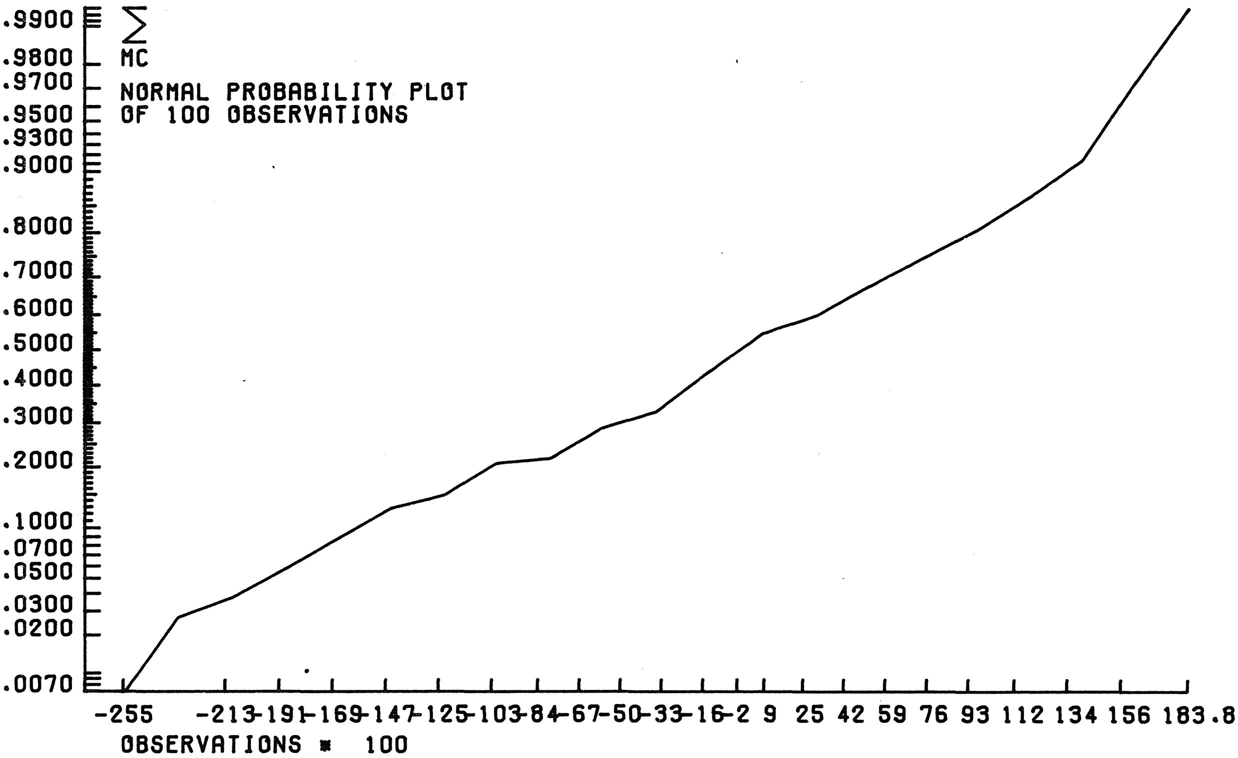
LB=0,UB=1,MODE=11,PART=20,SIZE=170110,OPTION=00

Figure 8.



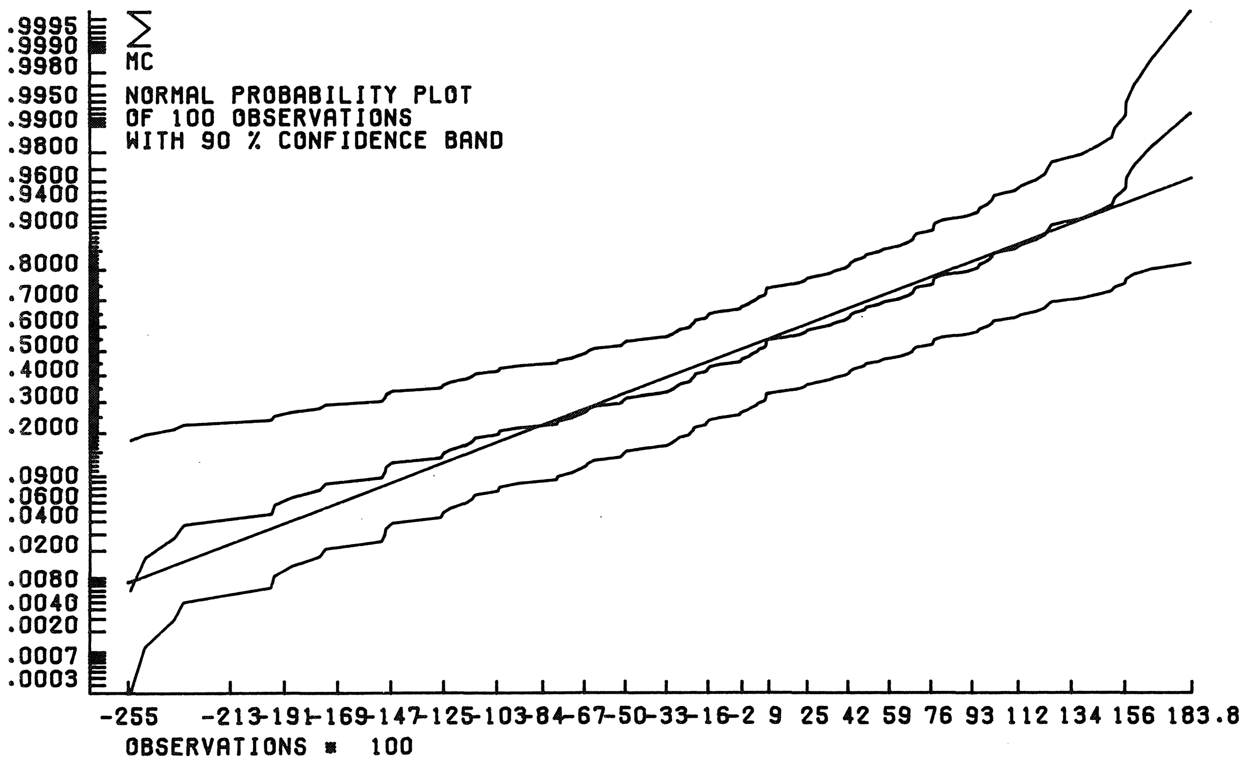
LB=0,UB=1,MODE=-1,PART=0,SIZE=170110,OPTION=00

Figure 9.



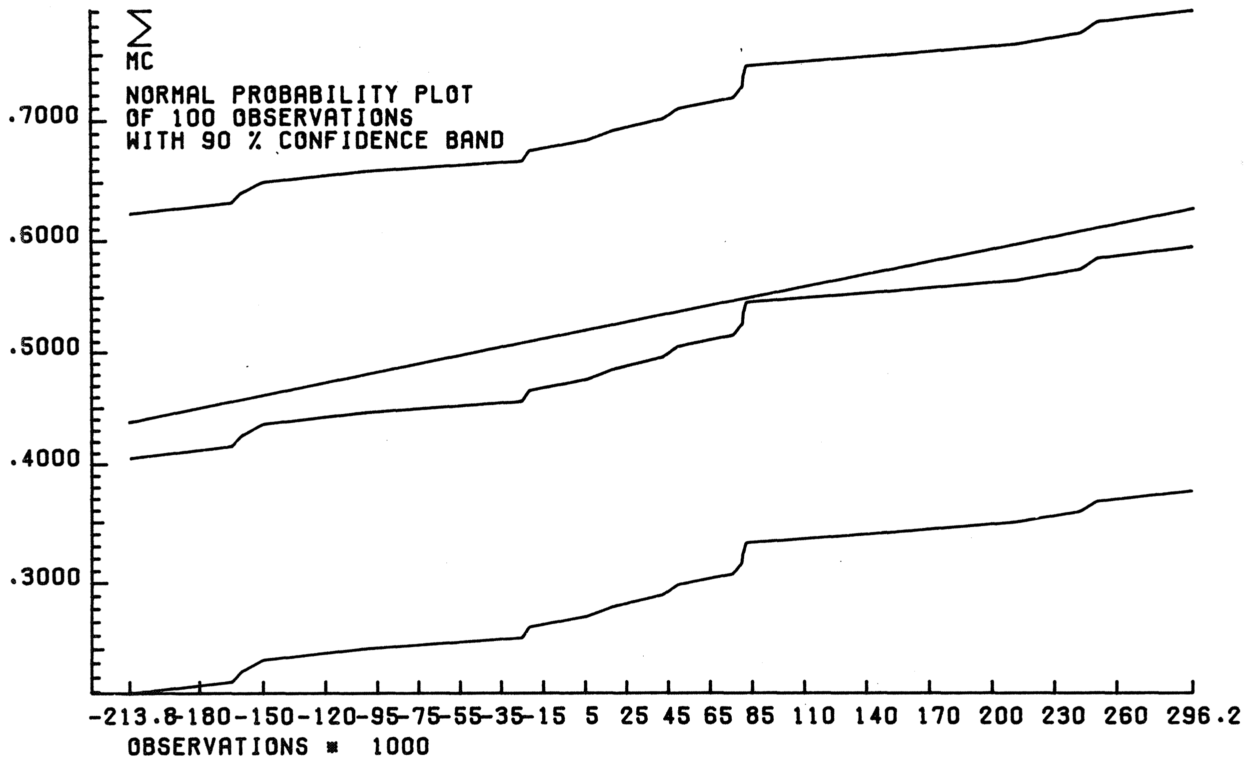
LB=0,UB=1,MODE=-1,PART=20,SIZE=170110,OPTION=00

Figure 10.



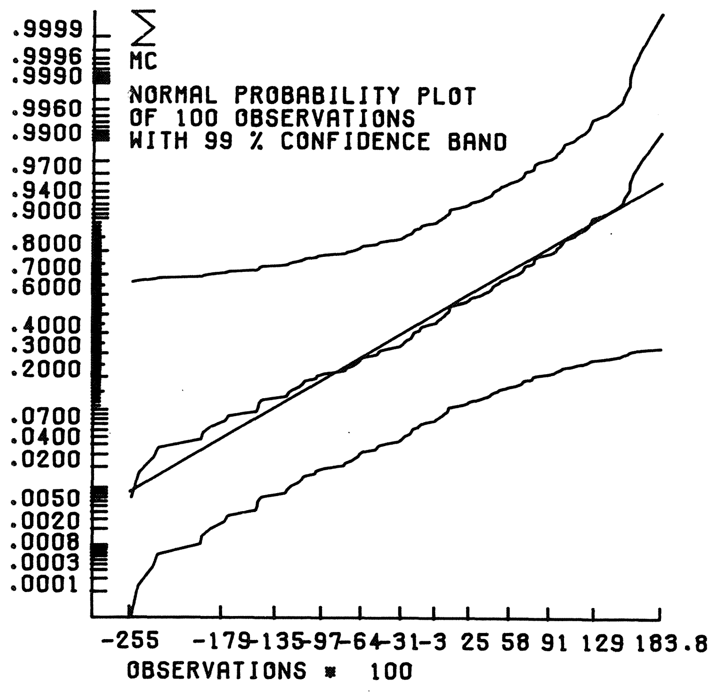
LB=0,UB=1,MODE=-1,PART=0,SIZE=170110,OPTION=11

Figure 11.



LB=0.4,UB=0.6,MODE=-1,PART=0,SIZE=170110,OPTION=11

Figure 12.



SIZE=100100,MODE=-1,PART=0,OPTION=11

Figure 13.

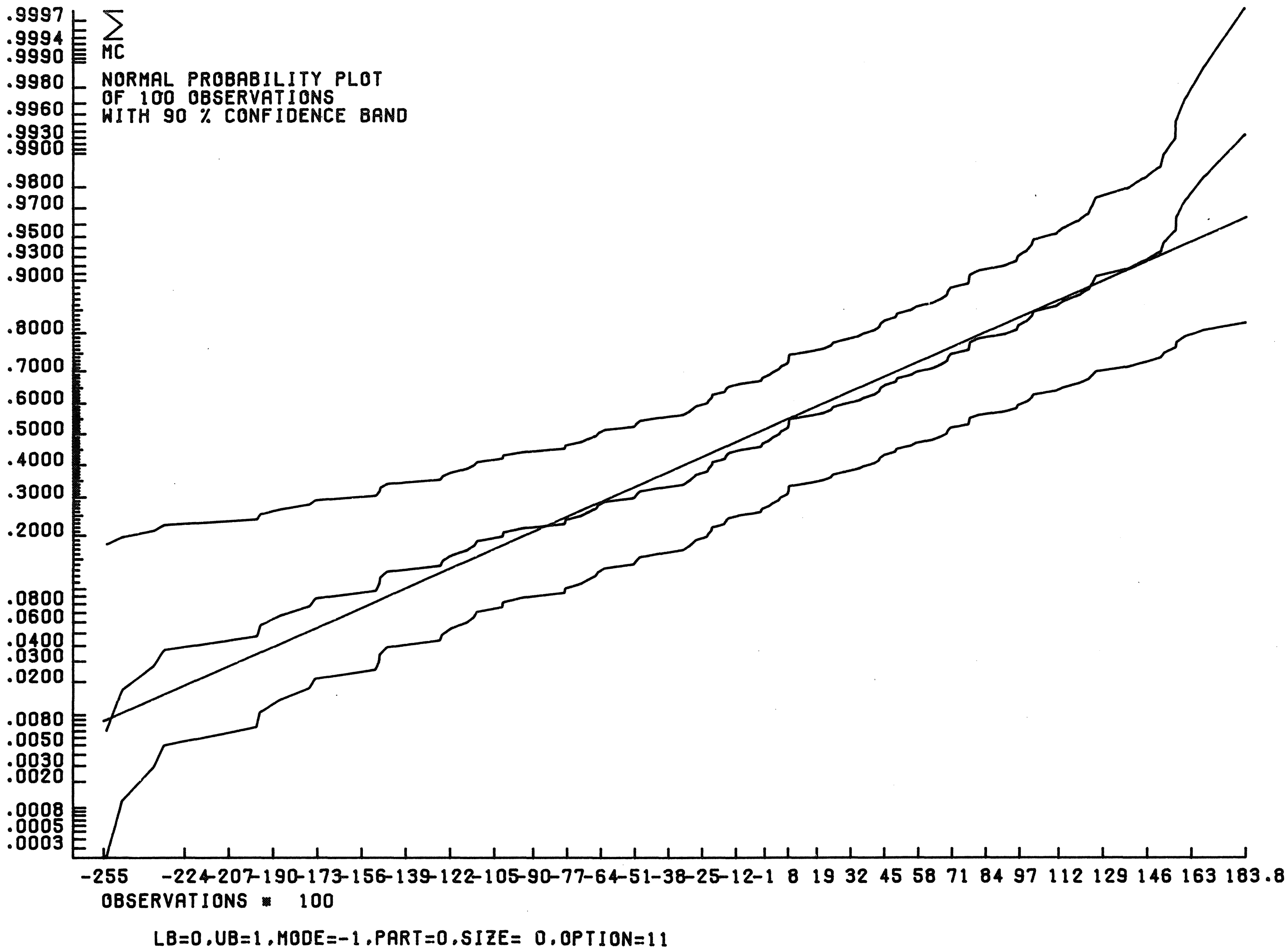

.9997
.9994
.9990
.9980
.9960
.9930
.9900
.9800
.9700
.9500
.9300
.9000
.8000
.7000
.6000
.5000
.4000
.3000
.2000
.0800
.0600
.0400
.0300
.0200
.0080
.0050
.0030
.0020
.0008
.0005
.0003
∑
MC
NORMAL PROBABILITY PLOT
OF 100 OBSERVATIONS
WITH 90 % CONFIDENCE BAND
-255 -224-207-190-173-156-139-122-105-90-77-64-51-38-25-12-1 8 19 32 45 58 71 84 97 112 129 146 163 183.8
OBSERVATIONS * 100
LB=0,UB=1,MODE=-1,PART=0,SIZE= 0,OPTION=11

Figure 14.

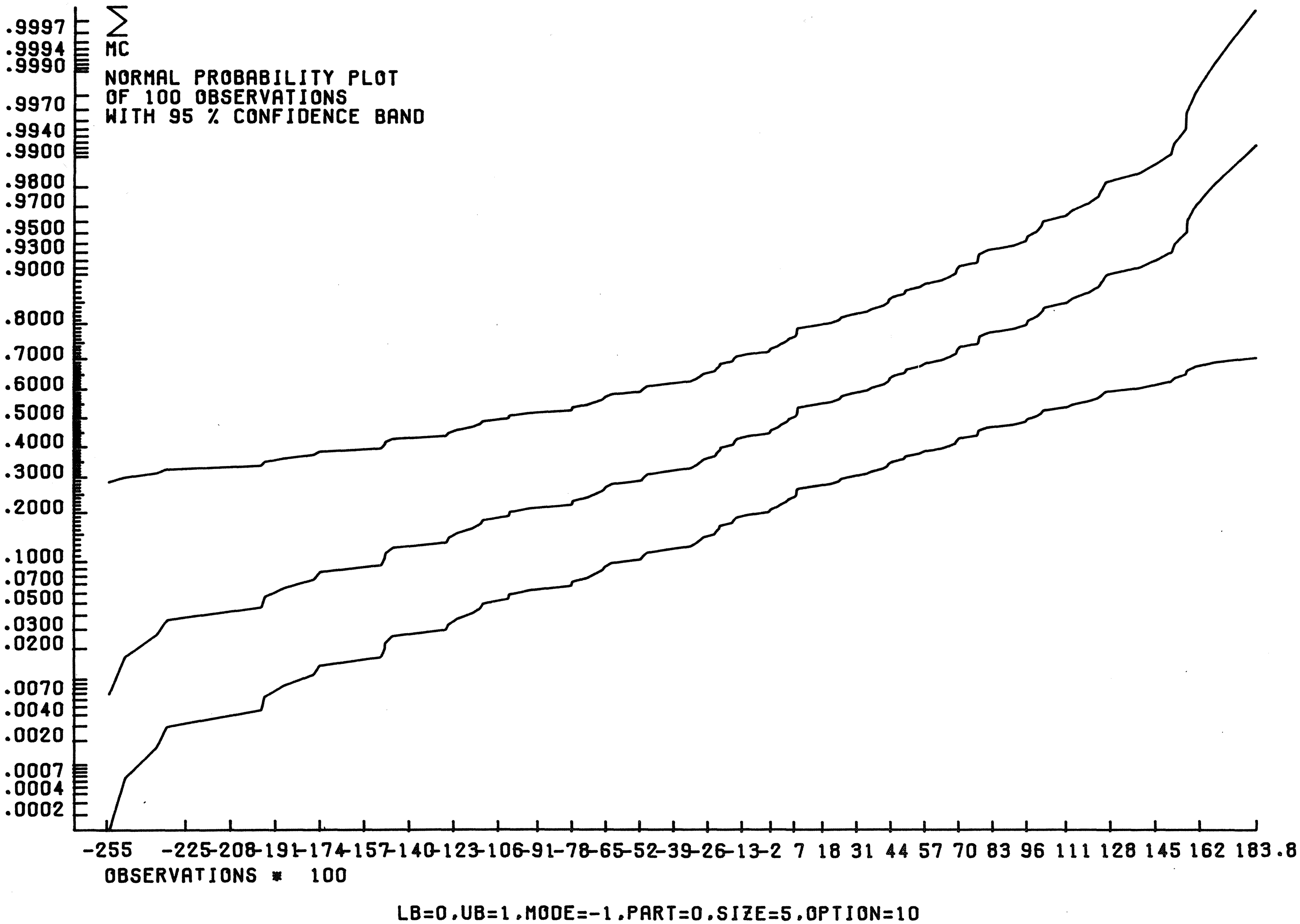

Σ
MC
NORMAL PROBABILITY PLOT
OF 100 OBSERVATIONS
WITH 95 % CONFIDENCE BAND
.9997
.9994
.9990
.9970
.9940
.9900
.9800
.9700
.9500
.9300
.9000
.8000
.7000
.6000
.5000
.4000
.3000
.2000
.1000
.0700
.0500
.0300
.0200
.0070
.0040
.0020
.0007
.0004
.0002
-255 -225-208-191-174-157-140-123-106-91-78-65-52-39-26-13-2 7 18 31 44 57 70 83 96 111 128 145 162 183.8
OBSERVATIONS * 100
LB=0,UB=1,MODE=-1,PART=0,SIZE=5,OPTION=10

Figure 15.

5.2.2. Effect of the sample size.

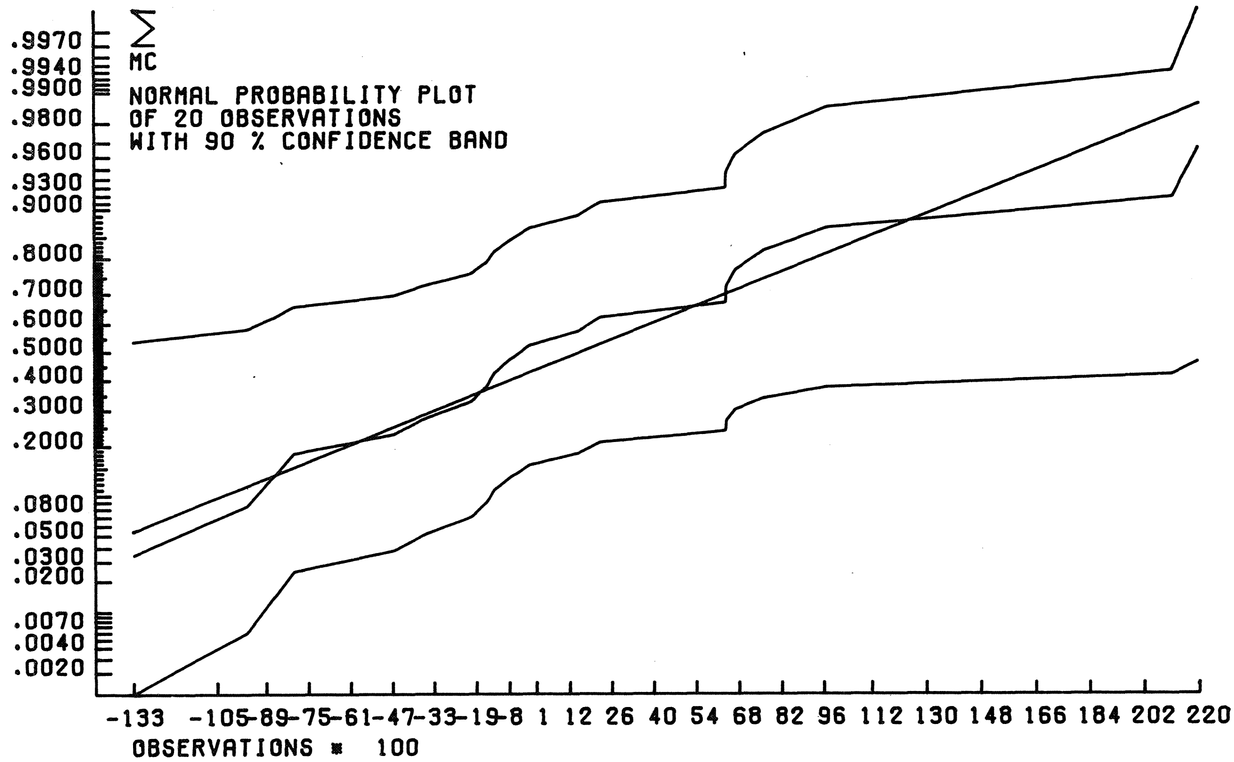


Figure 16. First sample from the standard normal distribution

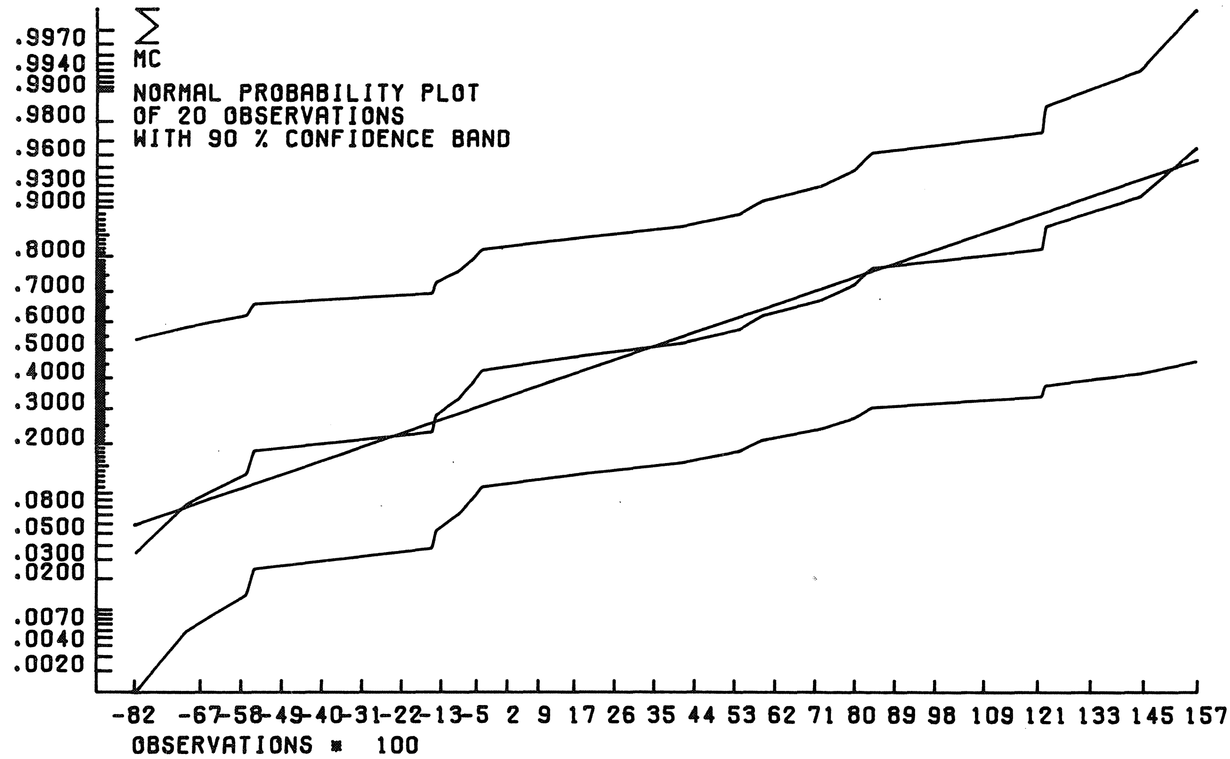


Figure 17. Second sample from the standard normal distribution

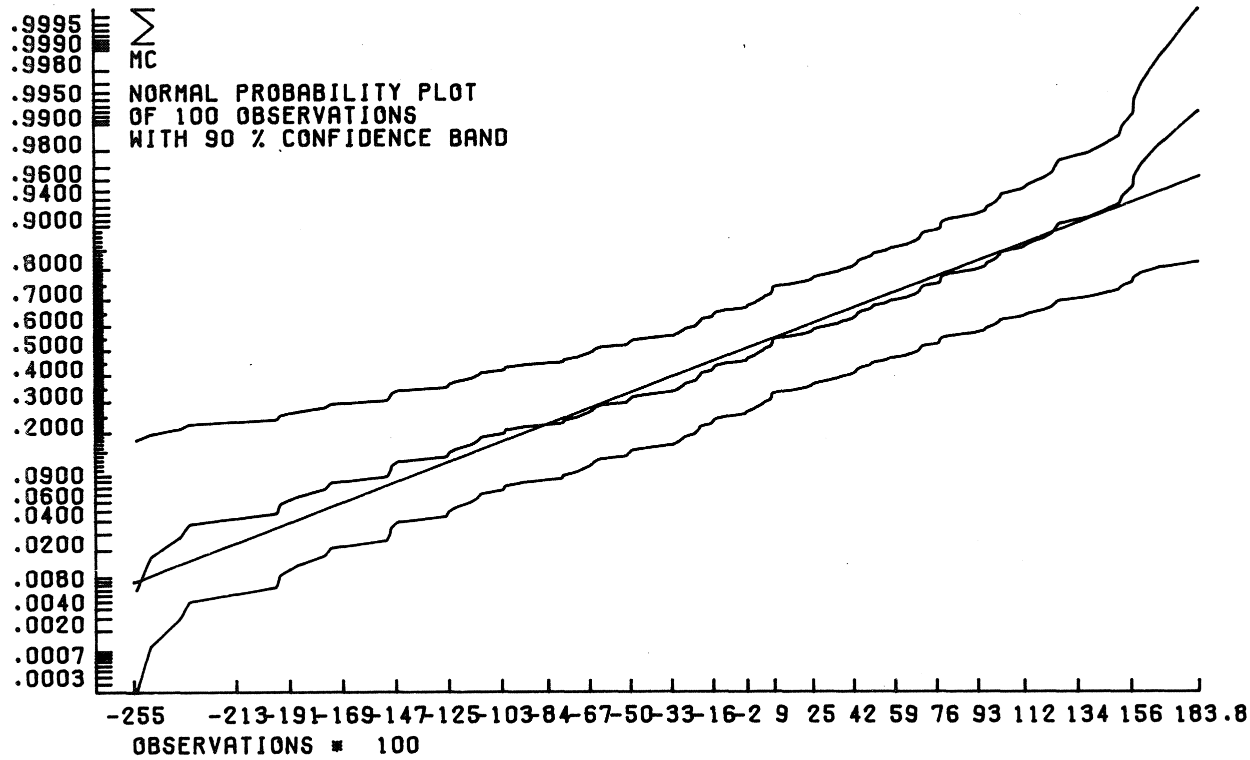


Figure 18. Third sample from the standard normal distribution

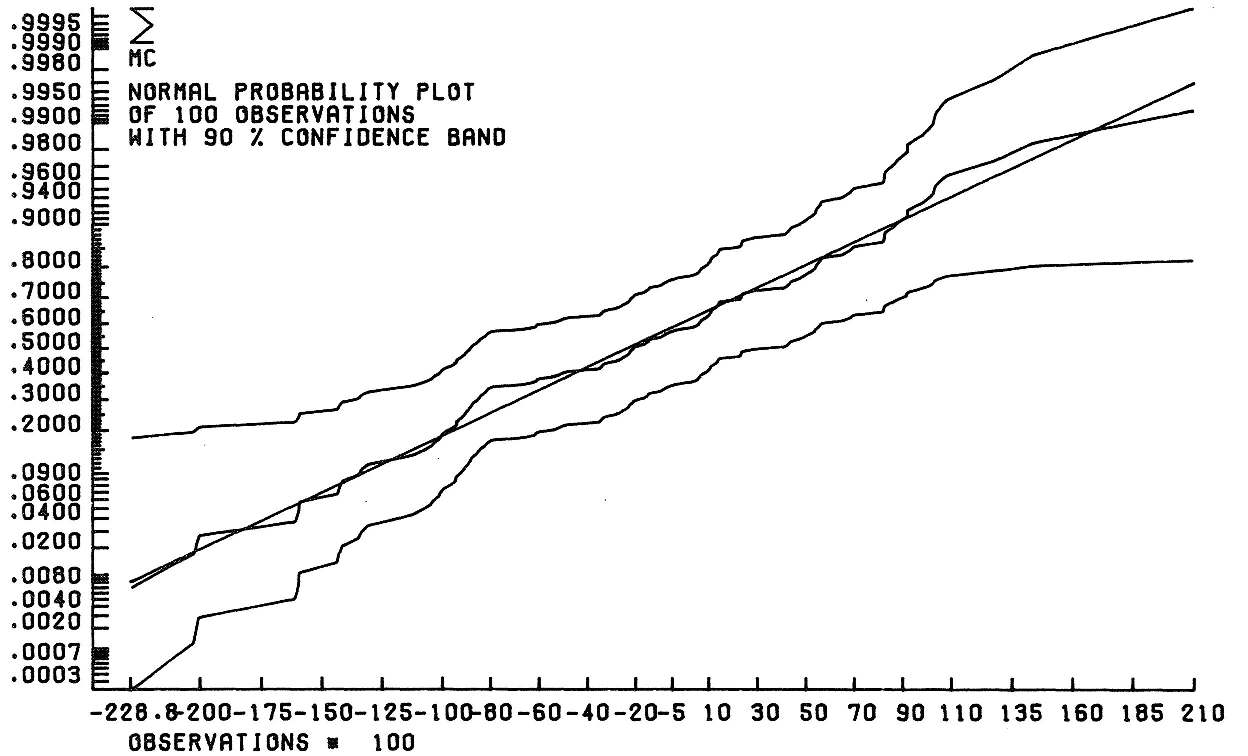


Figure 19. Fourth sample from the standard normal distribution

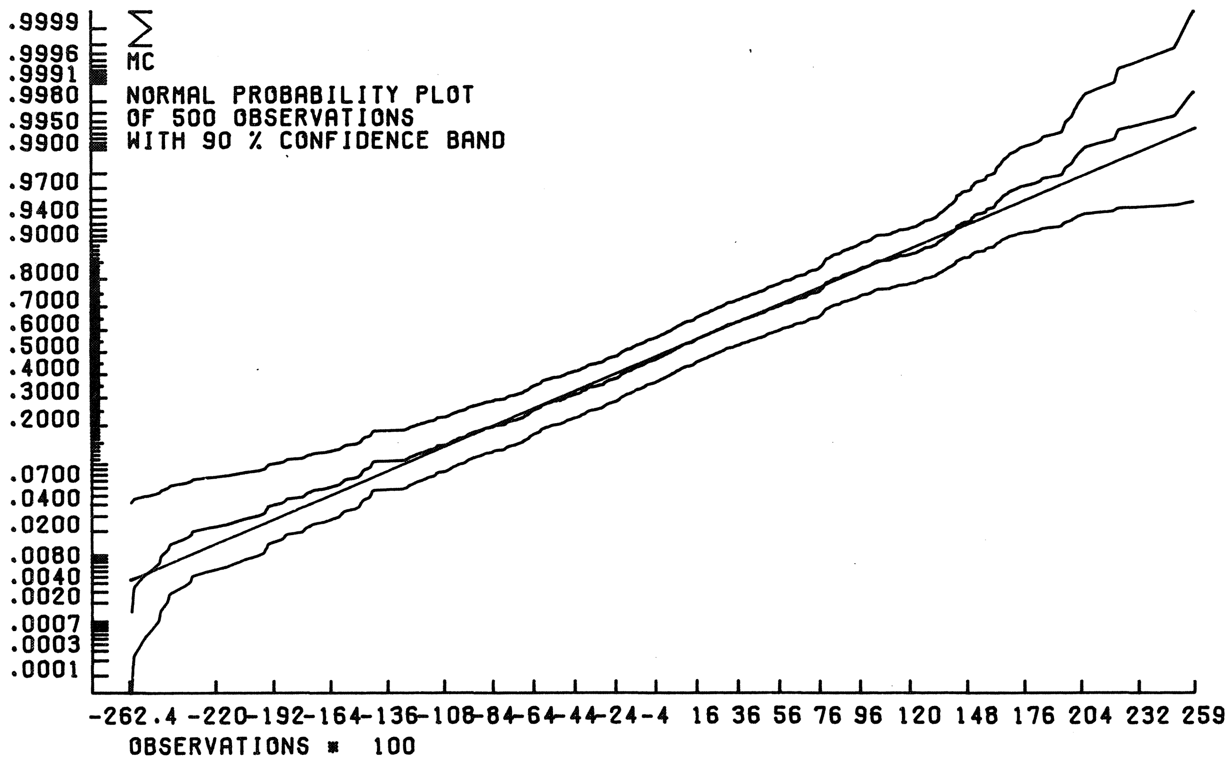


Figure 20. Fifth sample from the standard normal distribution

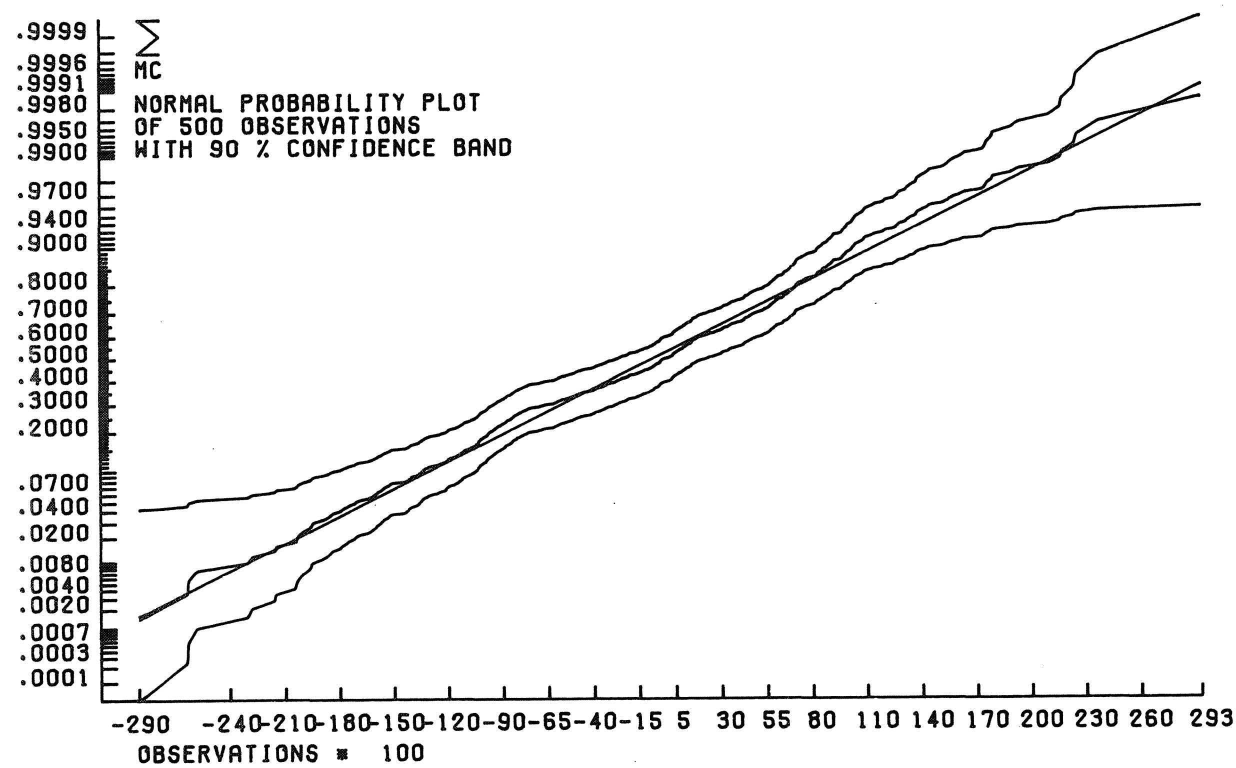


Figure 21. Sixth sample from the standard normal distribution

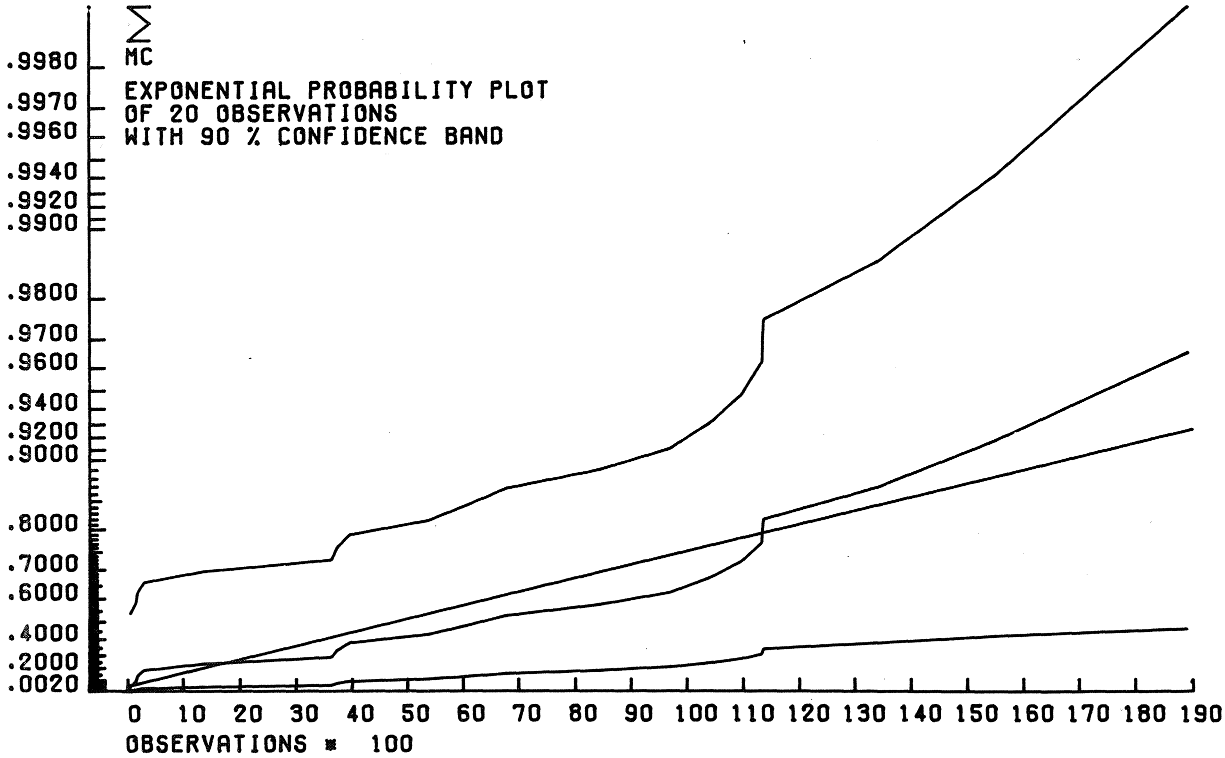


Figure 22. First sample from the standard exponential distribution, i.e. with parameters 0 and 1

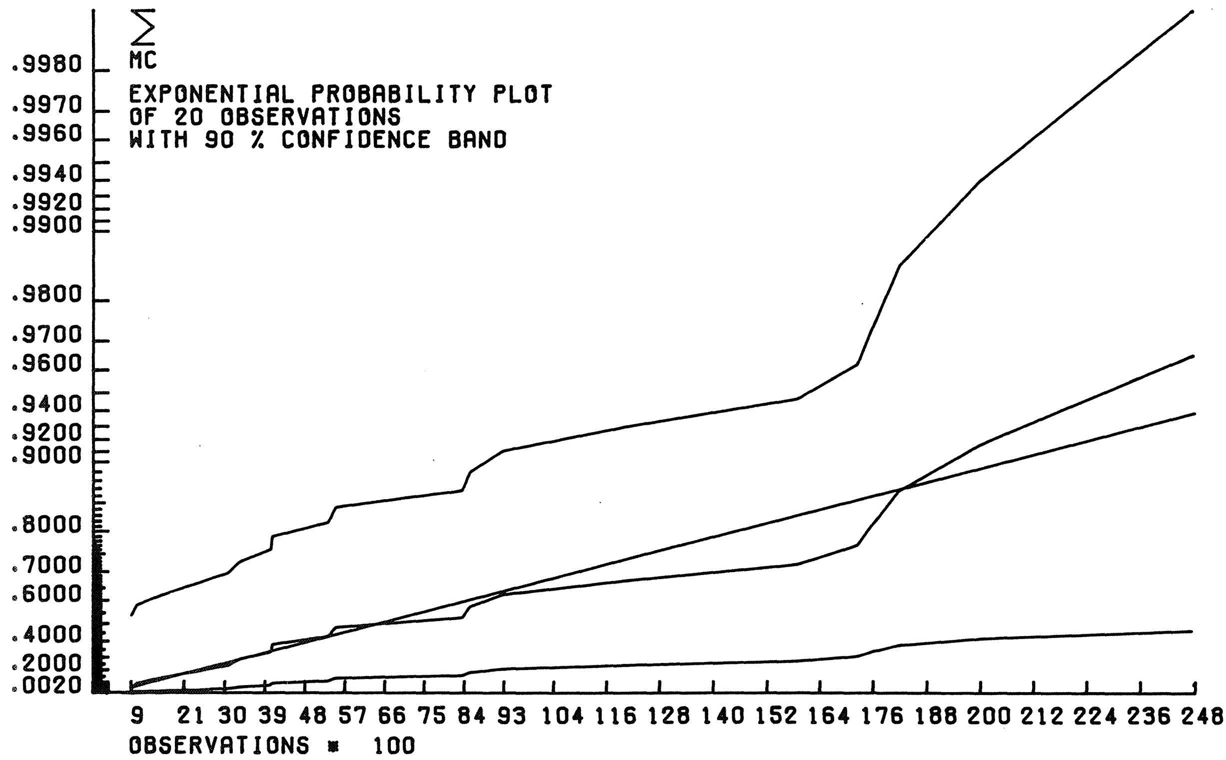


Figure 23. Second sample from the standard exponential distribution

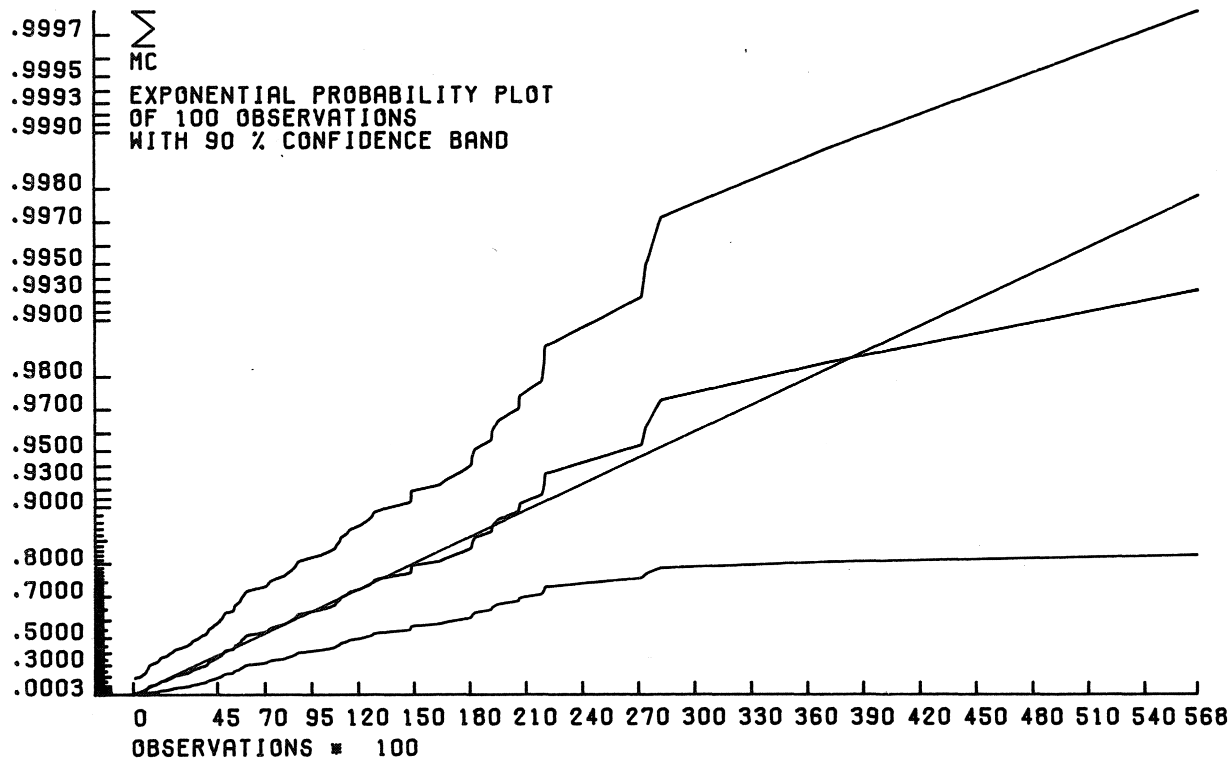


Figure 24. Third sample from the standard exponential distribution

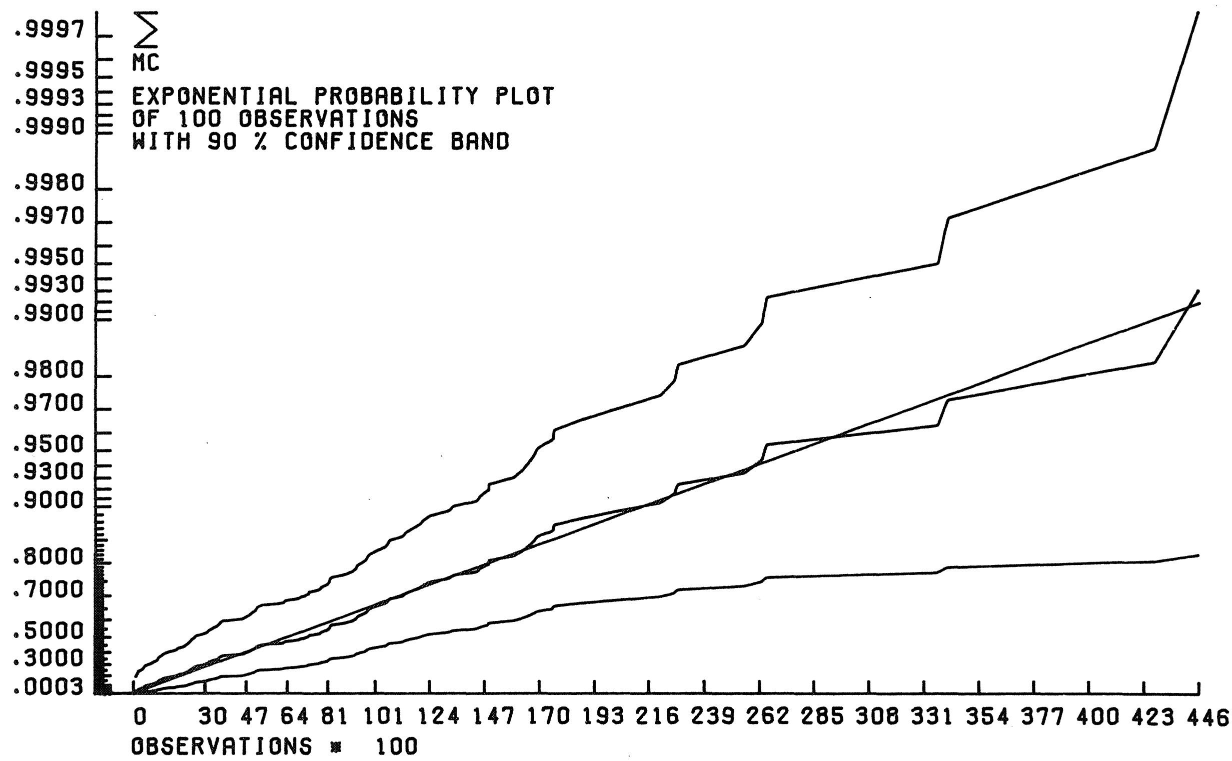


Figure 25. Fourth sample from the standard exponential distribution

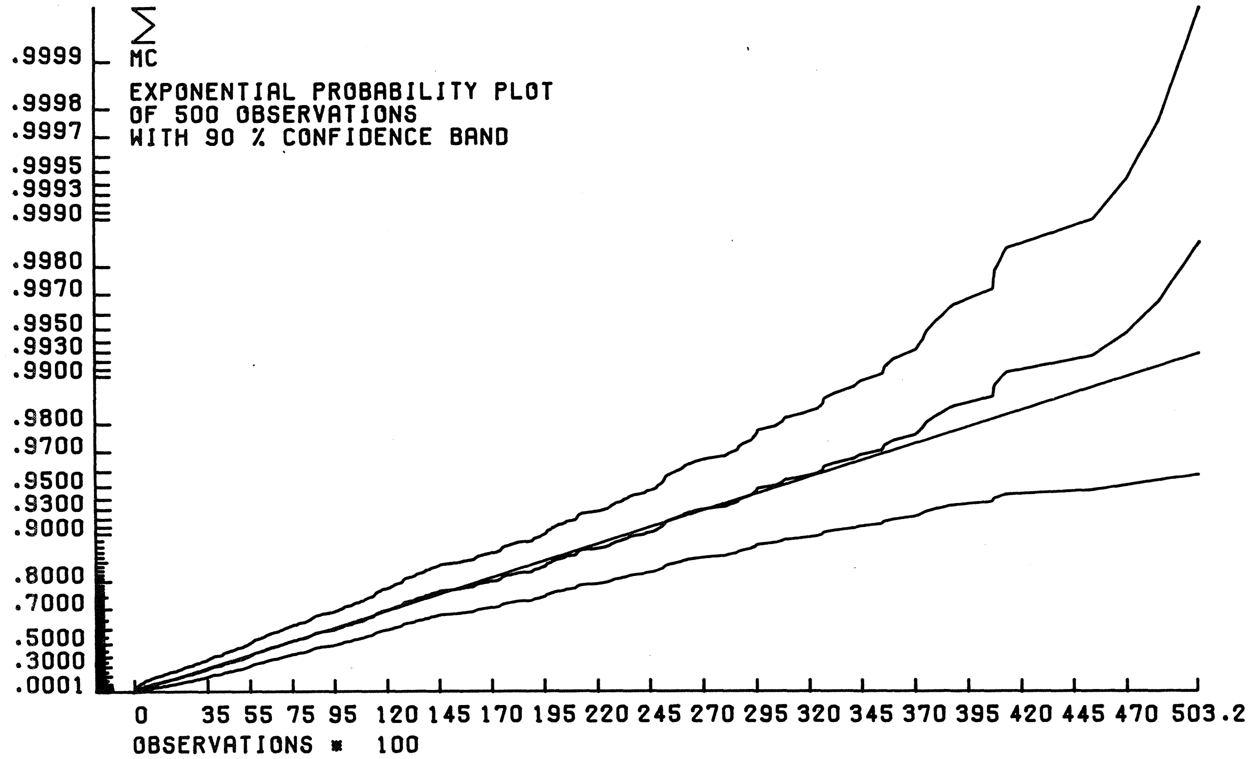


Figure 26. Fifth sample from the standard exponential distribution

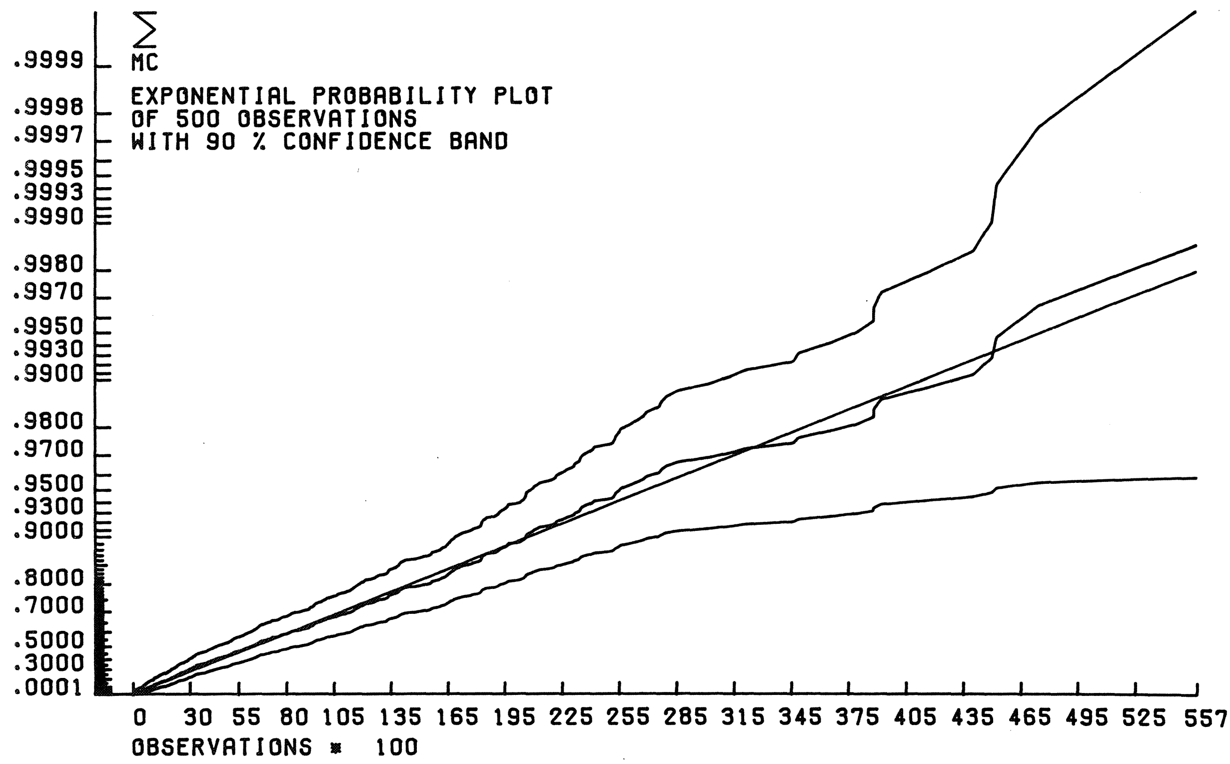


Figure 27. Sixth sample from the standard exponential distribution

5.2.3. Samples from several distributions in several types of plots

5.2.3.1. Samples in plots of corresponding type

In the first part of this subsection uniform, normal, exponential, Laplace, Gumbel and Cauchy probability plots, with sample sizes 20, 50, 100 and 500, of corresponding samples are presented, followed by numerous examples of Weibull probability plots of samples from the Weibull (1,1,2) distribution, which has fairly "normal" tails, and the Weibull (0,1,0.25) distribution, which has heavy tails.

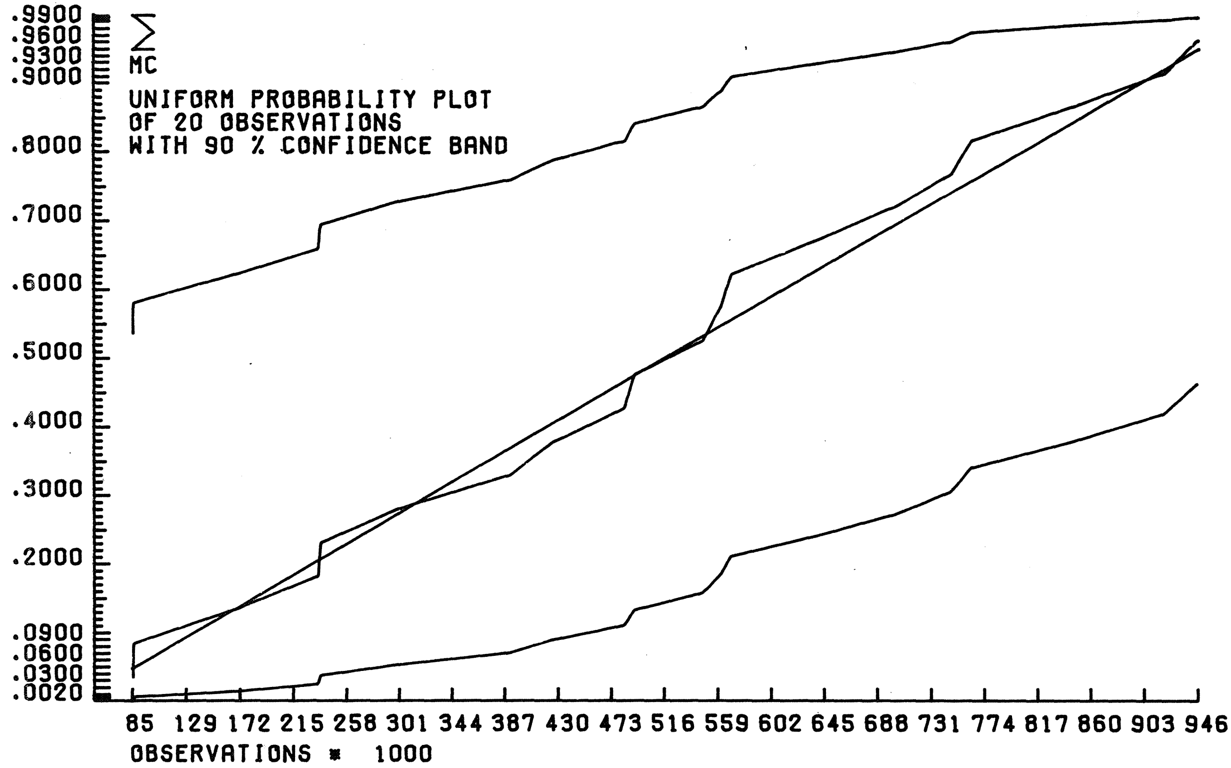


Figure 28. A sample from the uniform (0,1) distribution

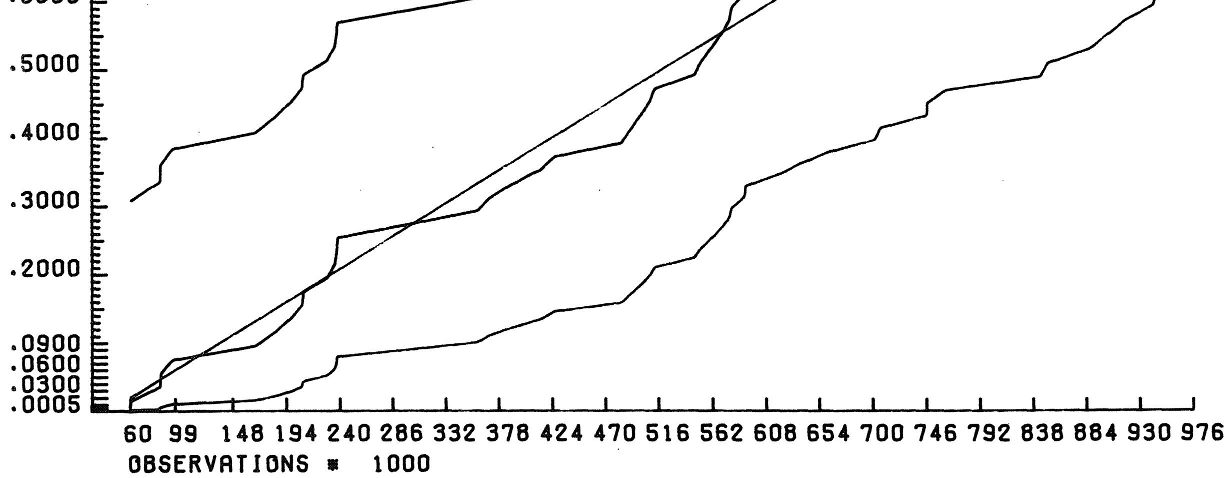


Figure 29. A sample from the uniform (0,1) distribution

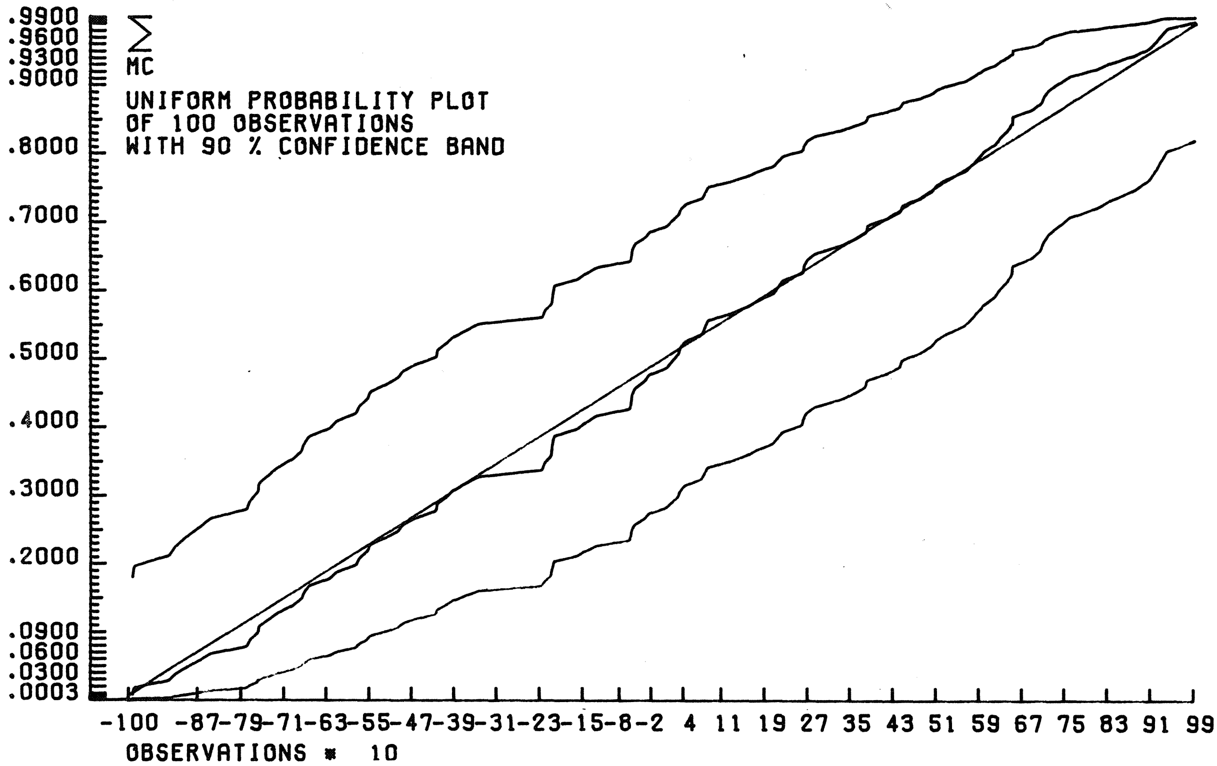


Figure 30. A sample from the uniform (-10,10) distribution

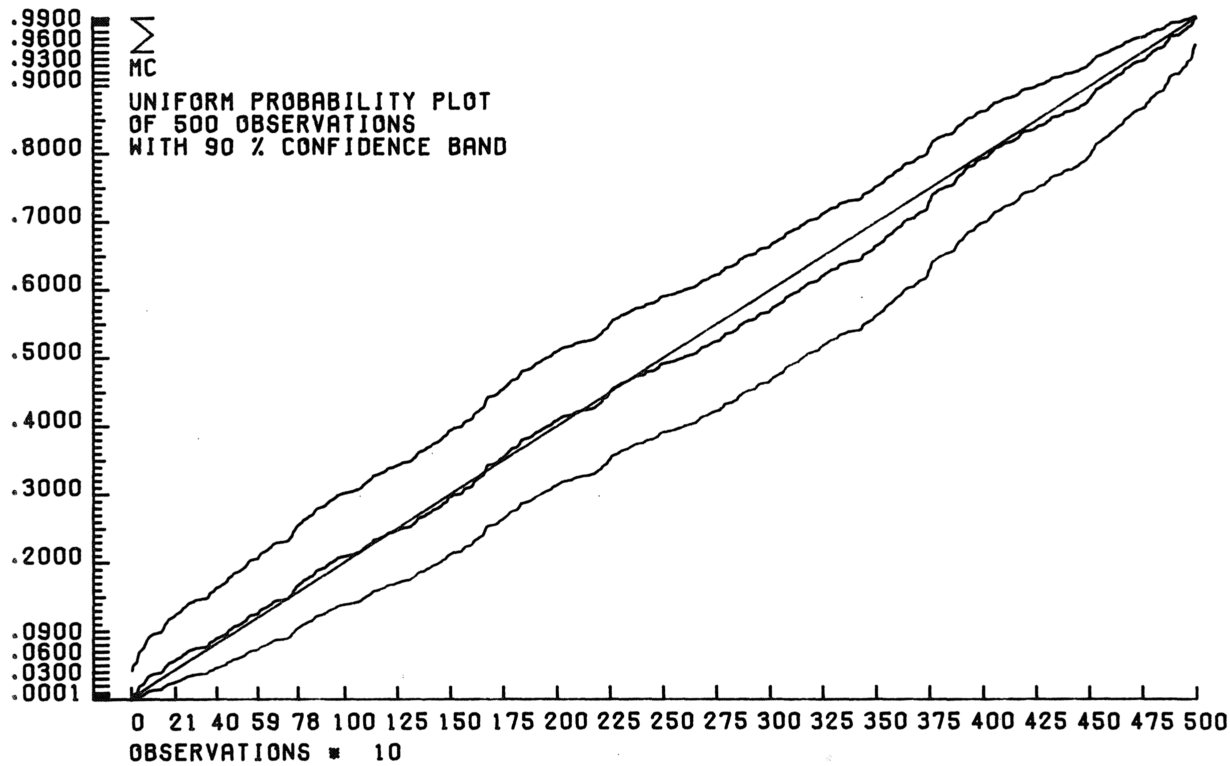


Figure 31. A sample from the uniform (0,50) distribution

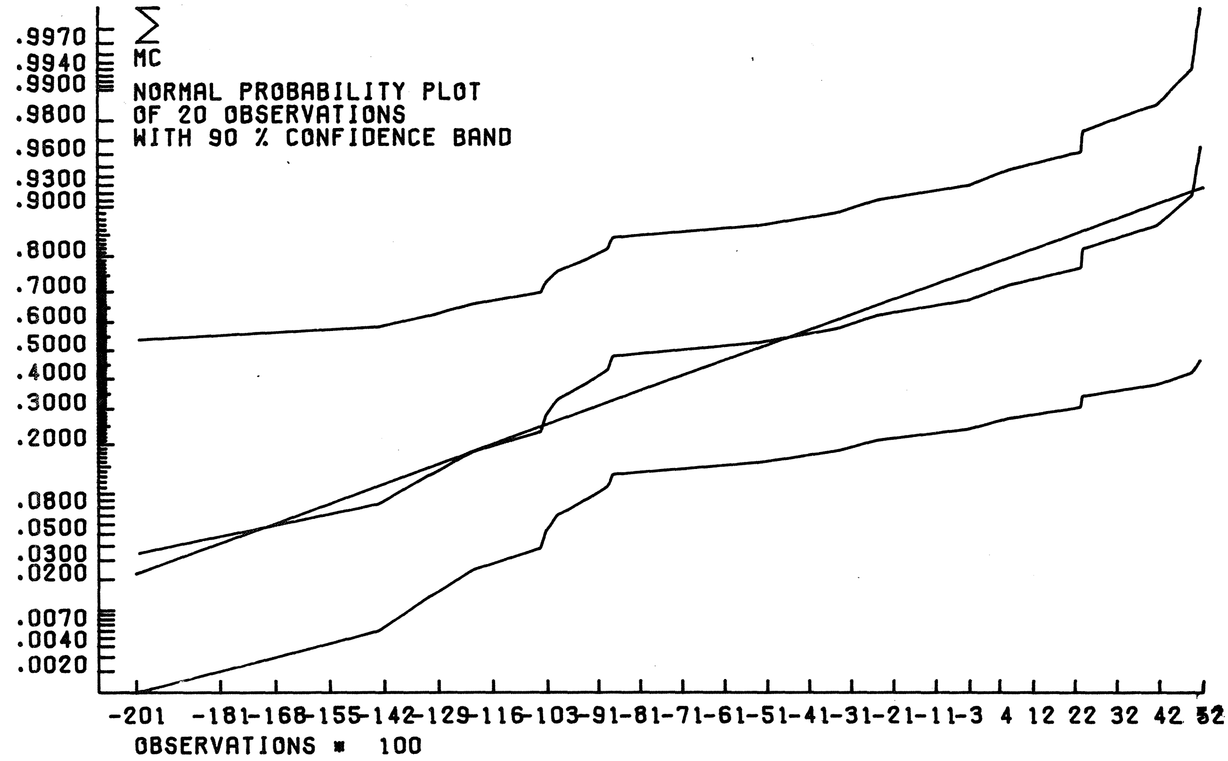


Figure 32. A sample from the normal (0,1) distribution

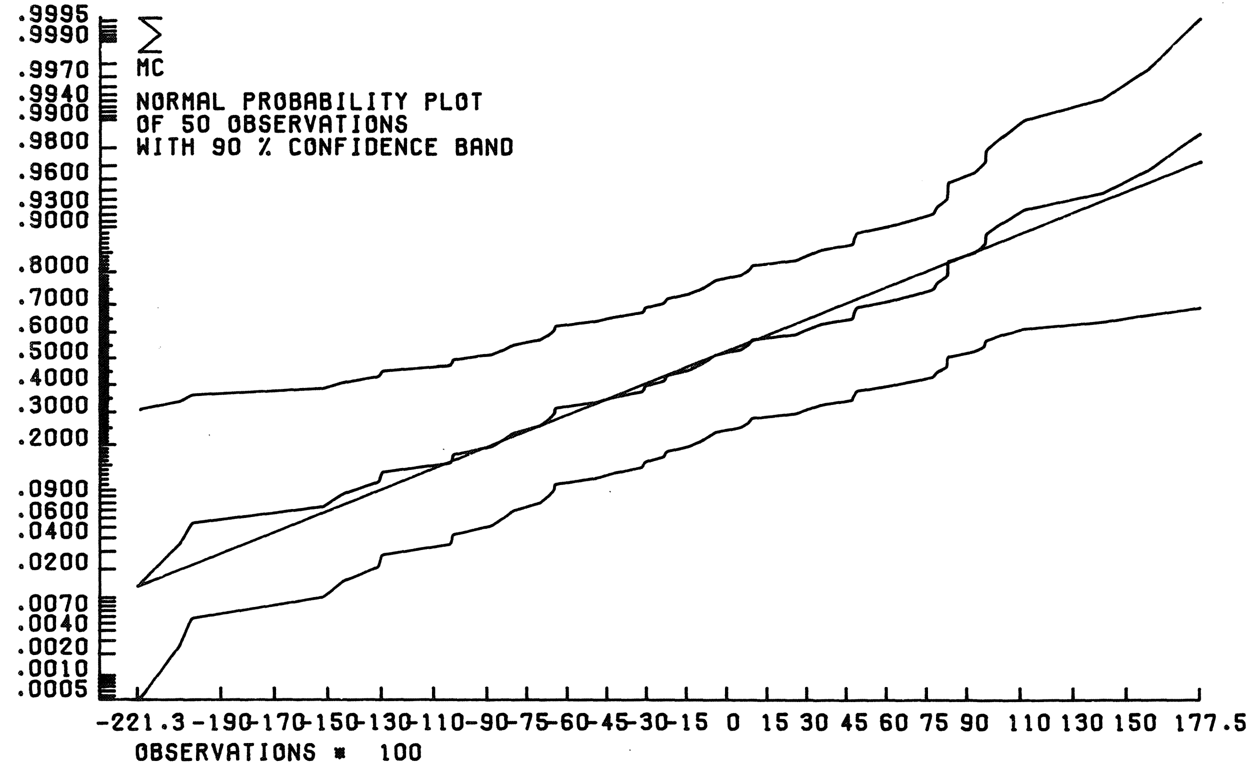


Figure 33. A sample from the normal (0,1) distribution

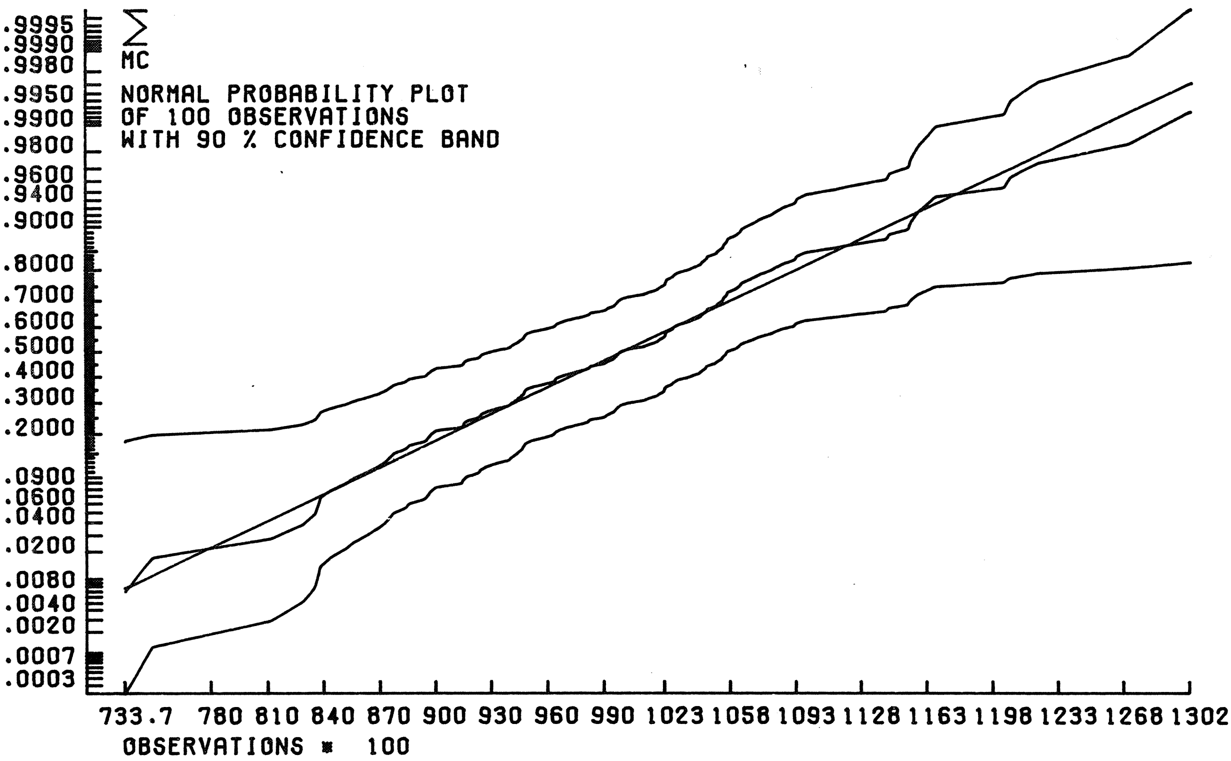


Figure 34. A sample from the normal (10,1) distribution

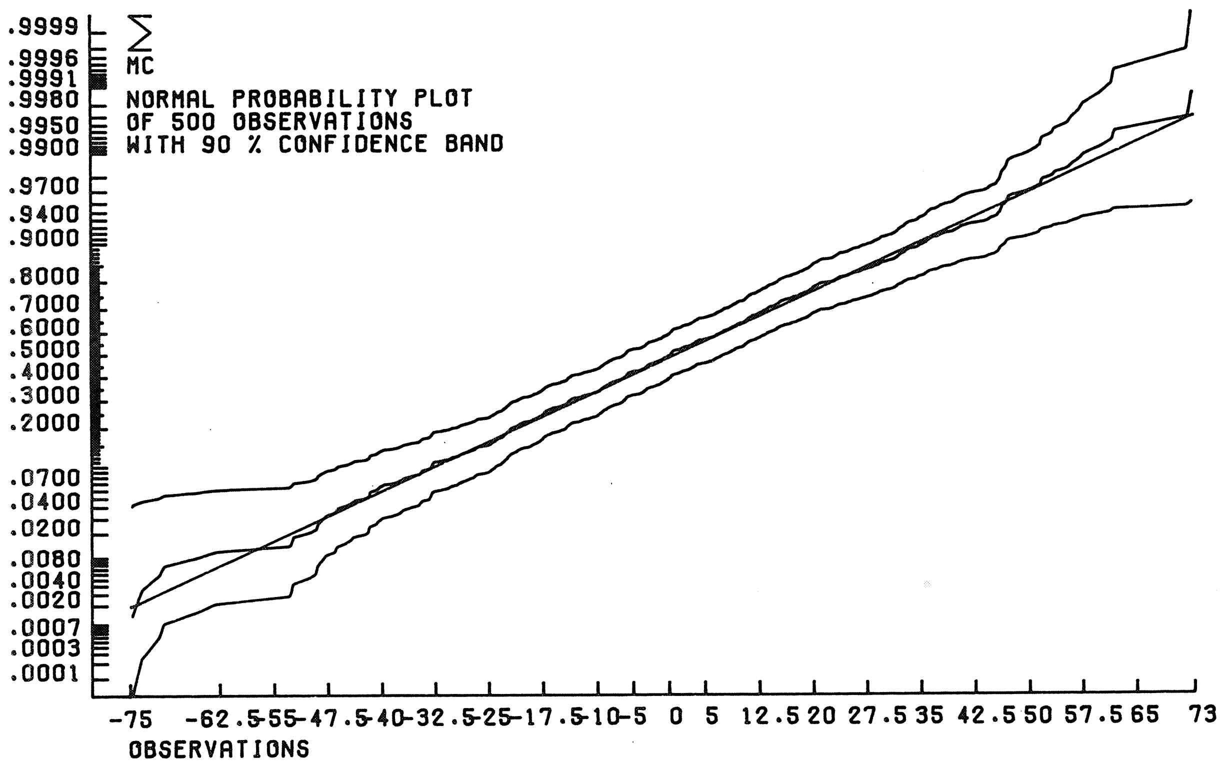


Figure 35. A sample from the normal (0,25) distribution

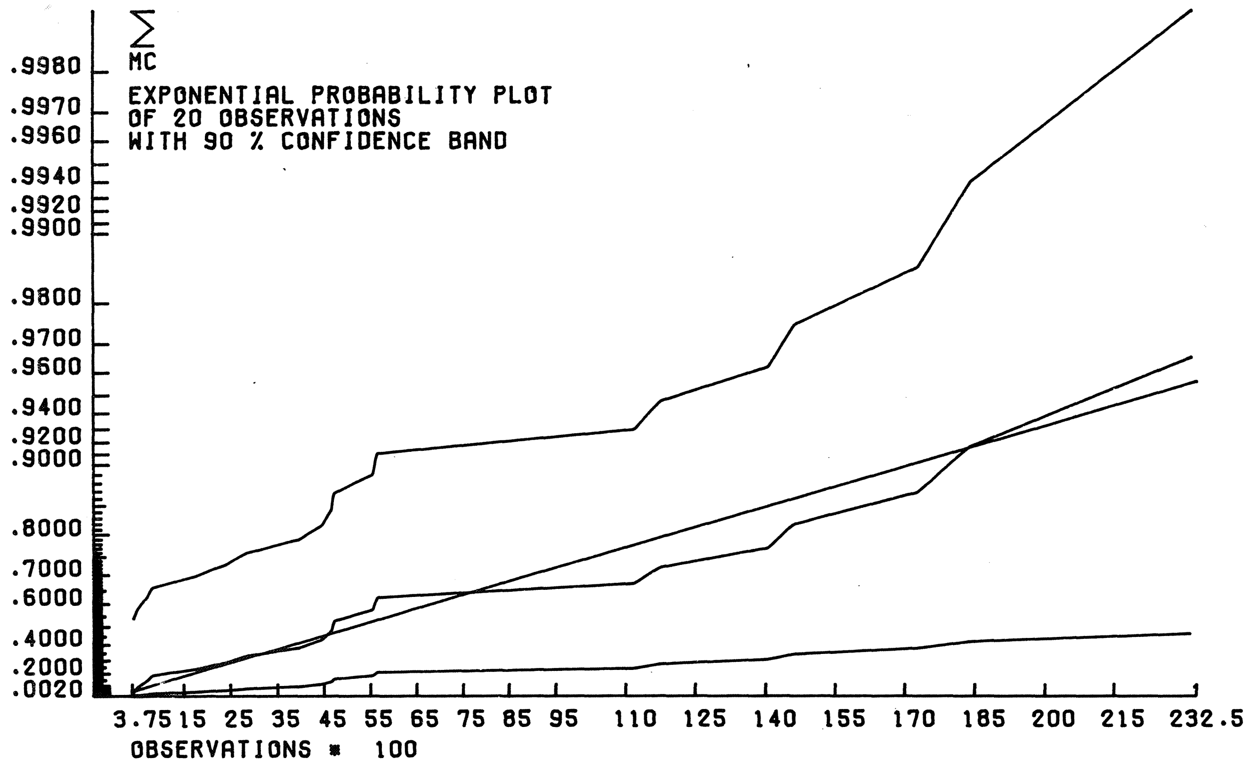


Figure 36. A sample from the exponential (0,1) distribution

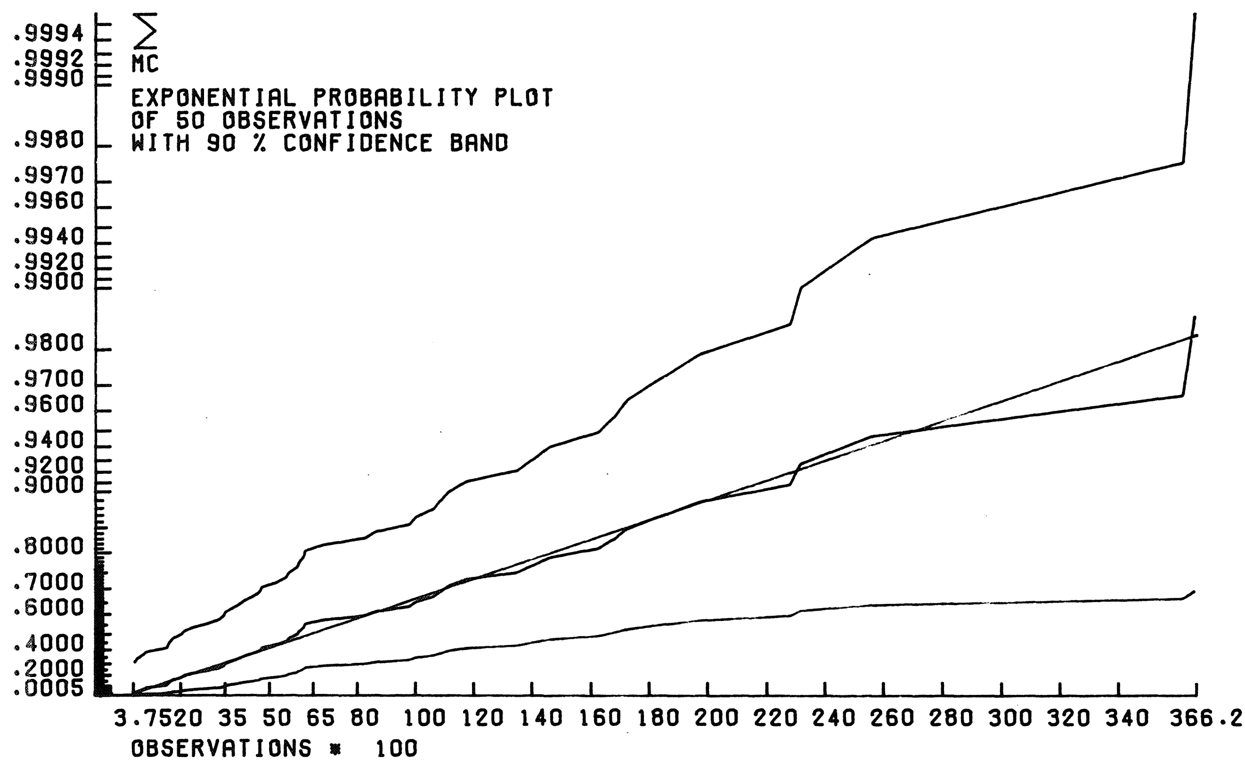


Figure 37. A sample from the exponential (0,1) distribution

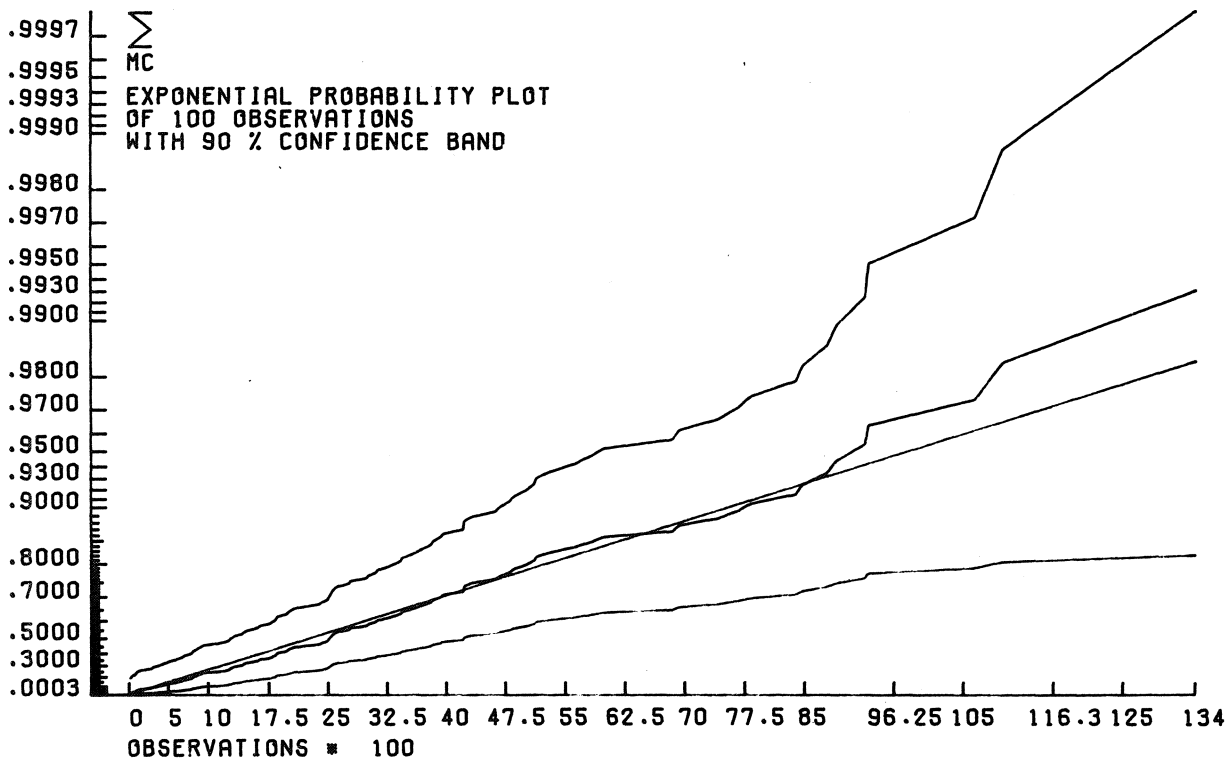


Figure 38. A sample from the exponential (0,3) distribution

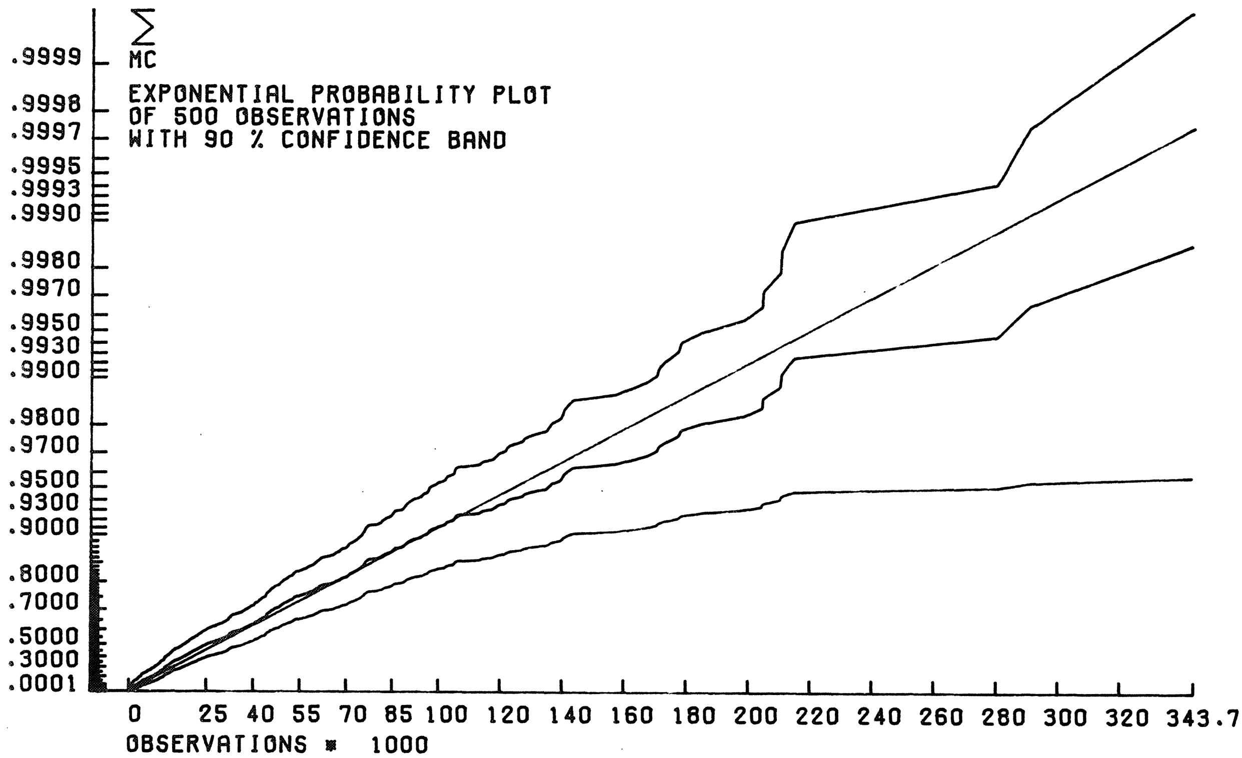


Figure 39. A sample from the exponential (0,25) distribution

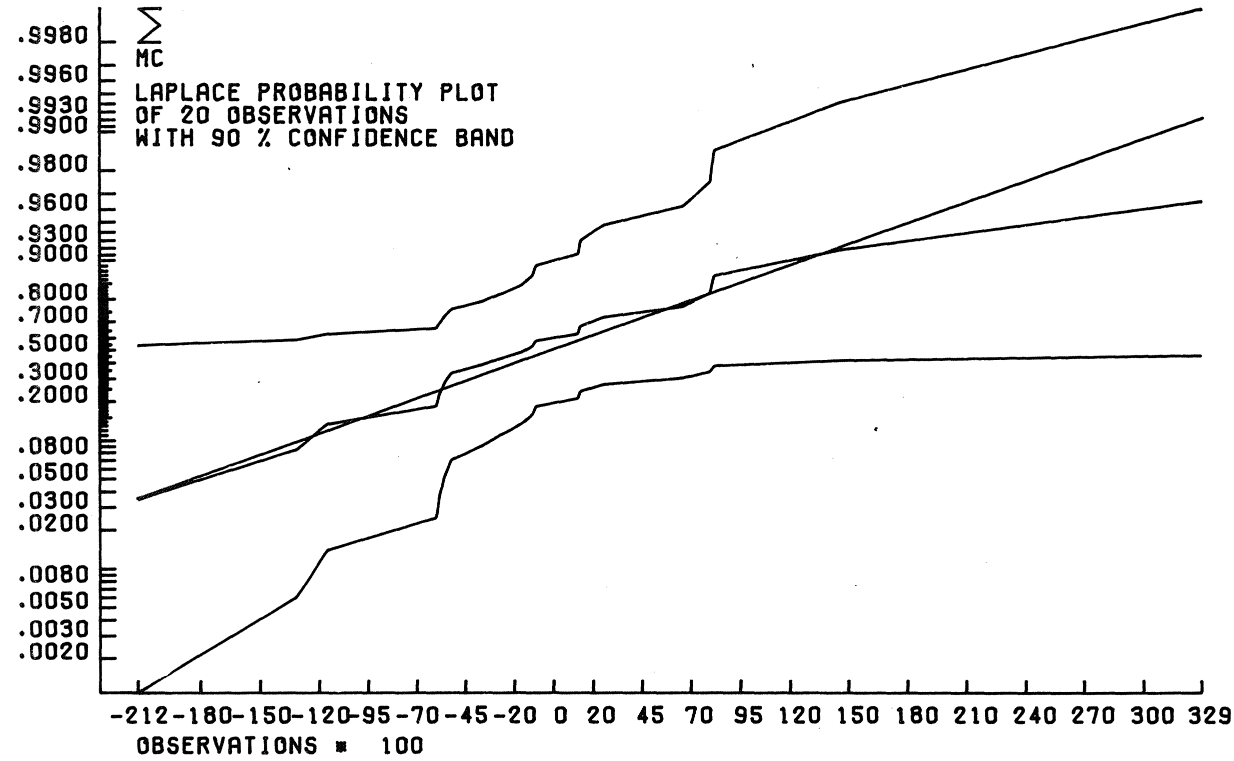


Figure 40. A sample from the Laplace (0,1) distribution

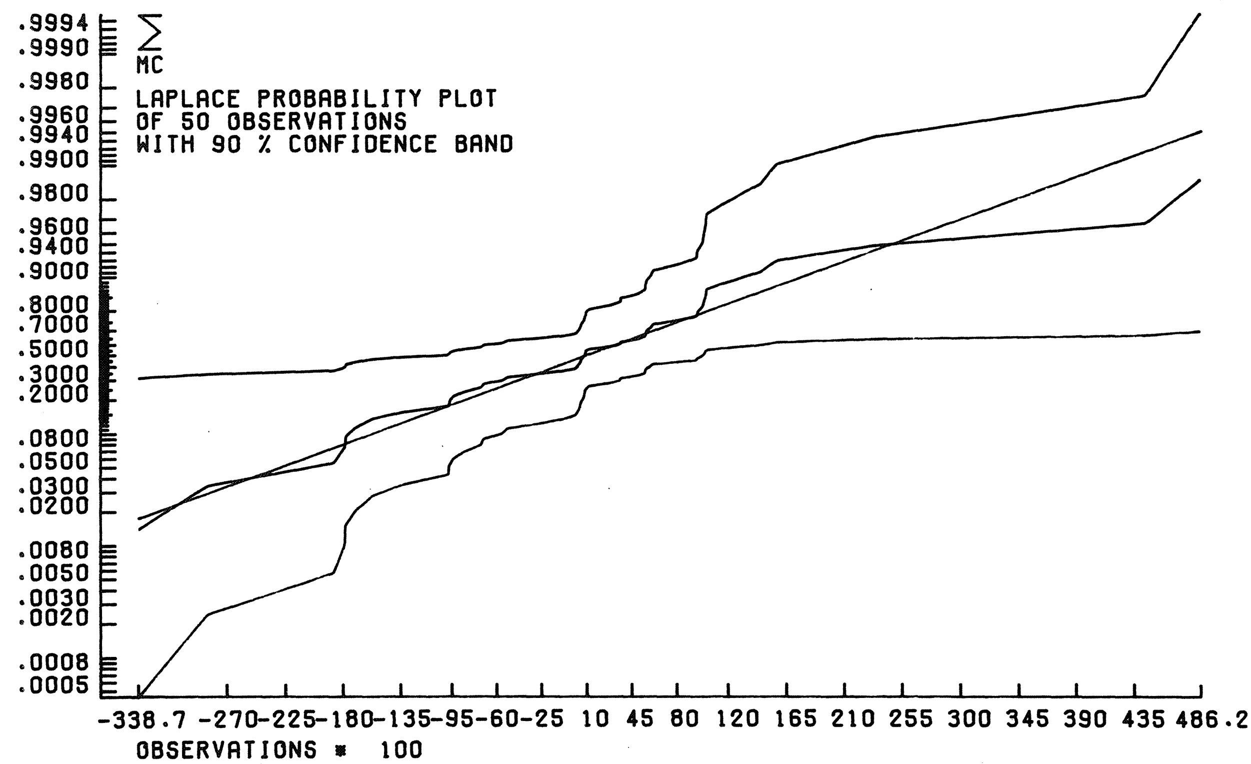


Figure 41. A sample from the Laplace (0,1) distribution

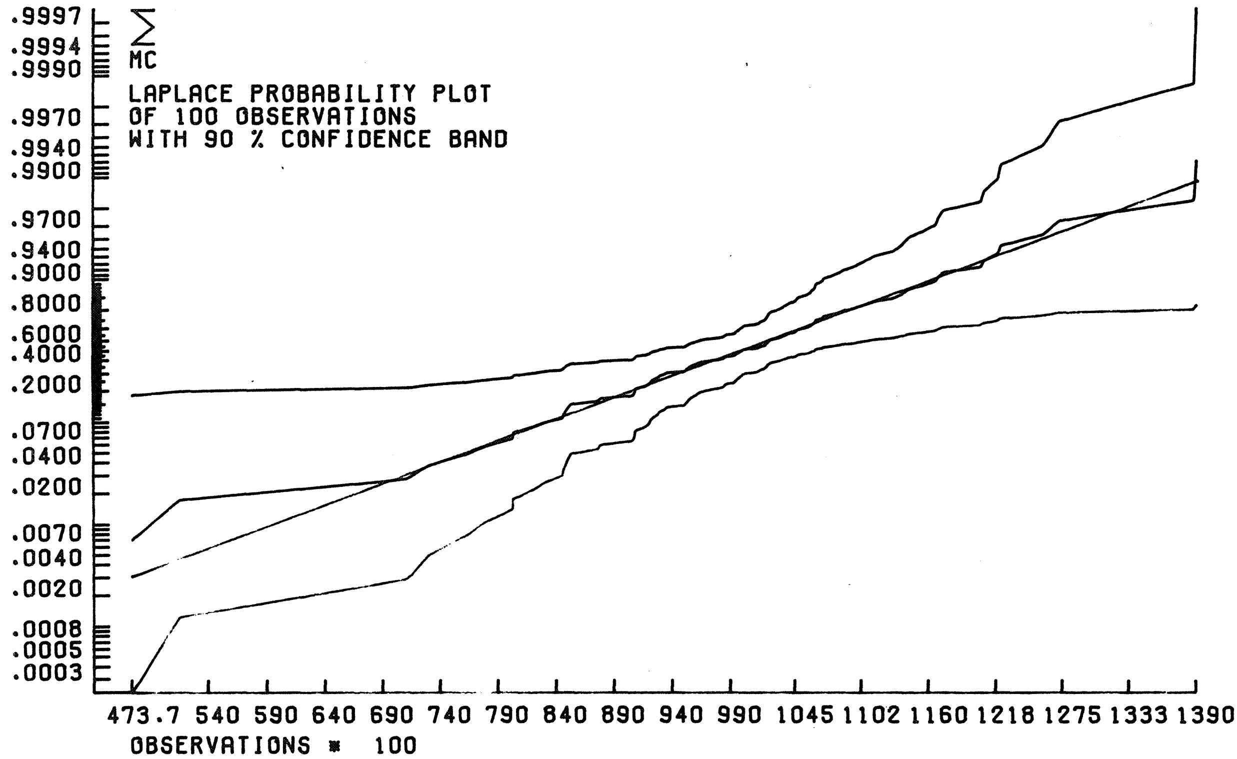


Figure 42. A sample from the Laplace (10,5) distribution

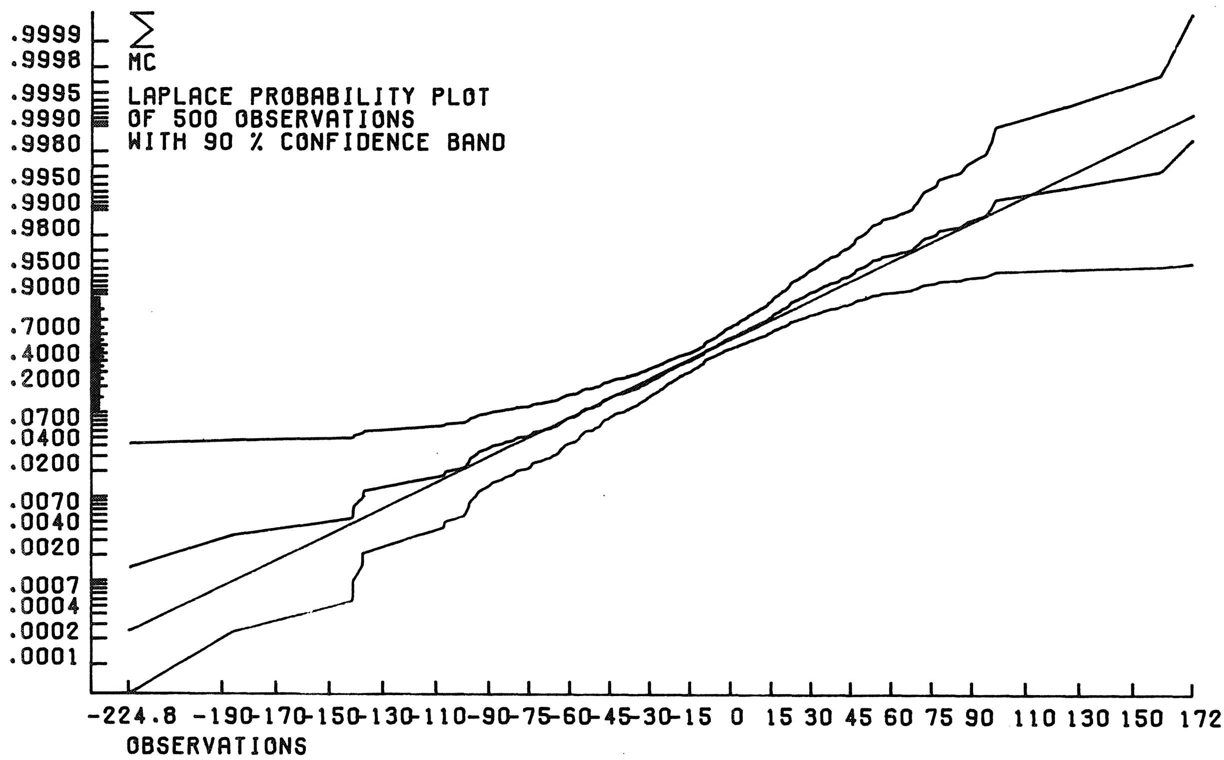


Figure 43. A sample from the Laplace (-10,25) distribution

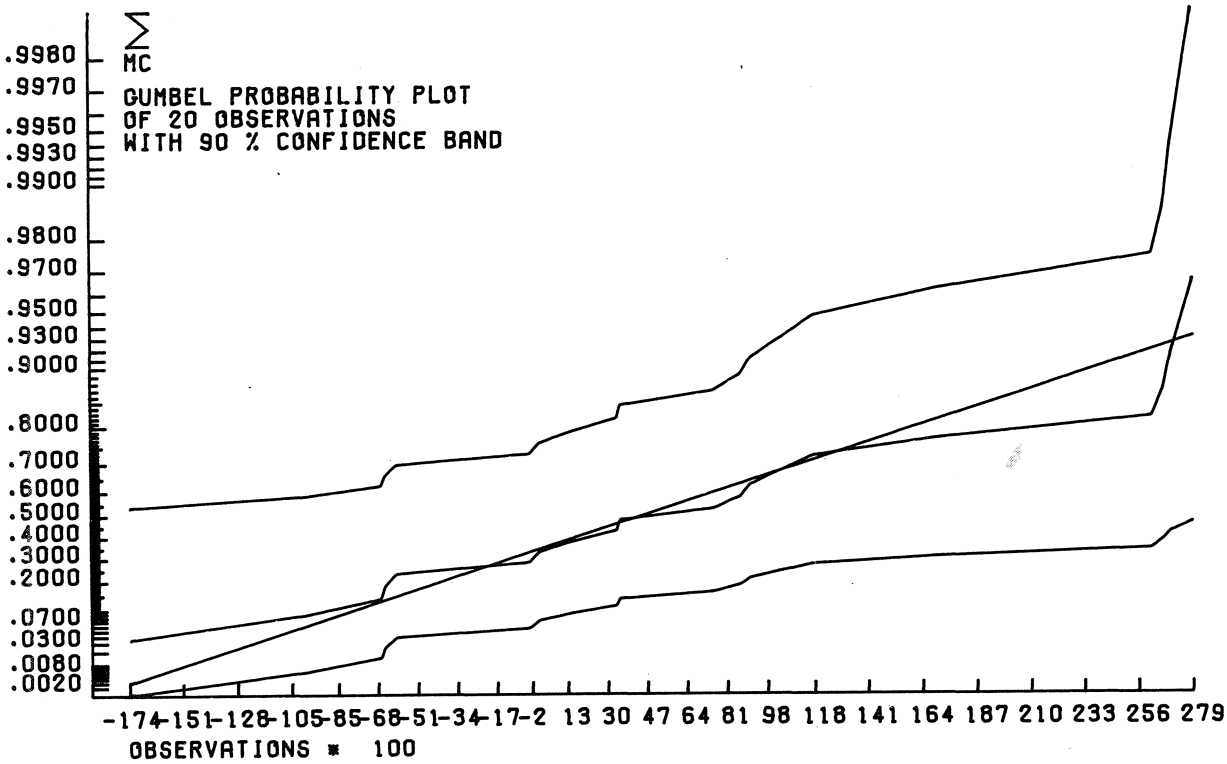


Figure 44. A sample from the Gumbel (0,1) distribution

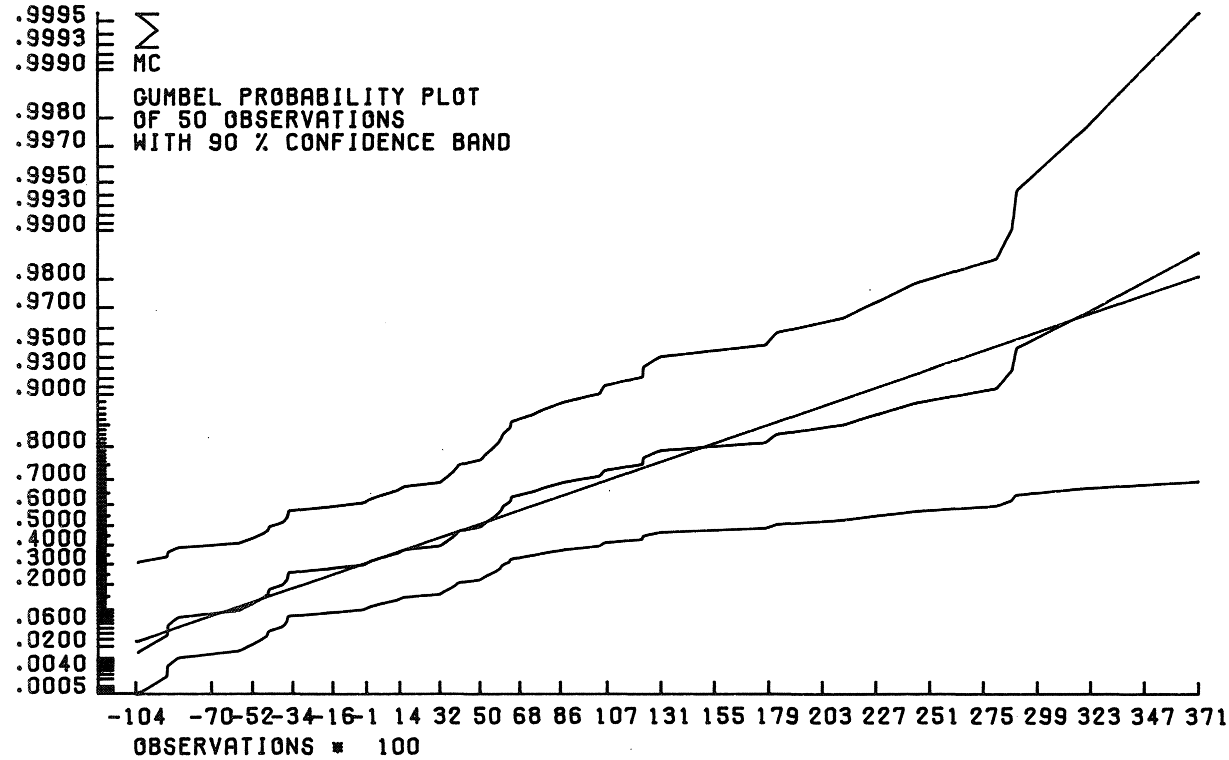


Figure 45. A sample from the Gumbel (0,1) distribution

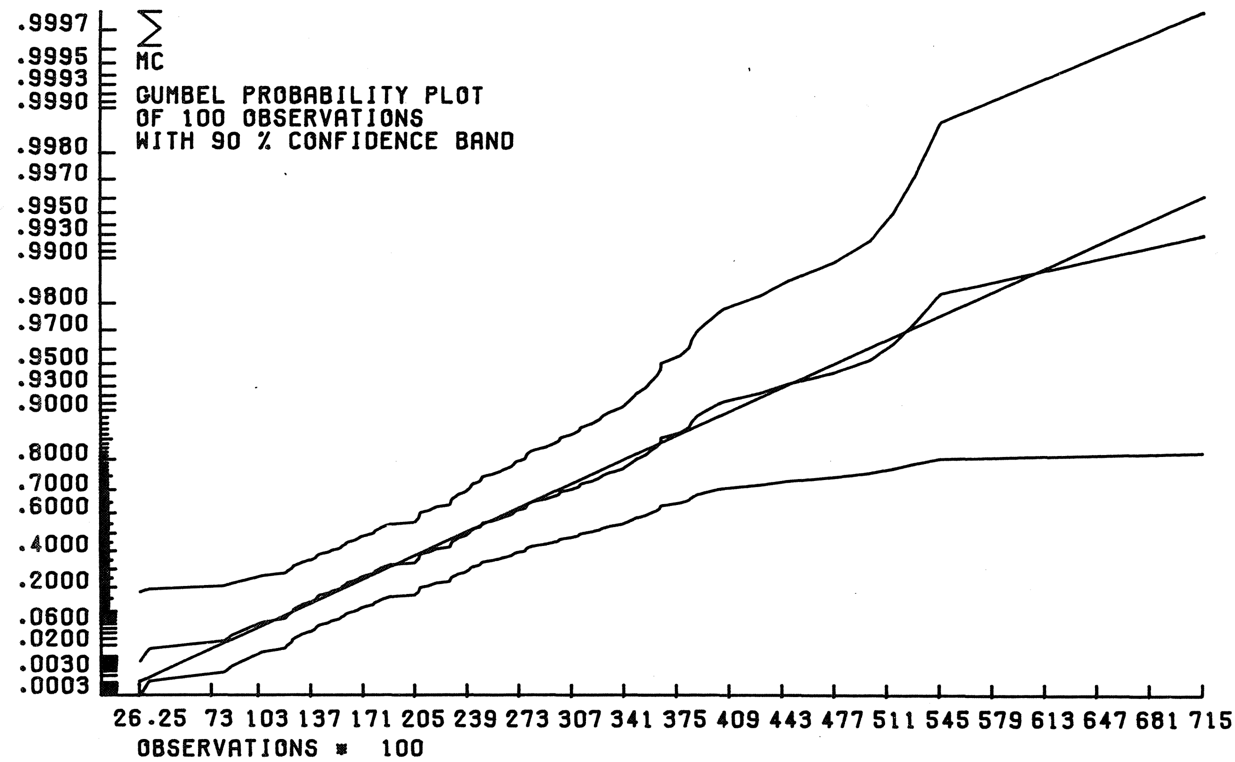


Figure 46. A sample from the Gumbel (2,1) distribution

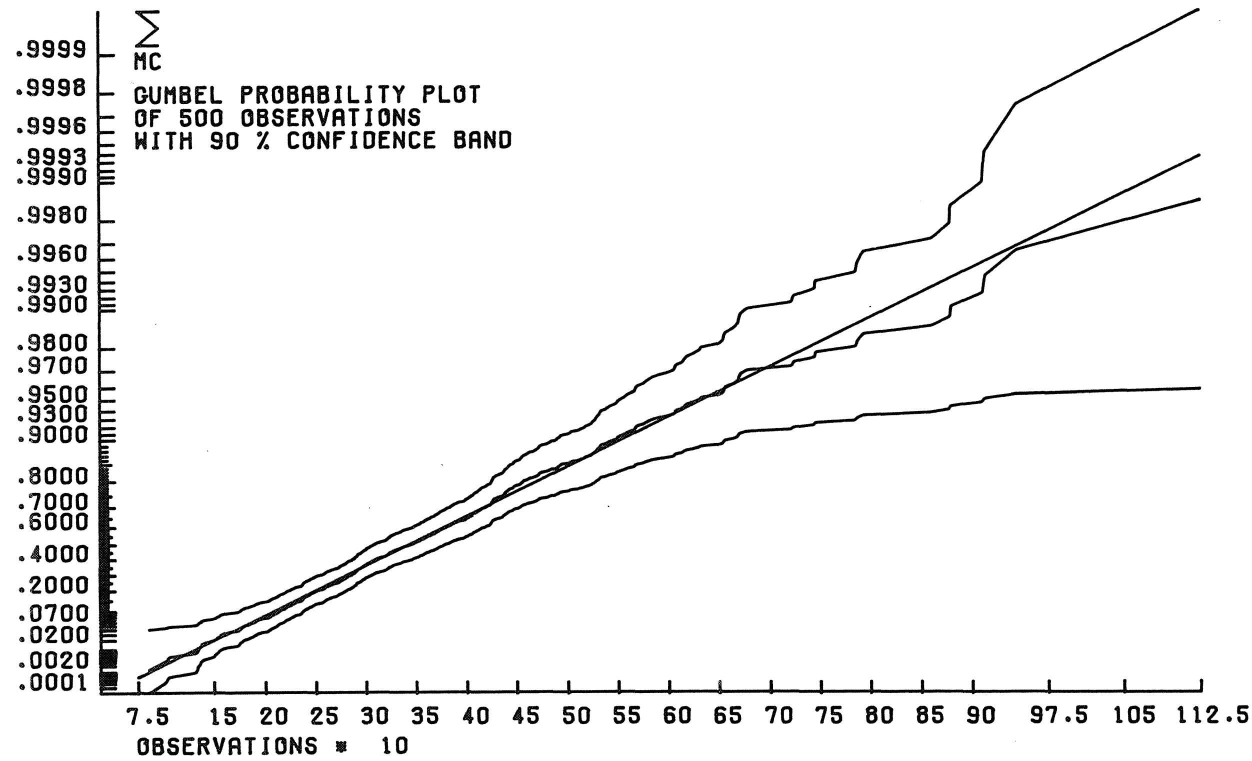


Figure 47. A sample from the Gumbel (3,1) distribution

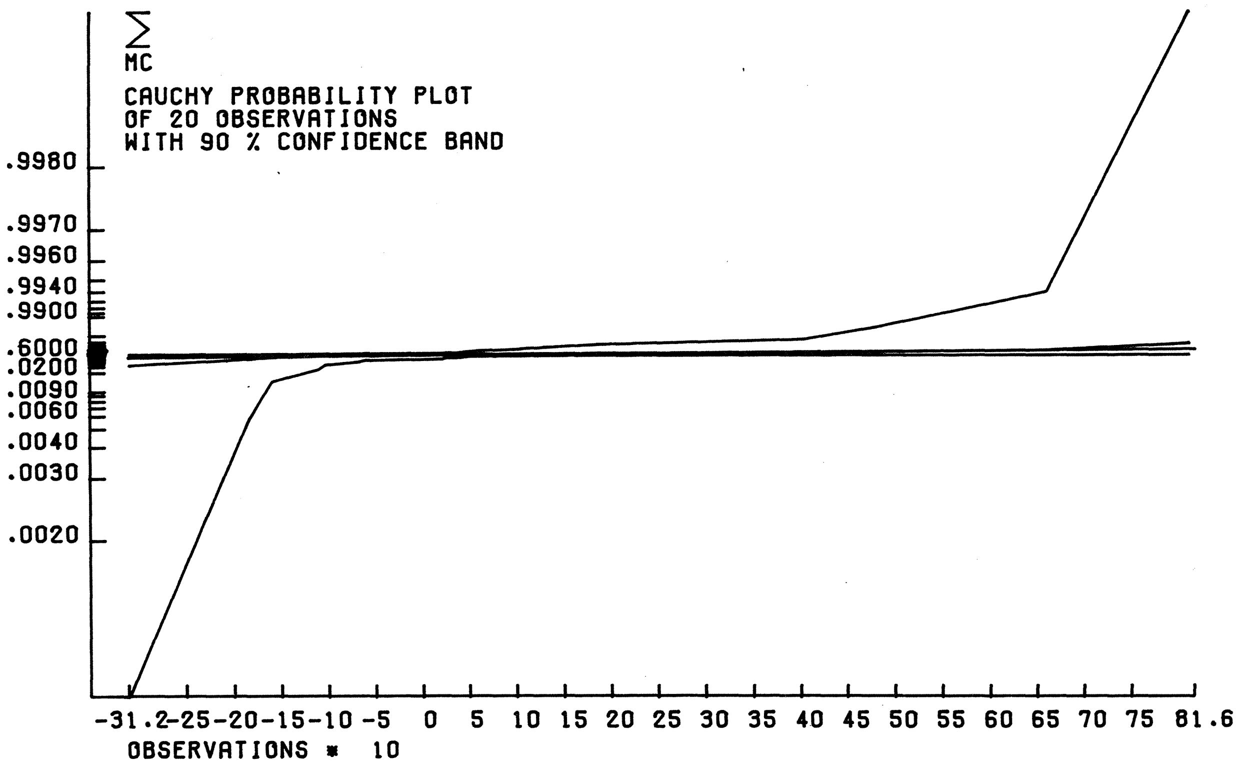


Figure 48. A sample from the Gumbel (0,1) distribution

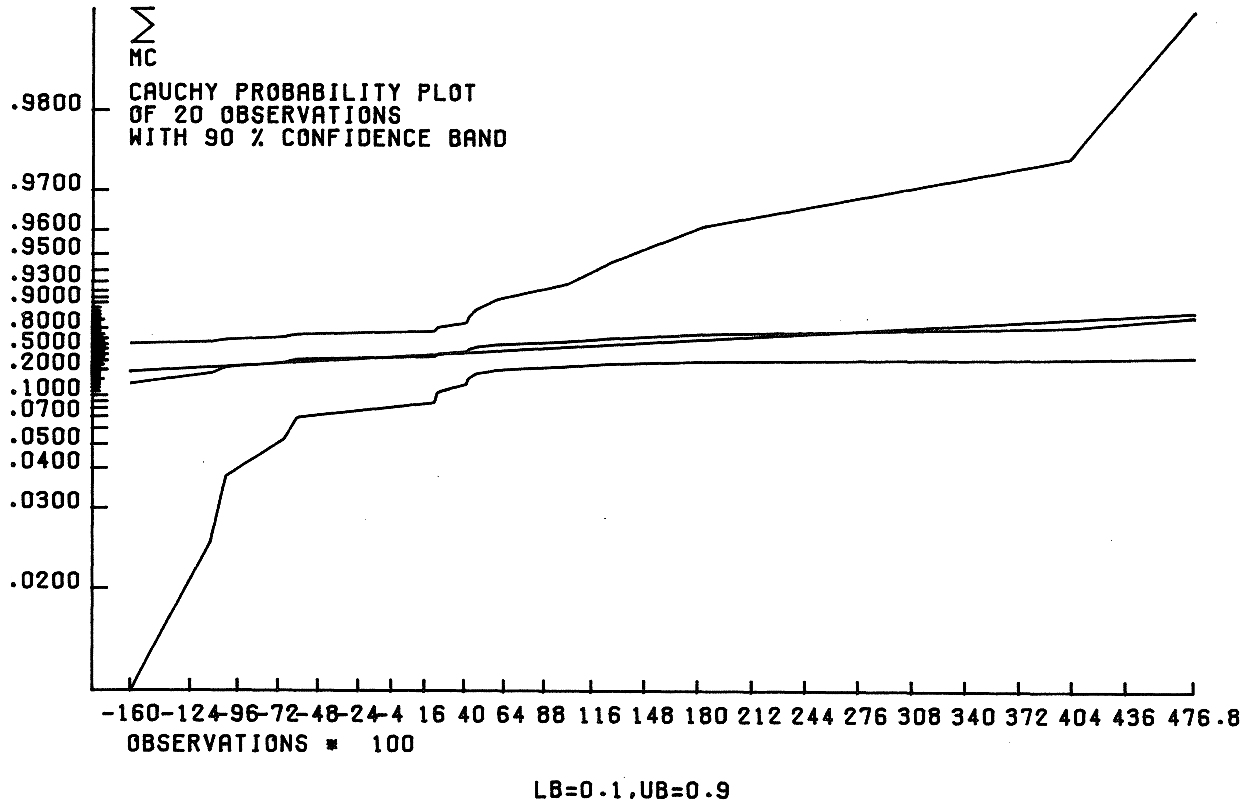


Figure 49. Enlargement of the previous plot

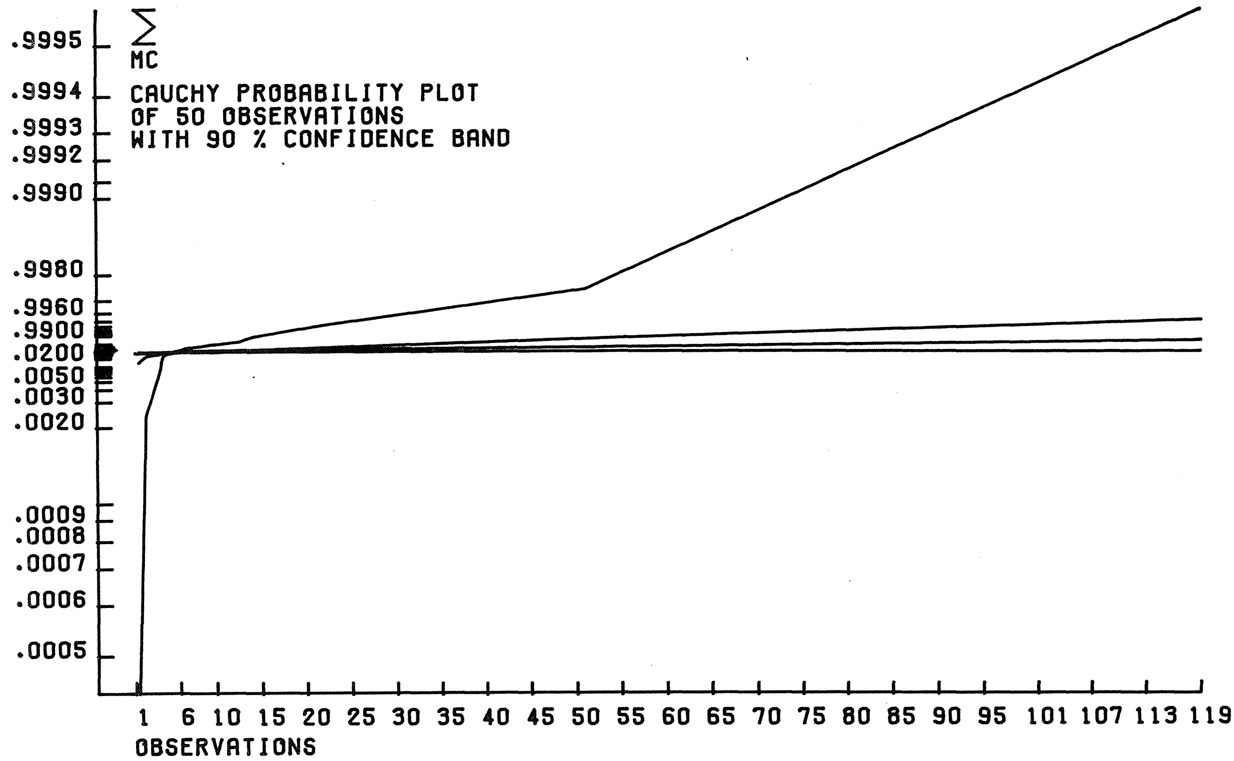


Figure 50. A sample of the Cauchy (5,1) distribution

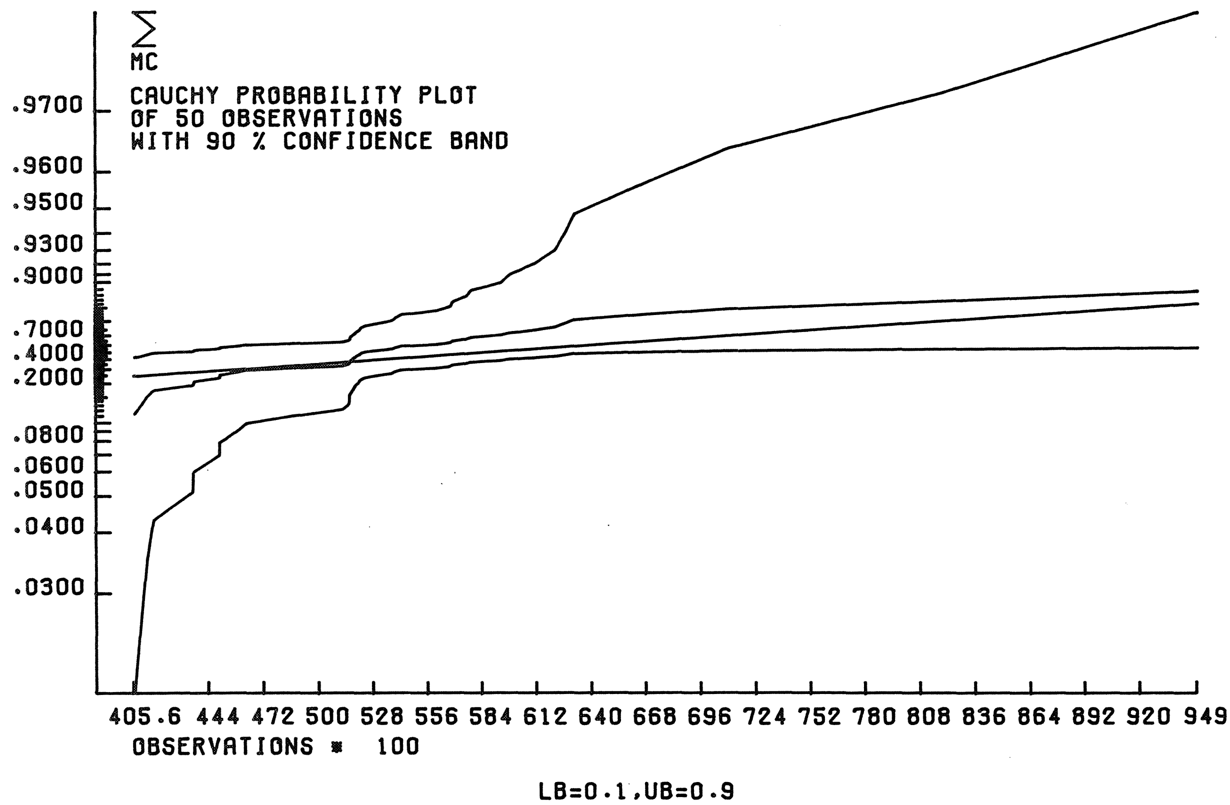


Figure 51. Enlargement of the previous plot

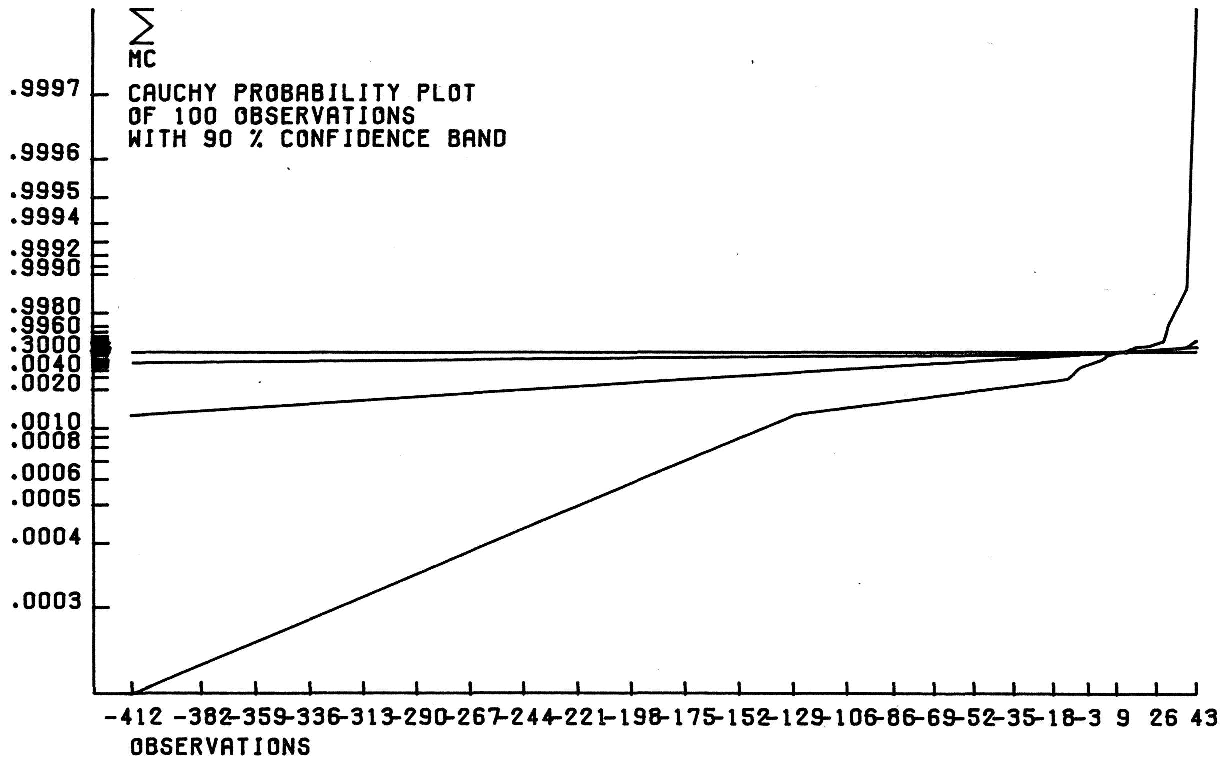


Figure 52. A sample from the Cauchy (10,1) distribution

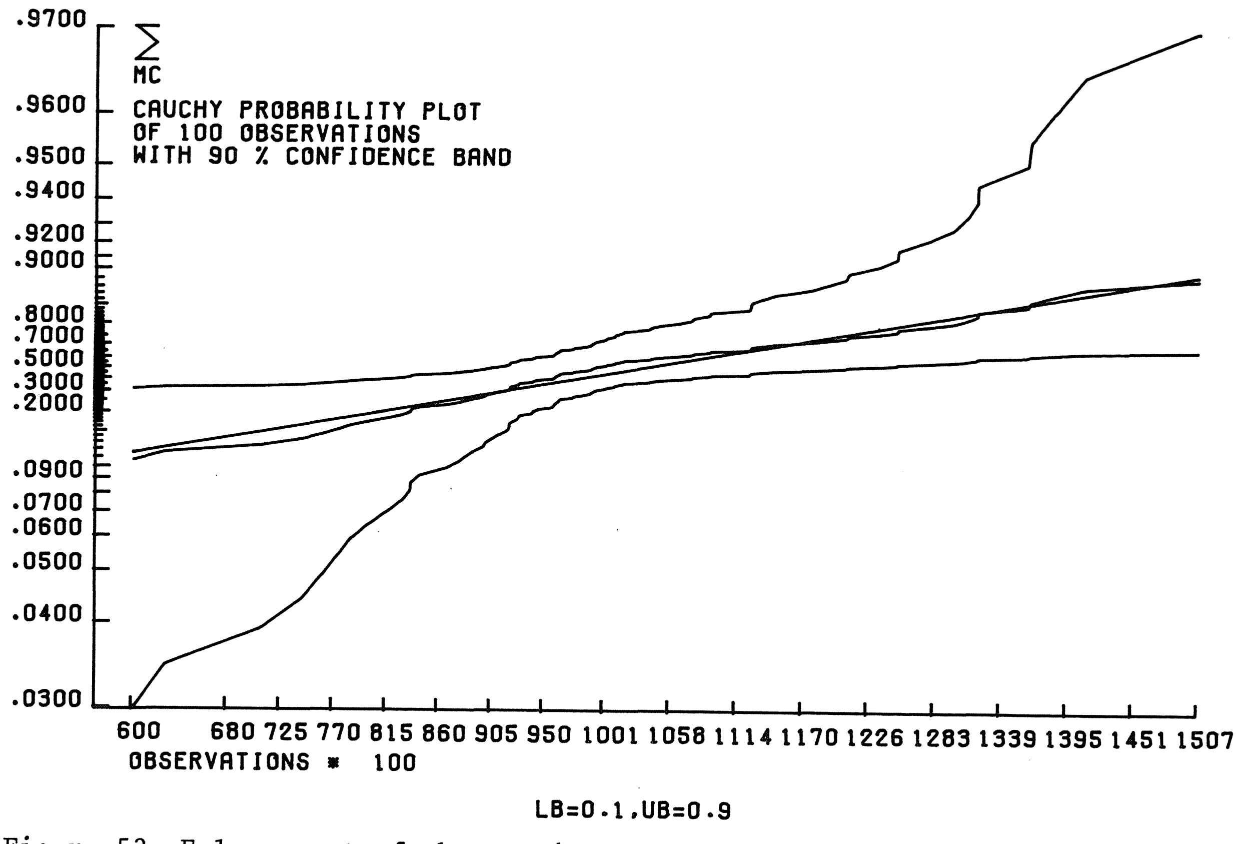


Figure 53. Enlargement of the previous plot

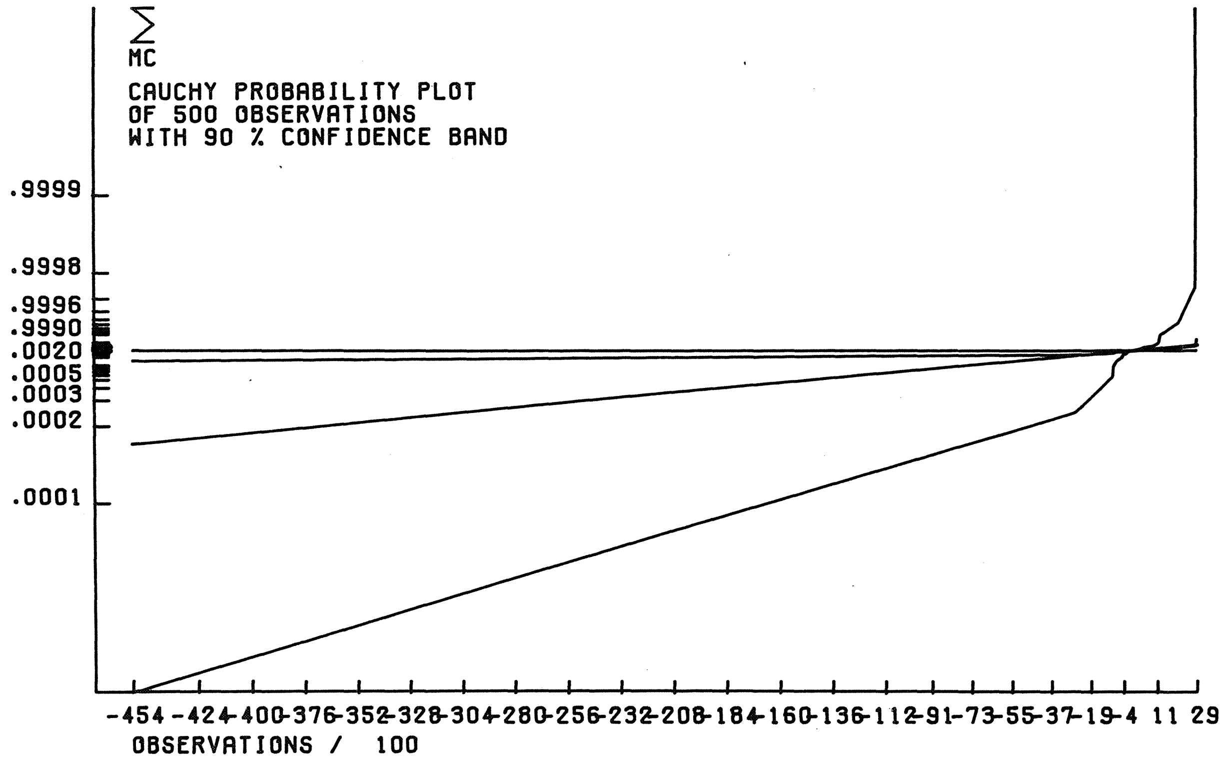


Figure 54. A sample from the Cauchy (-10,25) distribution

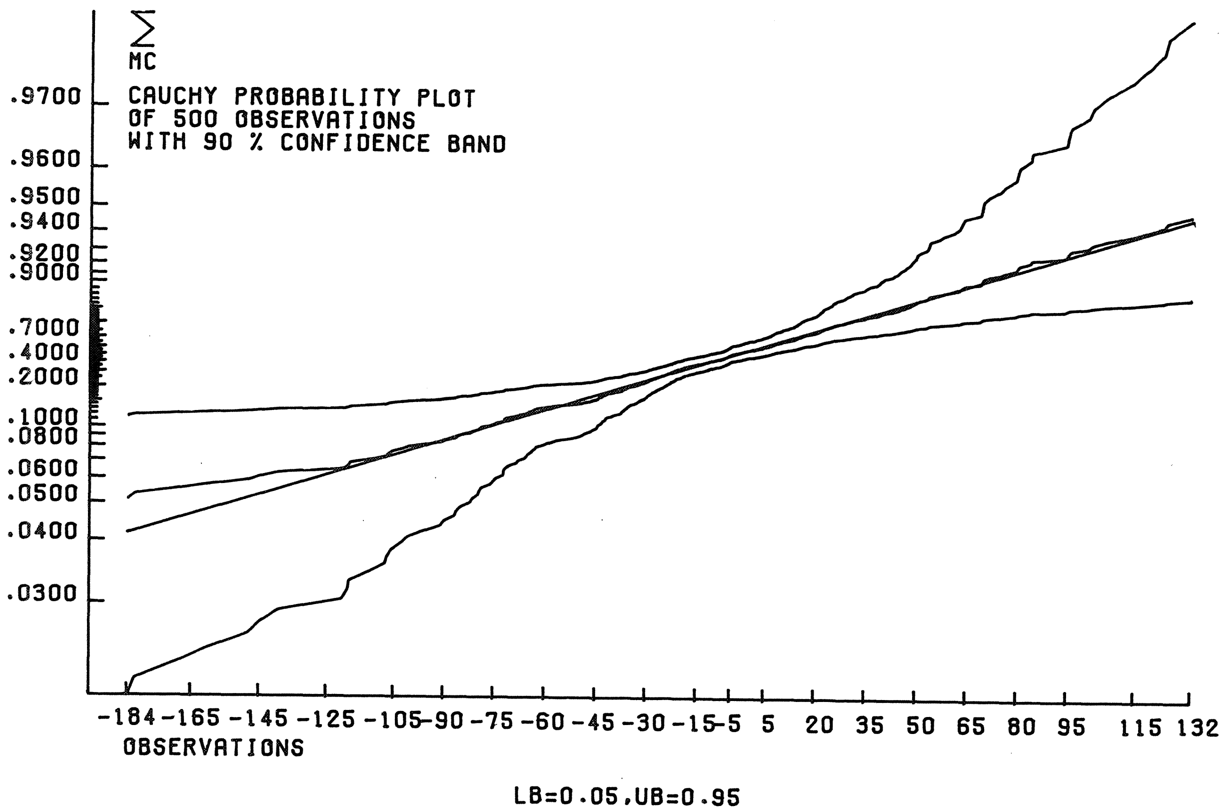


Figure 55. Enlargement of the previous plot

In the following plots the underlying distribution of the sample is denoted by WEIBULL (LOC, SCALE, C), where the parameters are defined in (3.7).

Let us summarize the effect of the parameter PAR

TYPE	PAR	ACTION
WEIBULL 2	> 0	C := *) PAR, observations are plotted in a WEIBULL (c fixed) prob. plot
WEIBULL 2	0	LOC := 0, -ln (observations) are plotted in a GUMBEL prob. plot
WEIBULL 3	≠ 0	LOC := PAR, -ln (observations-LOC) are plotted in a GUMBEL prob. plot
WEIBULL 3	0	LOC is estimated, -ln(observations-LOC) are plotted in a GUMBEL prob. plot

*) read: becomes, attains the value

Note that in the remaining part of 5.2.3.1 the samples are equal iff

1. the sample sizes are equal
2. the corresponding parameters LOC, SCALE and C are equal.

Hence the samples in the first, second and third plot are equal etc..

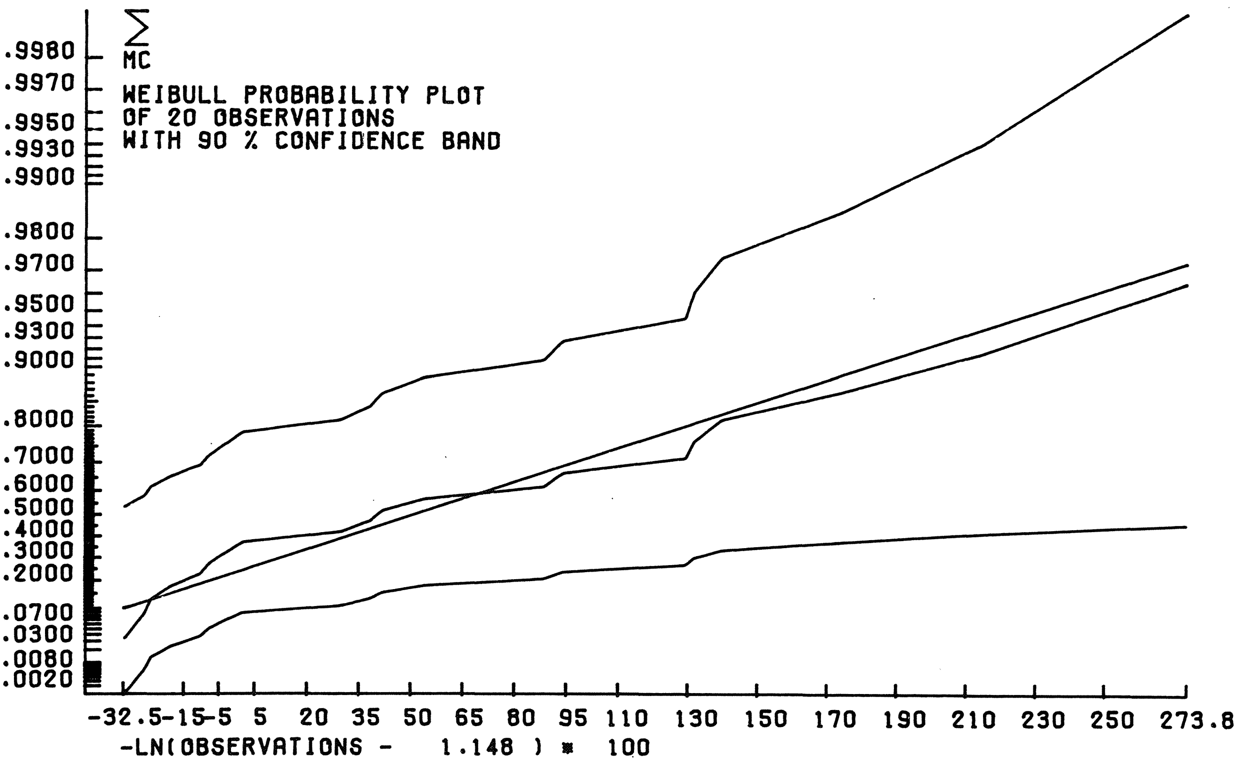


Figure 56.

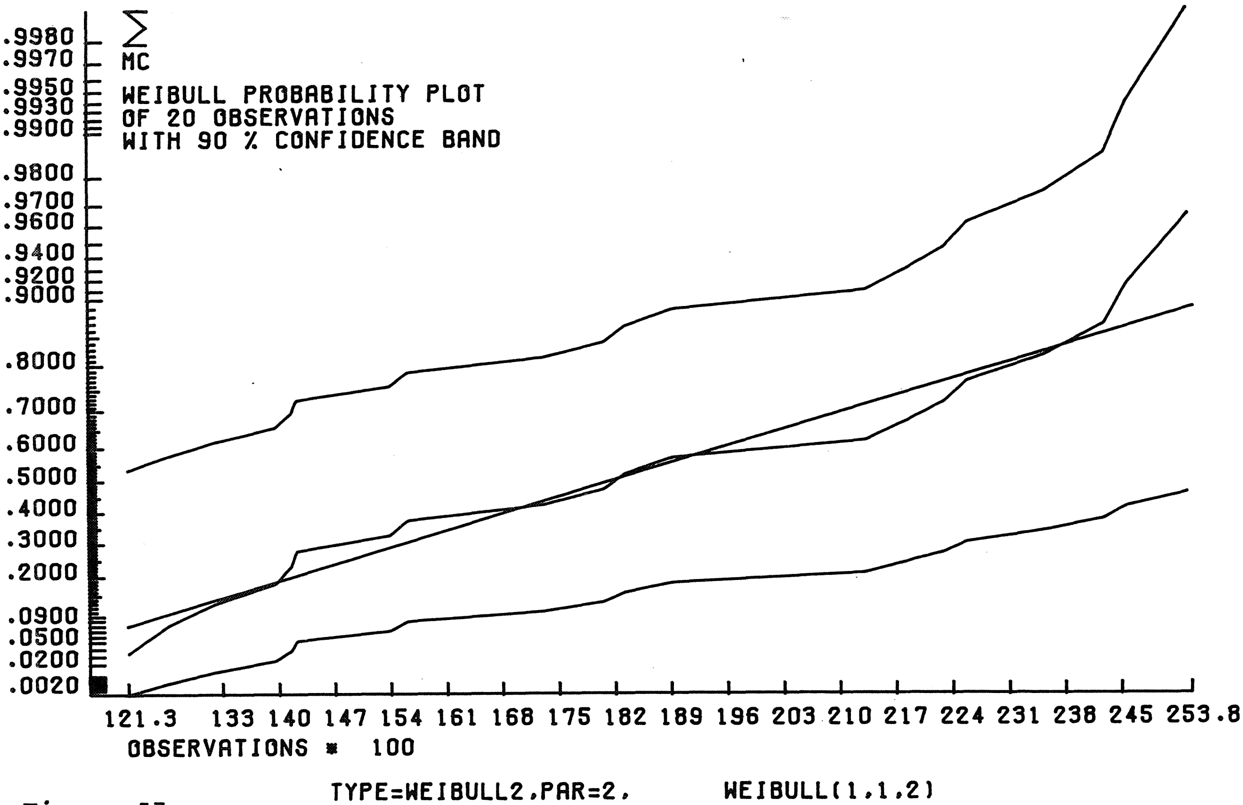


Figure 57.

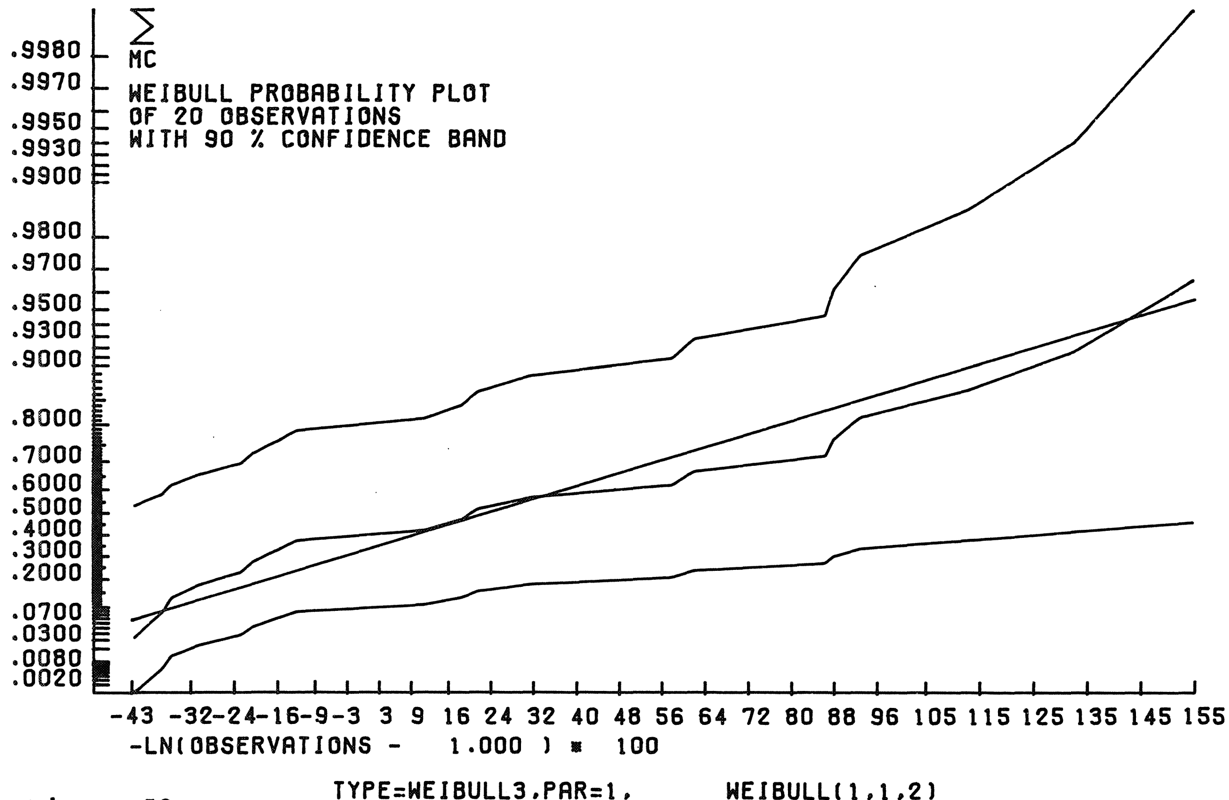


Figure 58.

The following three plots suggest how a preliminary guess of the parameter C may be obtained when one is sure to have a sample from a Weibull distribution.

The next four plots (Figures 62,63,64,65) show that estimating LOC graphically seems more difficult.

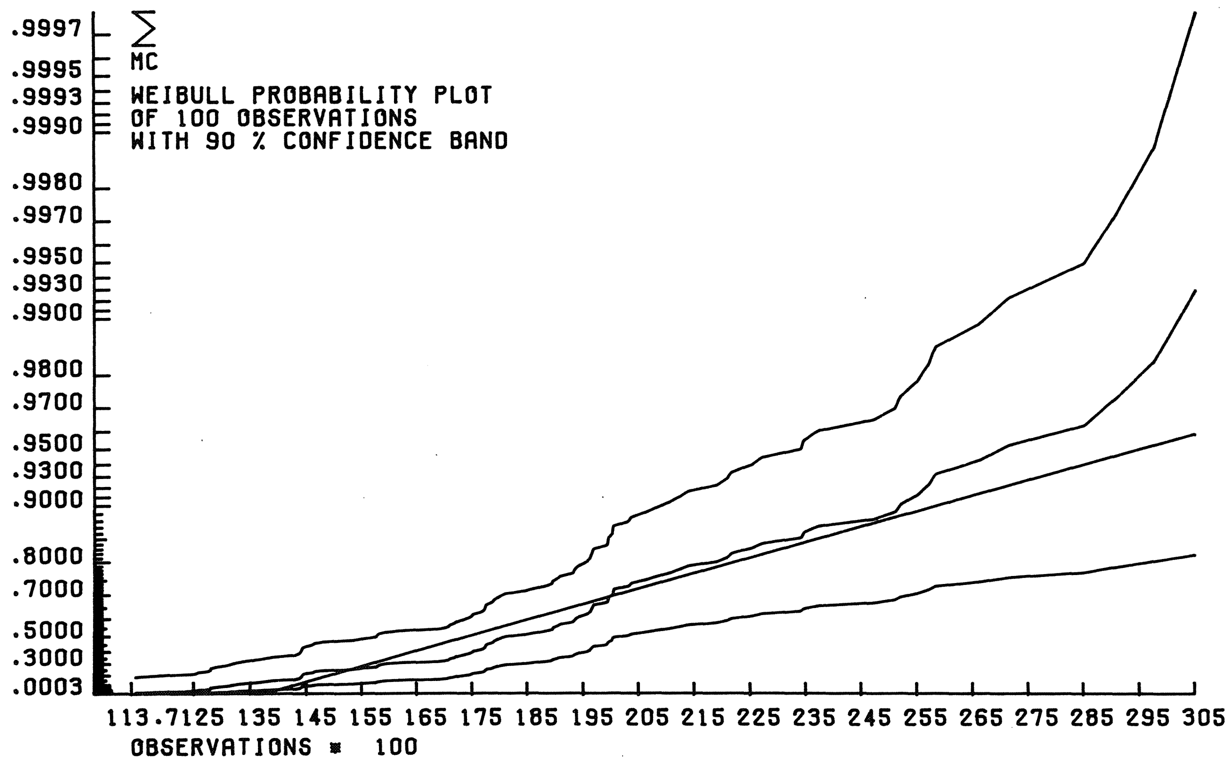


Figure 59.

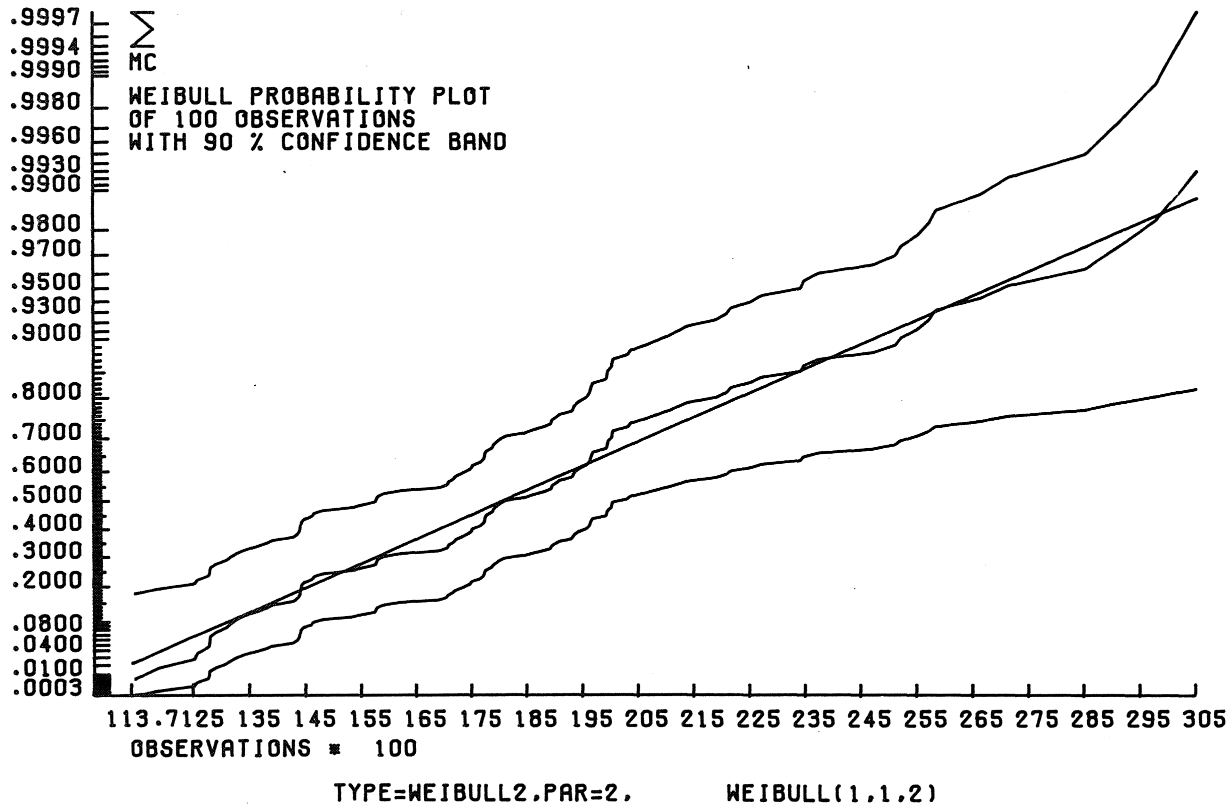


Figure 60.

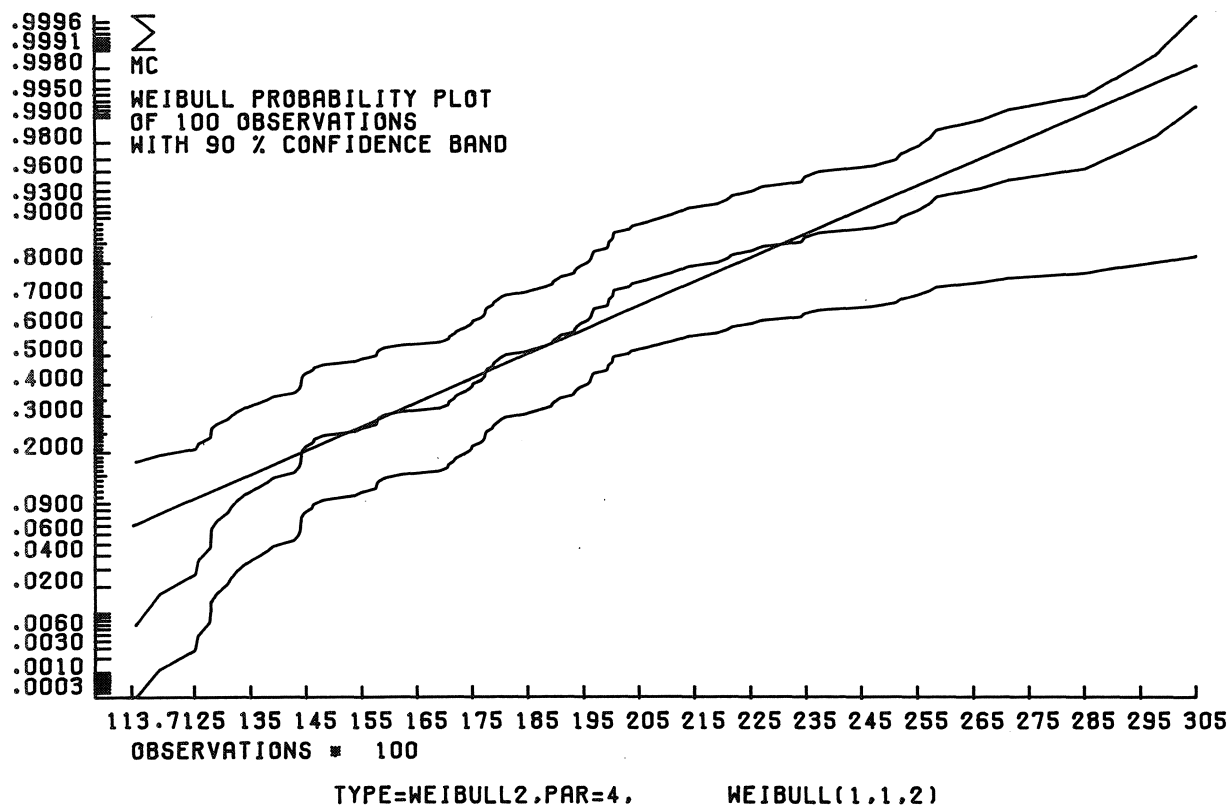


Figure 61.

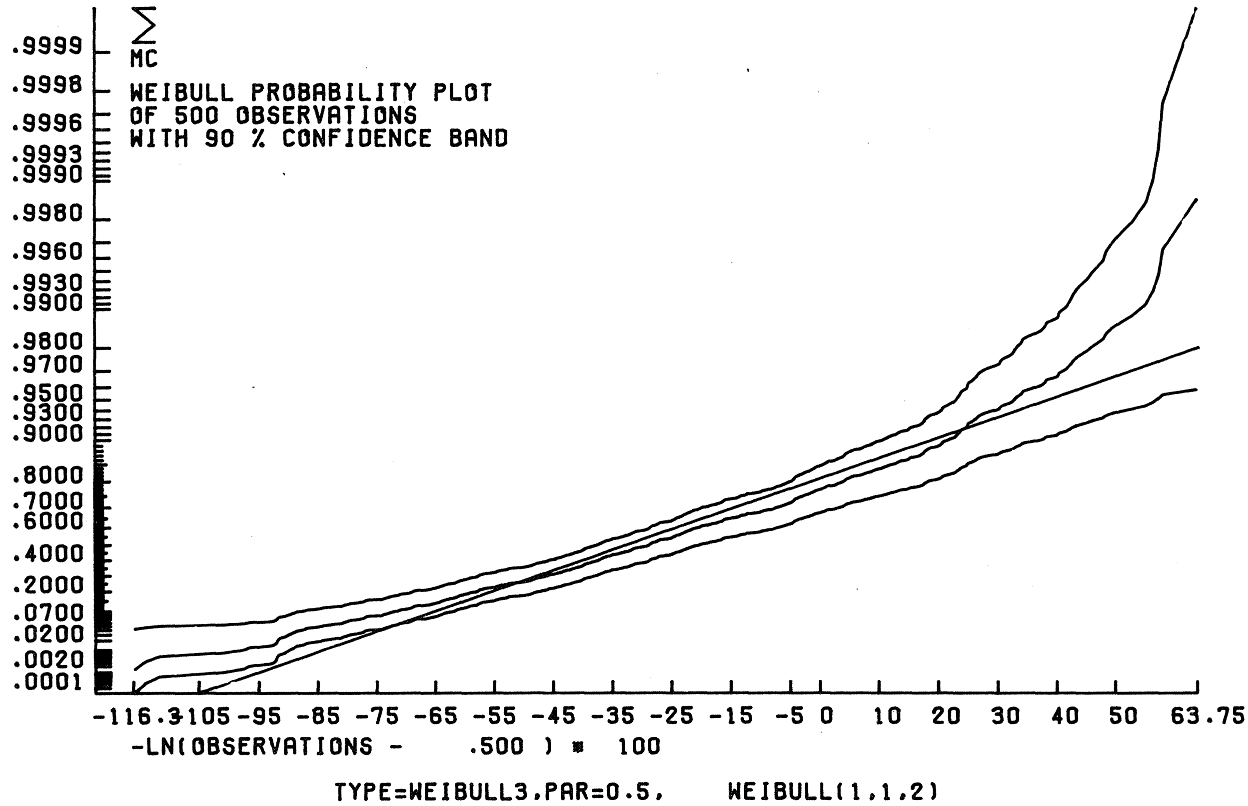


Figure 62.

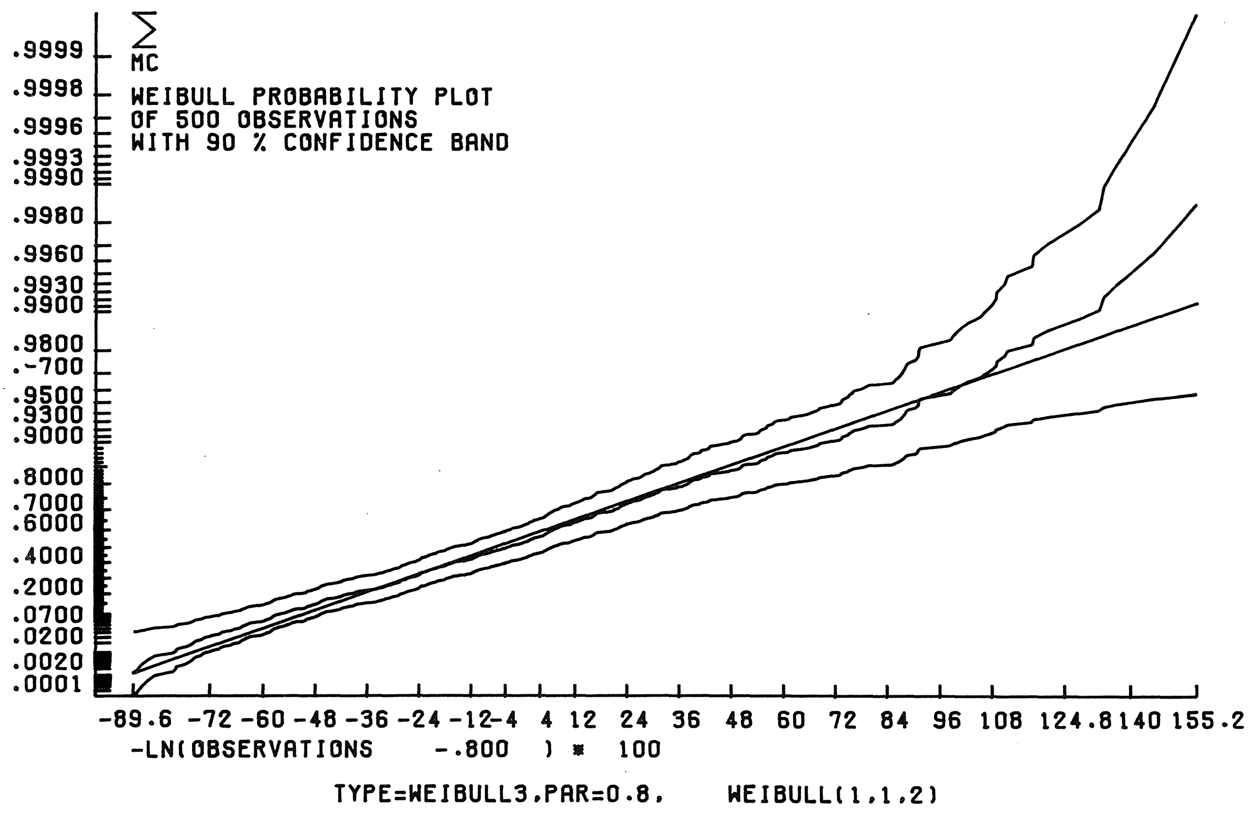


Figure 63.

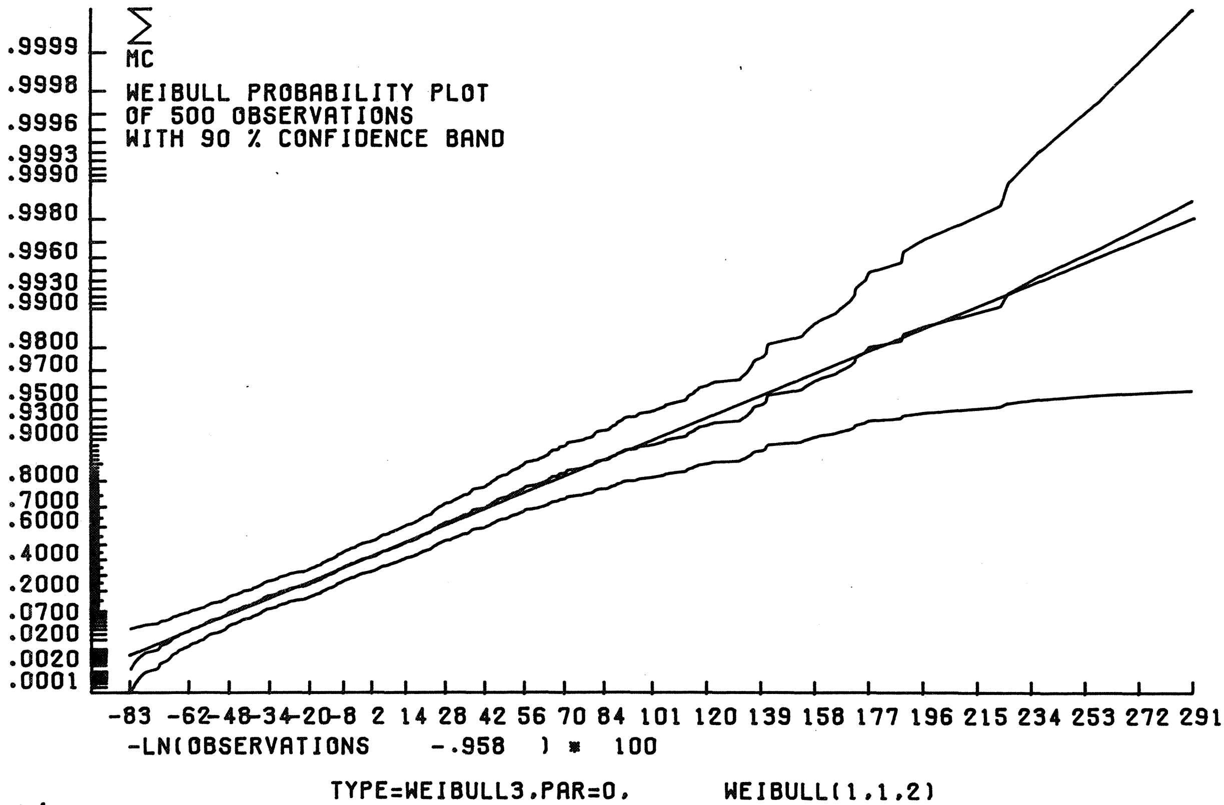


Figure 64.

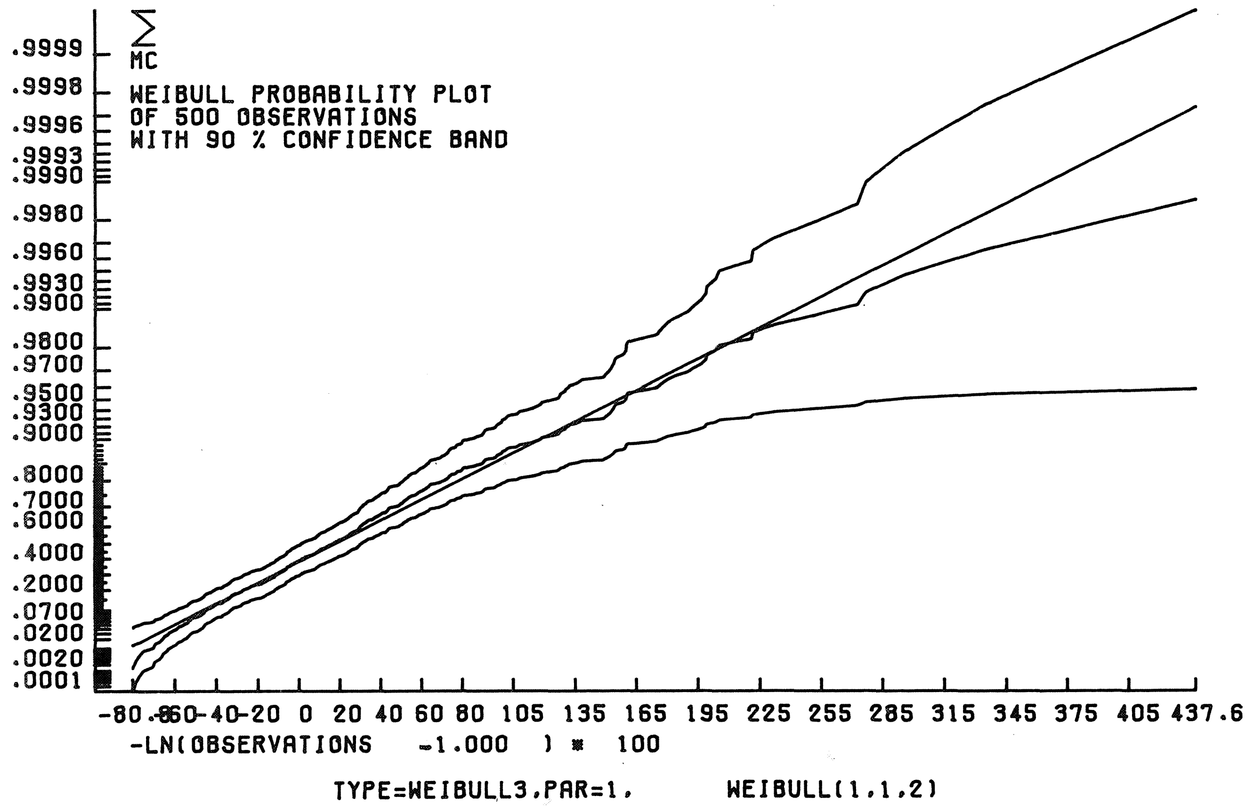


Figure 65.

Since the Weibull (25,1,0.25) distribution has heavy tails, its probability plot of type WEIBULL2 is not very informative (Figure 66). It may be improved by deleting the confidence band and by enlargement (Figures 67 and 68).

Note that -ln(observations-LOC) is a decreasing transformation of the observations, which implies that their order is reversed in WEIBULL 3 plots (Figures 69 and 70).

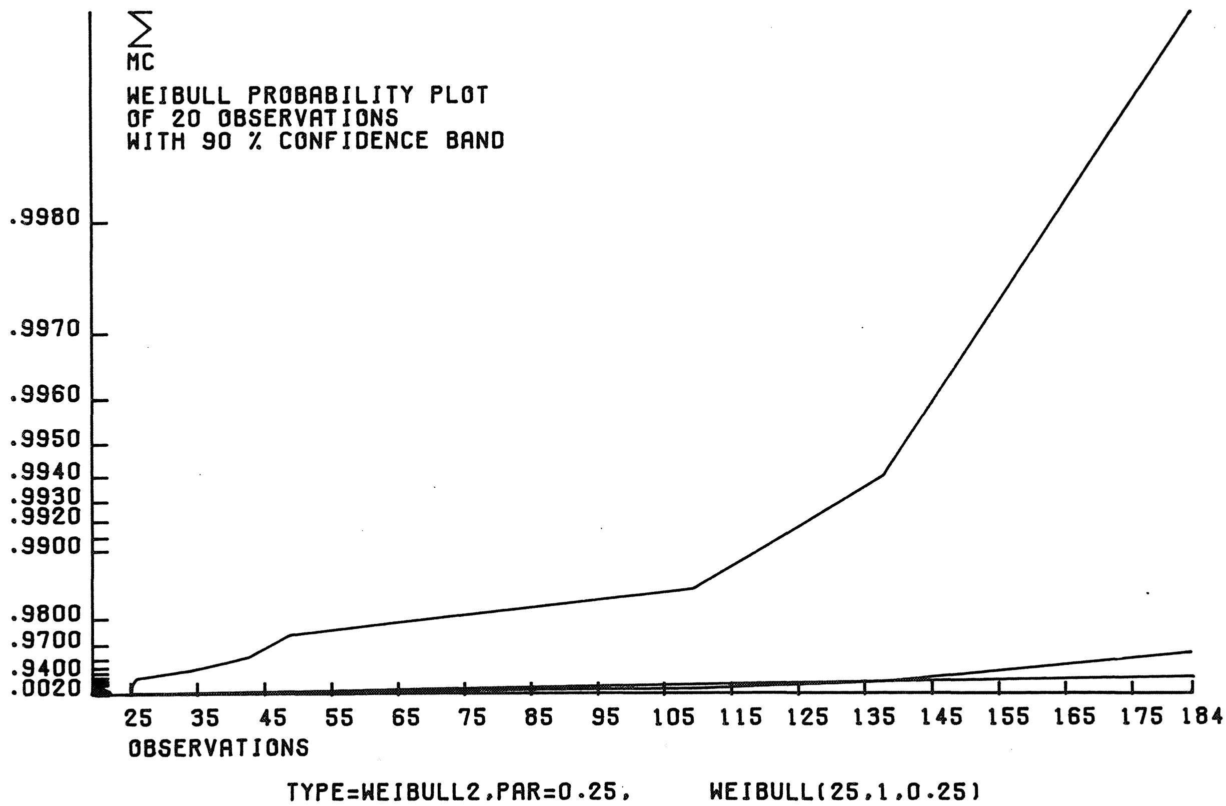


Figure 66.

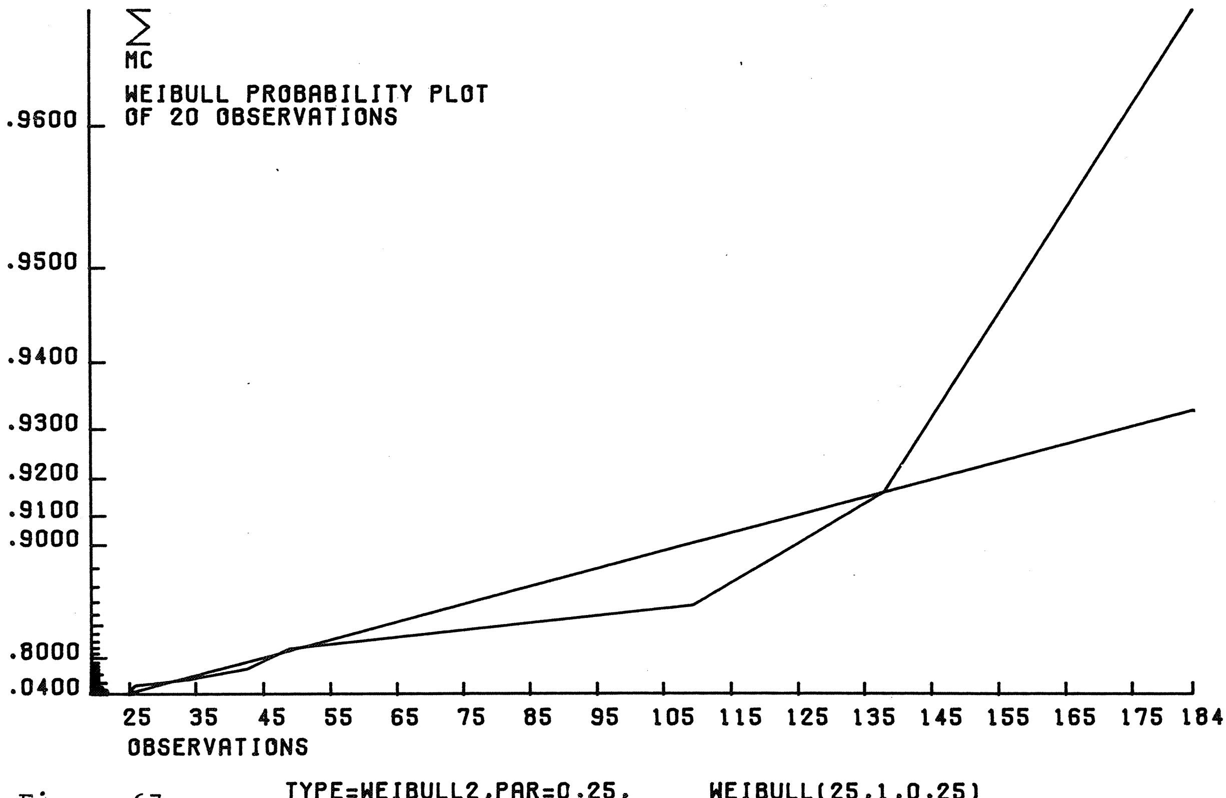


Figure 67.

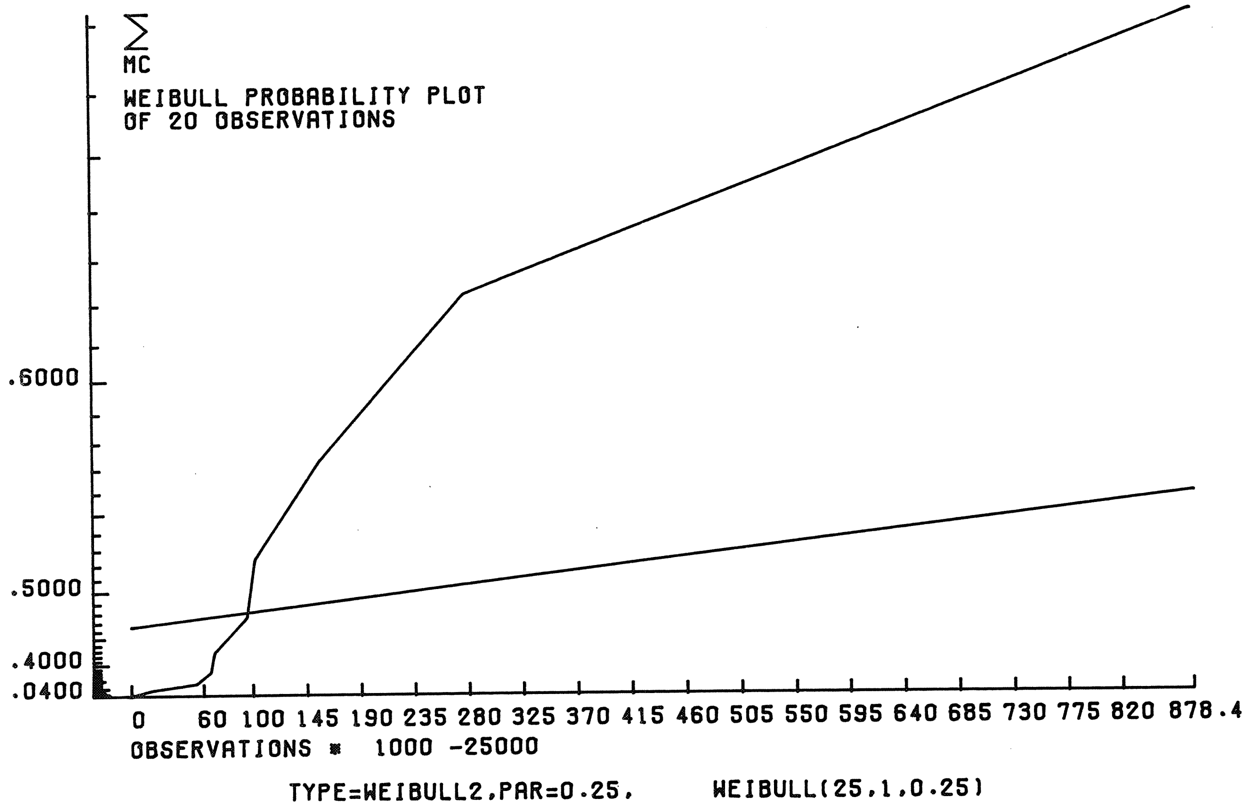


Figure 68. Enlargement of the previous plot

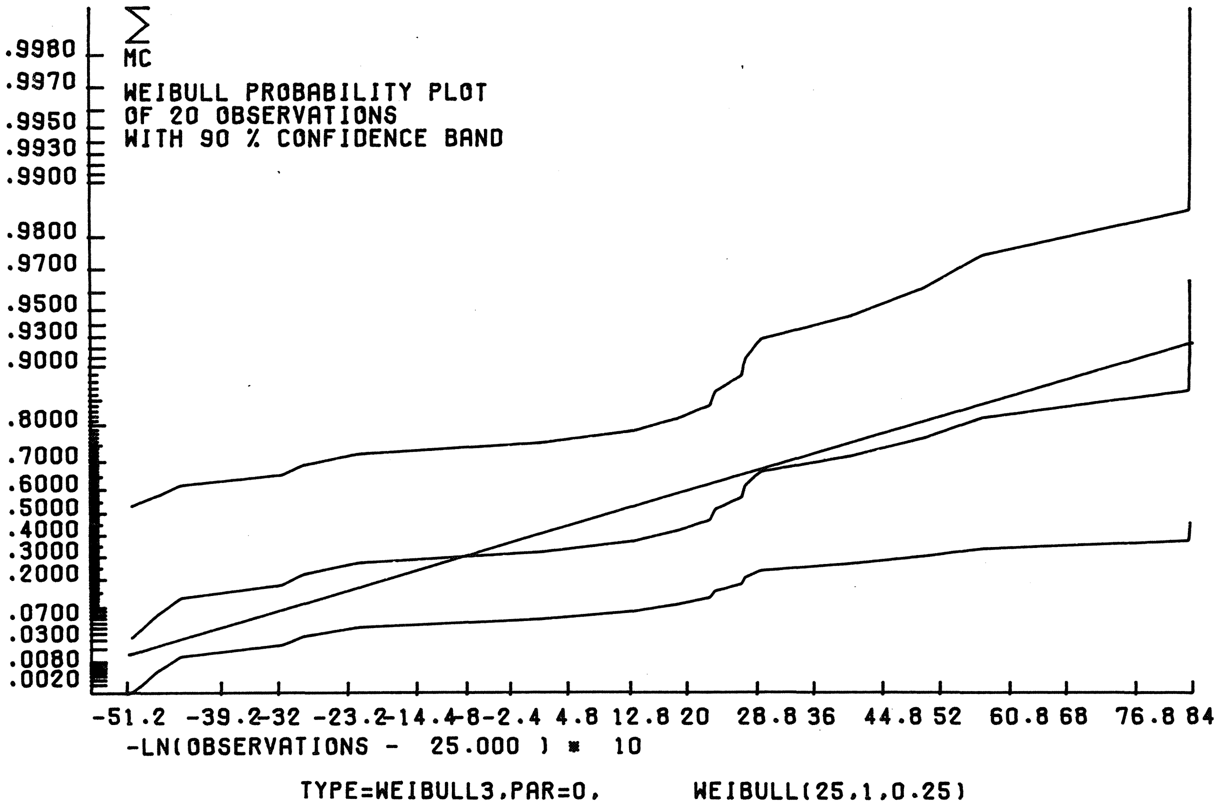


Figure 69. Location estimate is 24.9997!

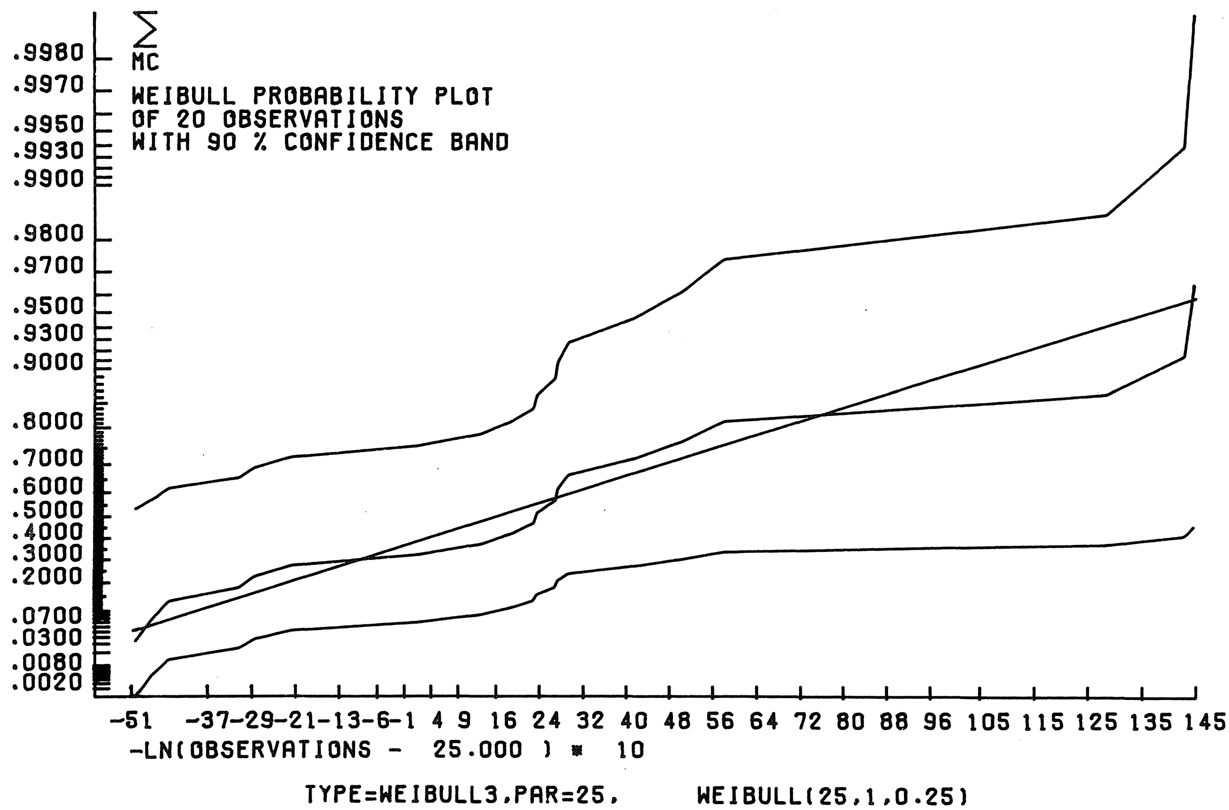


Figure 70.

The following nine plots show what may happen when one tries to estimate the parameter C graphically in this case, where the three parameters of the distribution are supposed to be unknown. Figures 71, 72 and 73 include a confidence band, unlike Figures 74, 75 and 76. Figures 77, 78 and 79 are enlargements of the three preceeding plots.

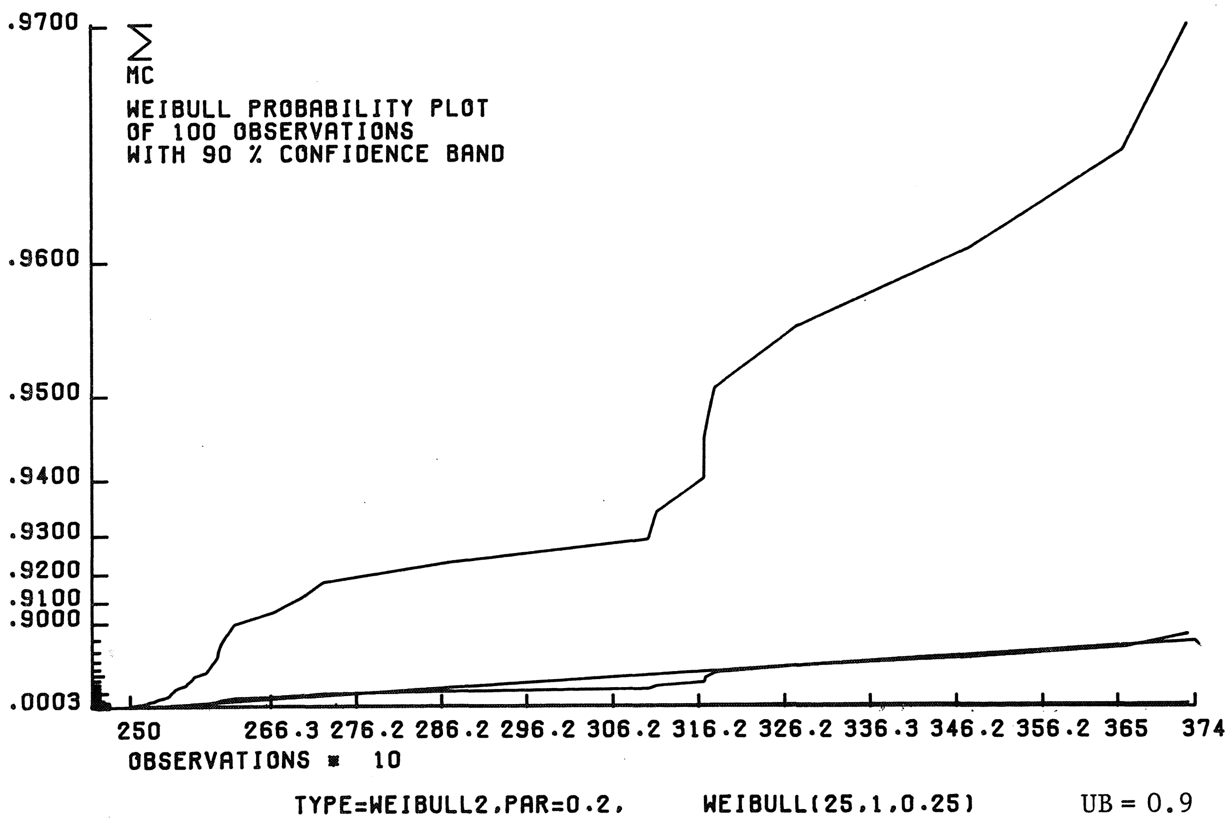


Figure 71.

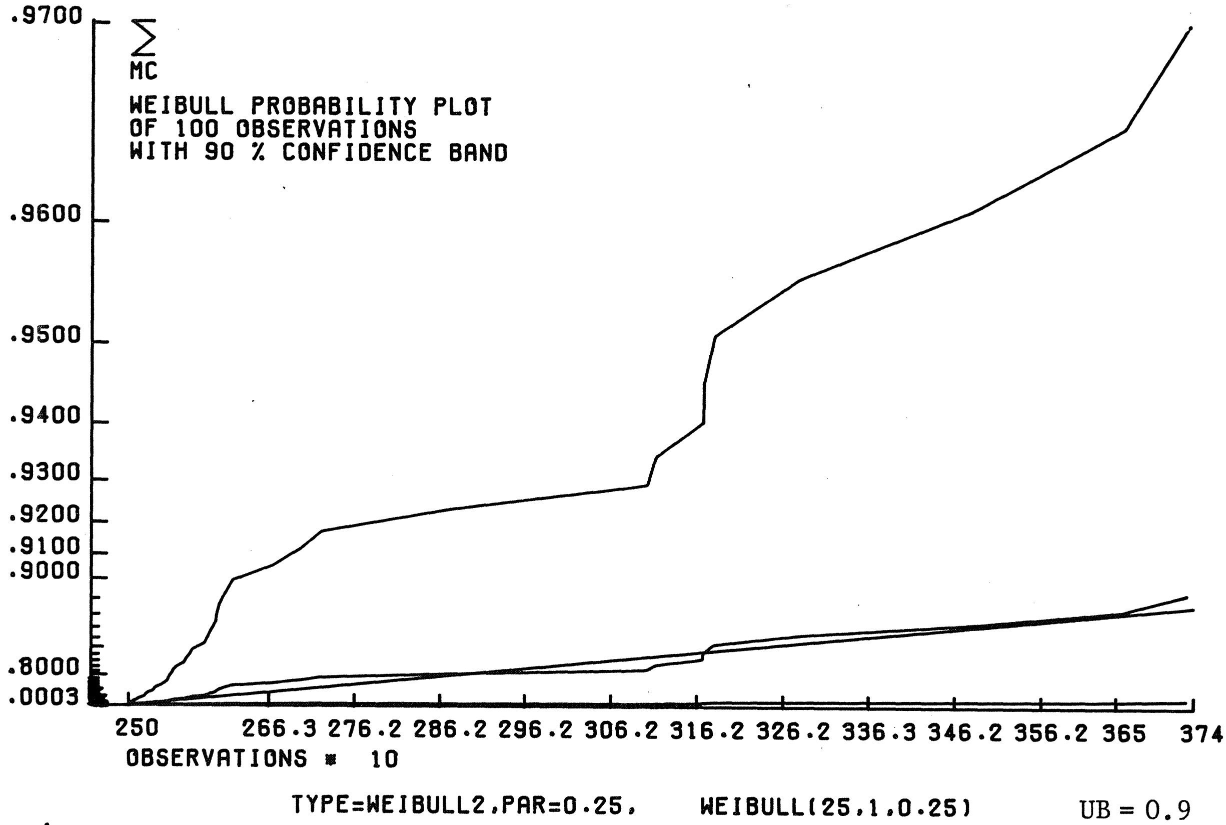


Figure 72.

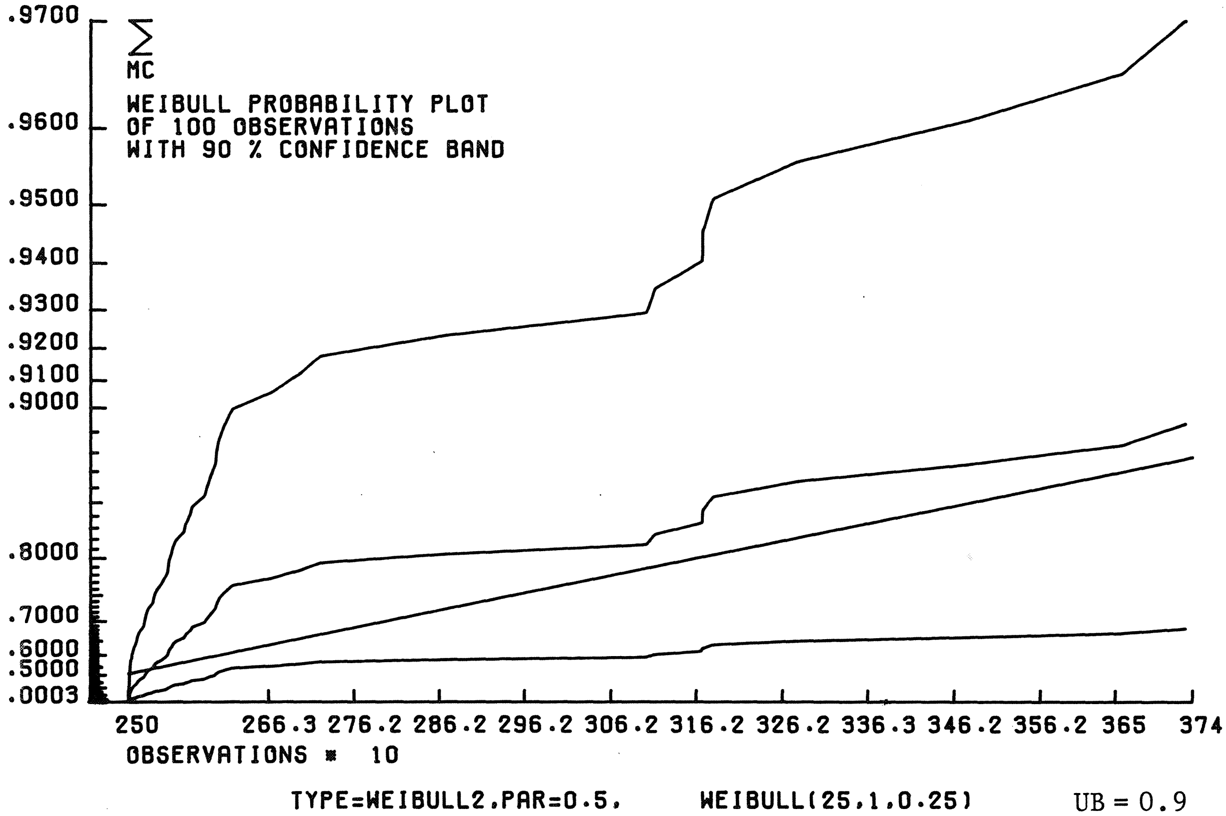


Figure 73.

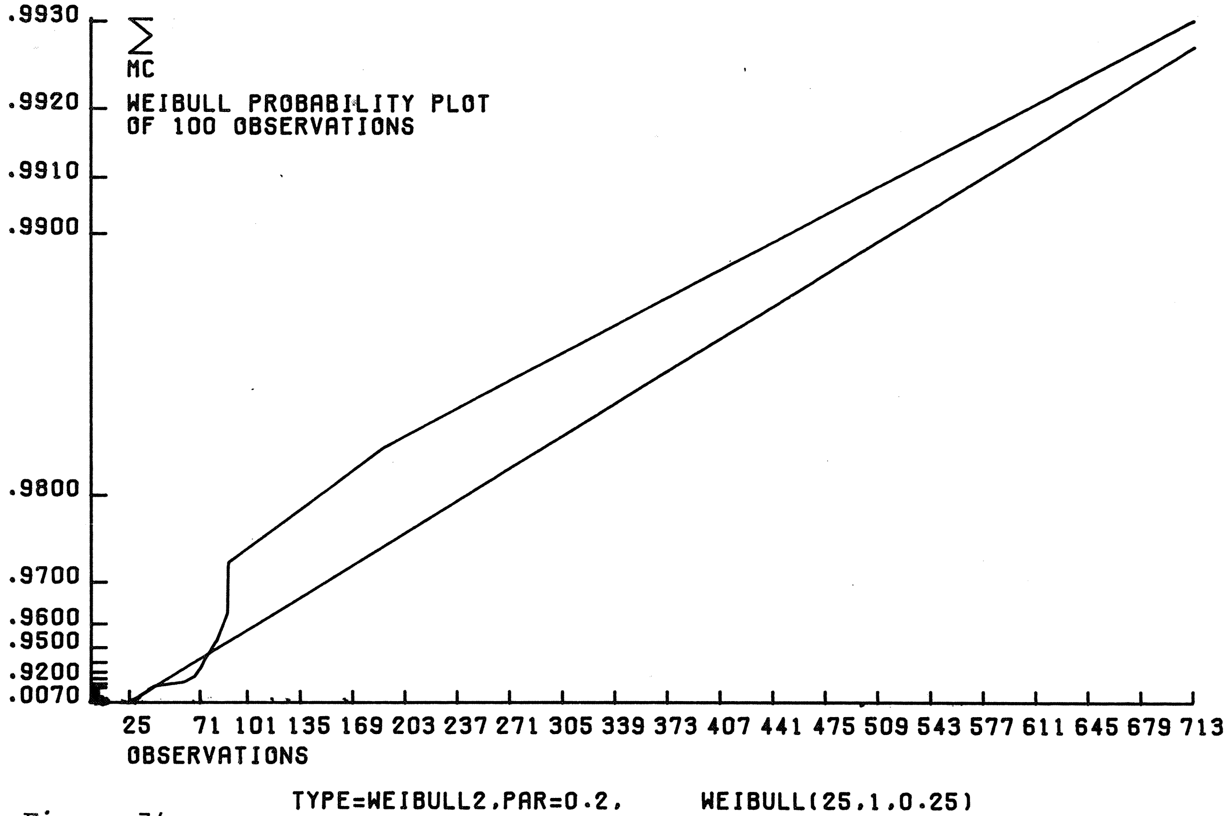


Figure 74.

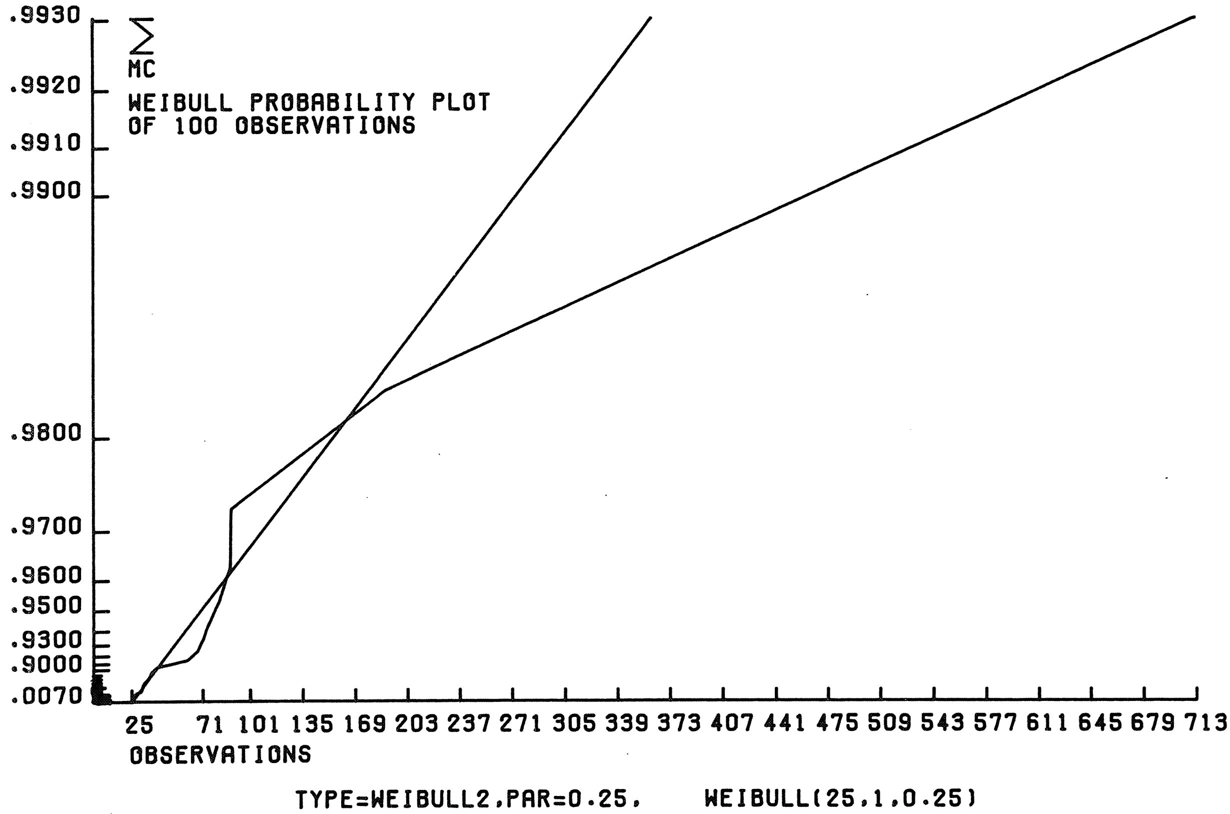


Figure 75.

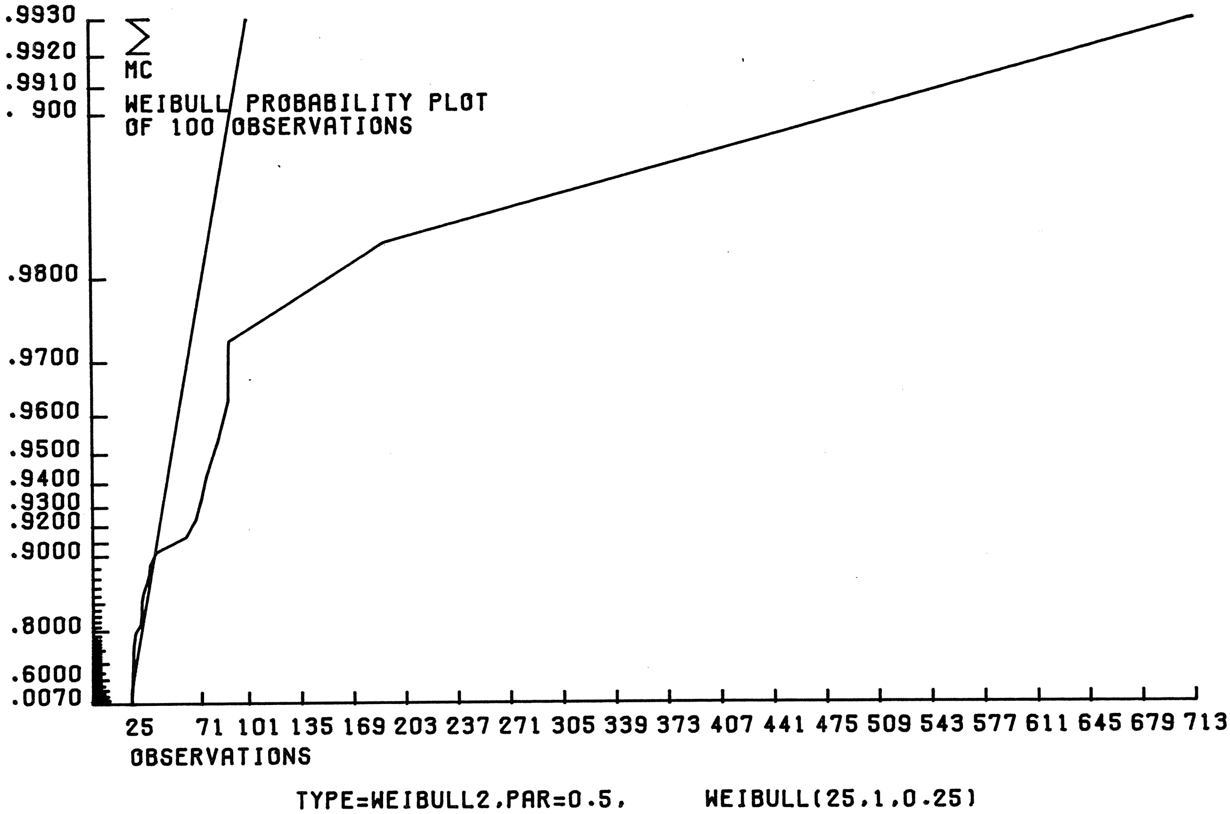


Figure 76.

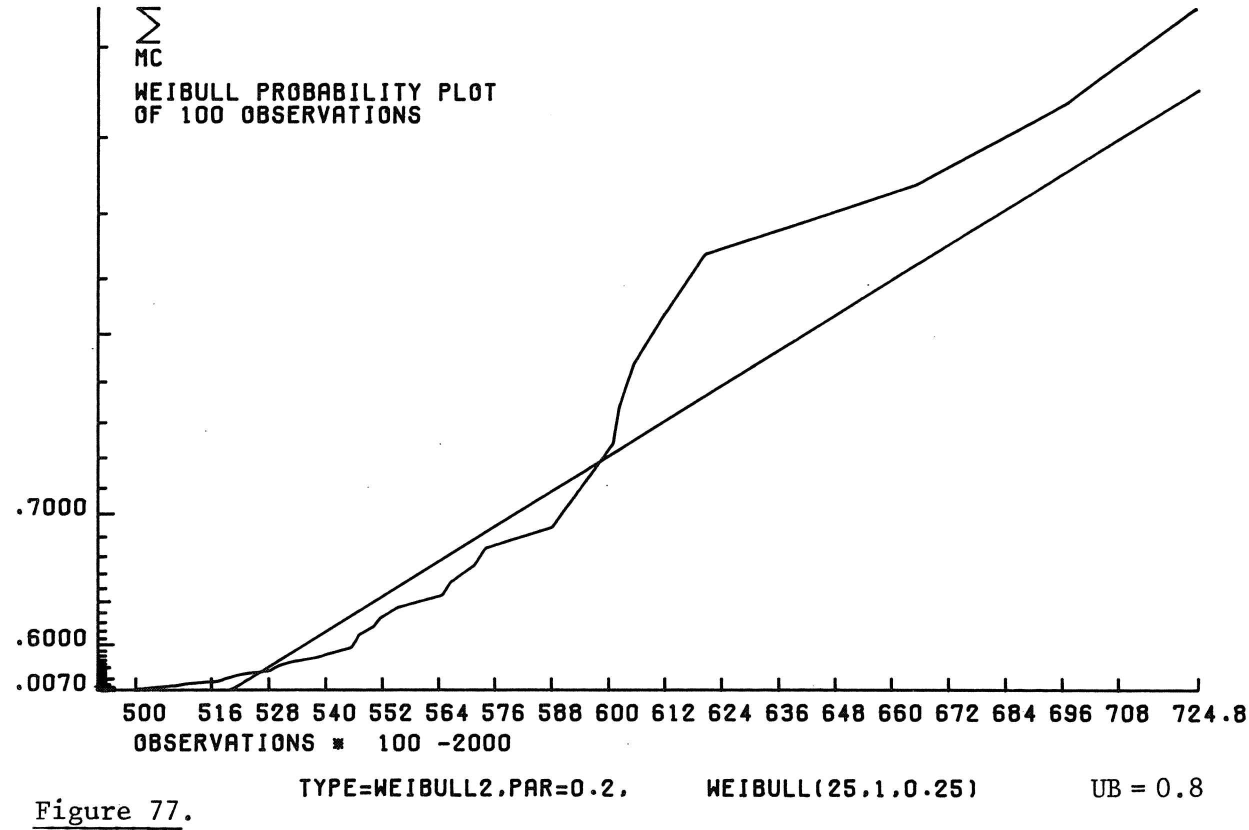
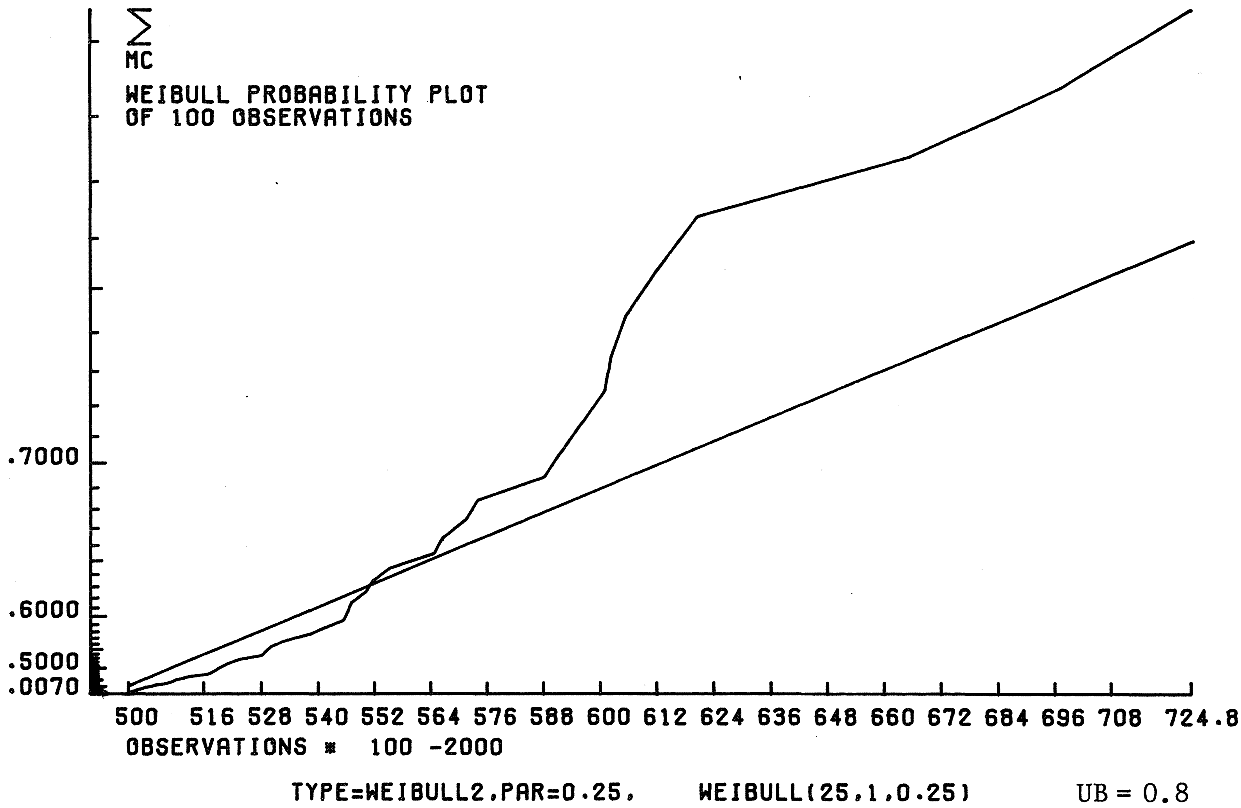
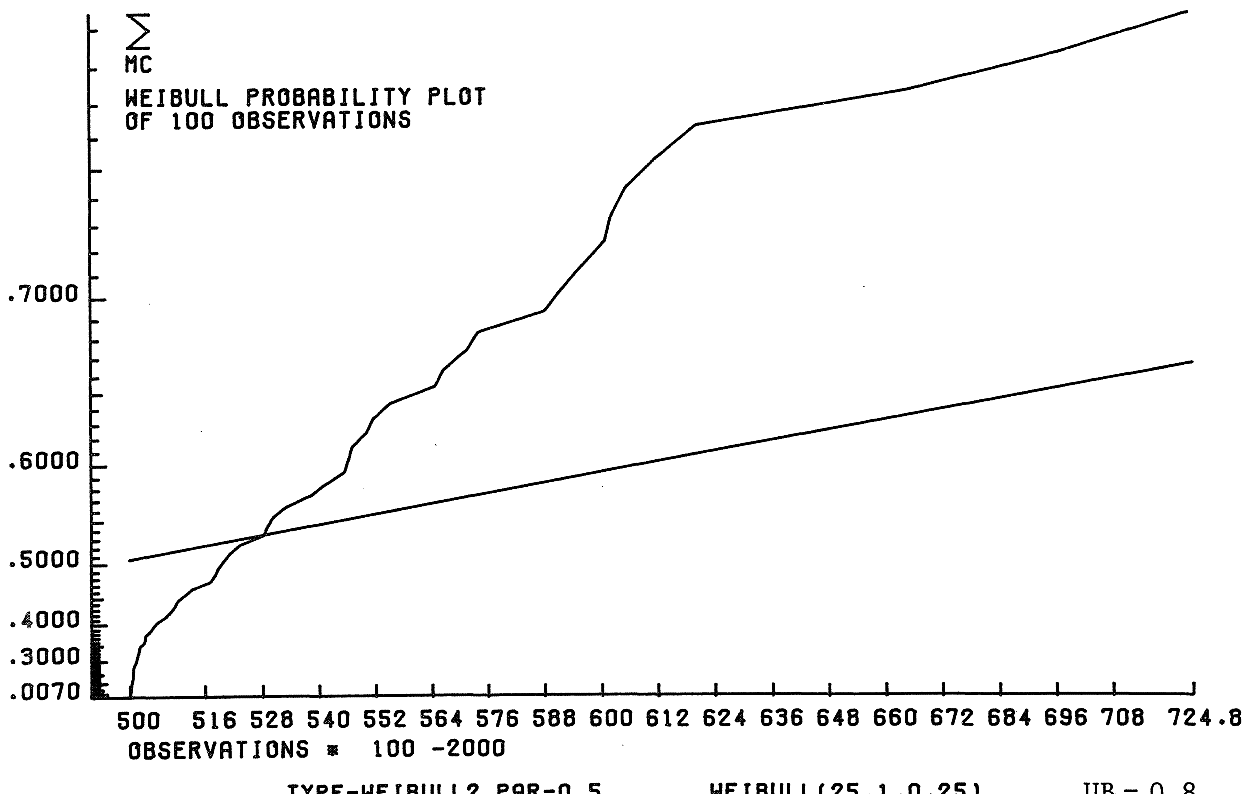


Figure 77.



UB = 0.8

Figure 78.



UB = 0.8

Figure 79.

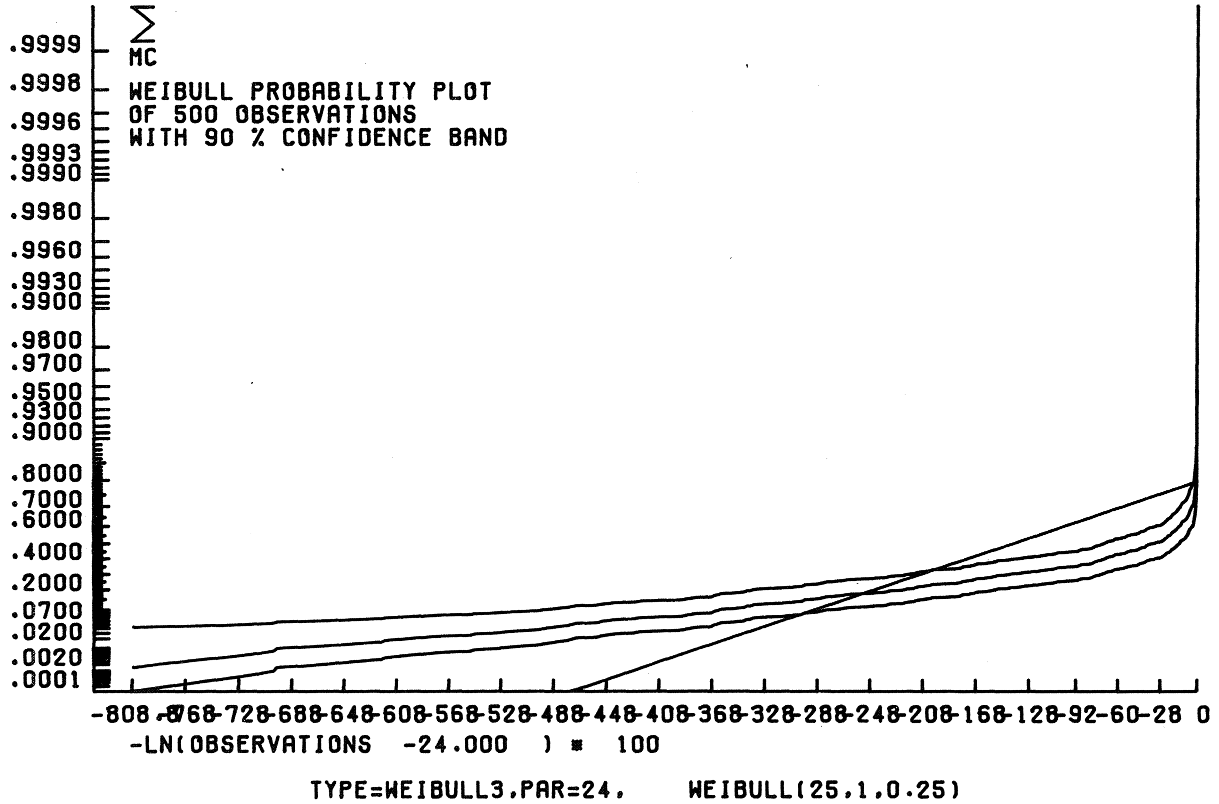


Figure 80.

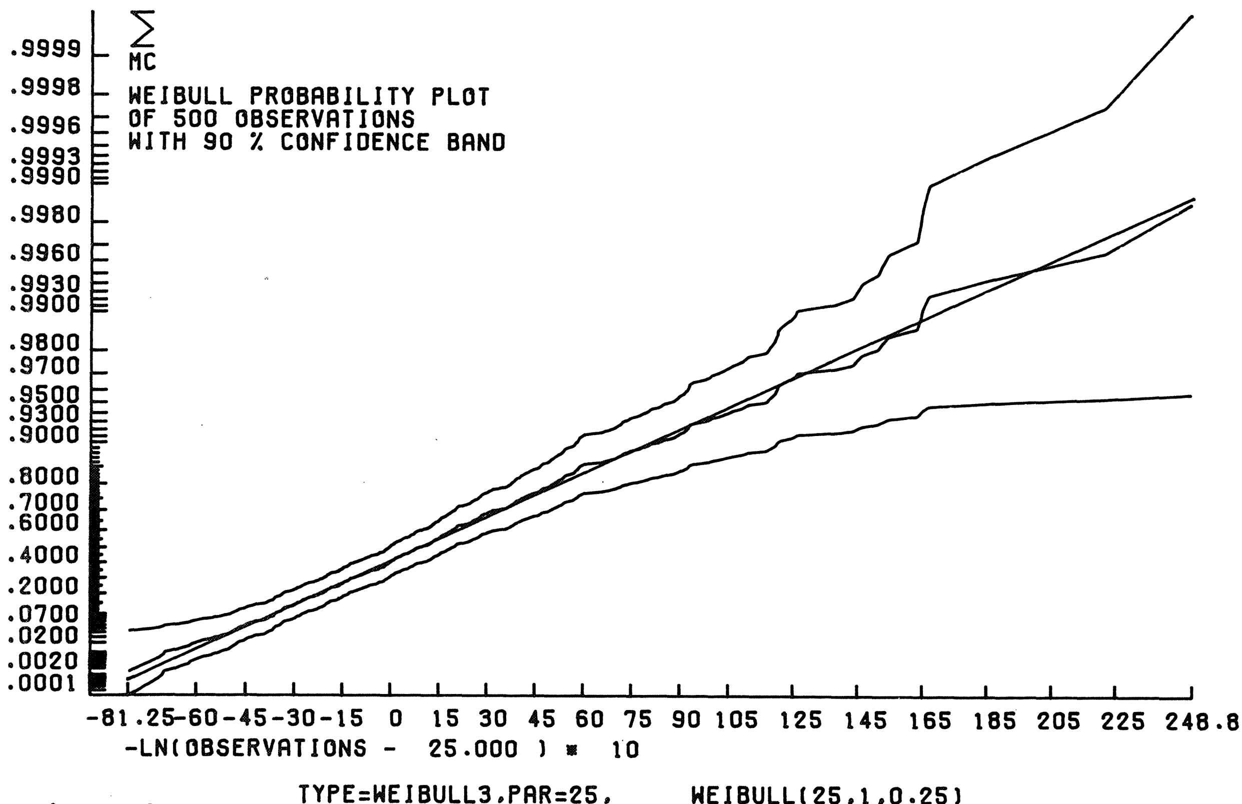


Figure 81.

5.2.3.2. Uniform, normal and exponential probability plots of samples from different distributions

Sample	Distribution
1	uniform (0,1)
2	normal (0,1)
3	exponential (0,1)
4	Laplace (0,1)
5	Gumbel (0,1)
6	Cauchy (0,1)
7	Weibull (0,1,0.25)
8	Weibull (0,1,3.6)

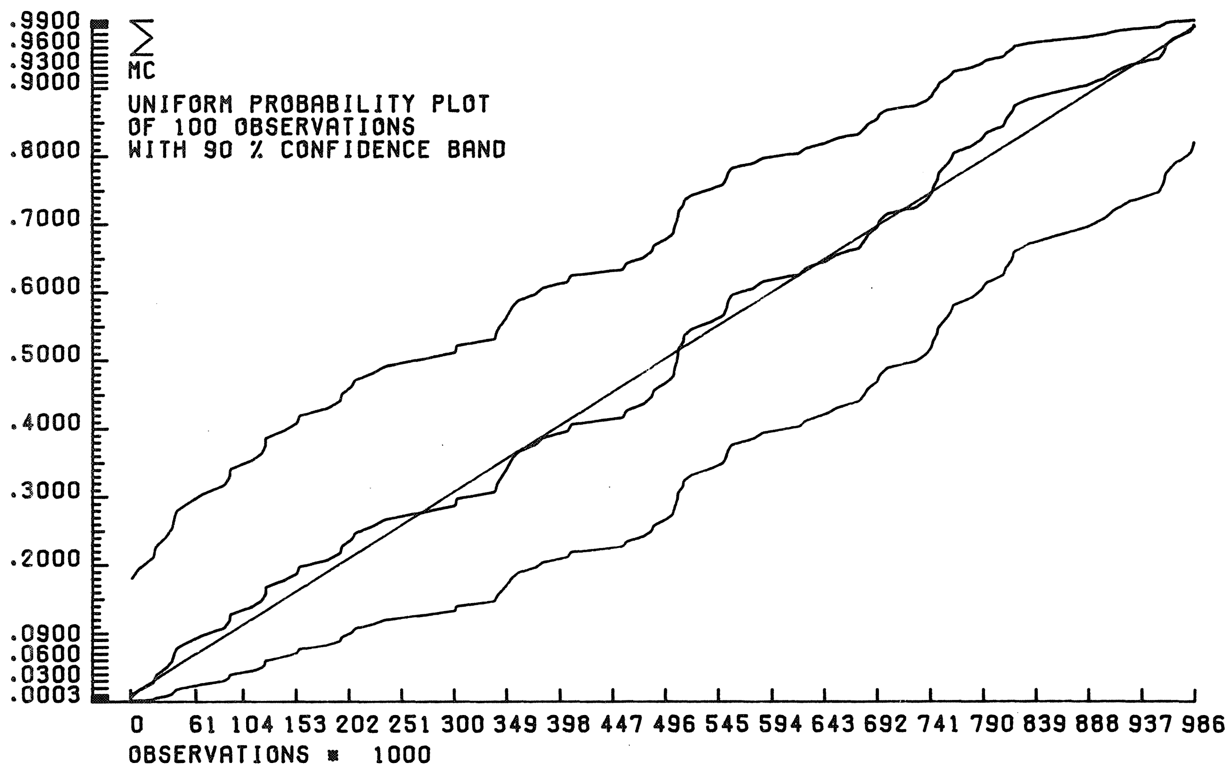


Figure 82. Sample 1

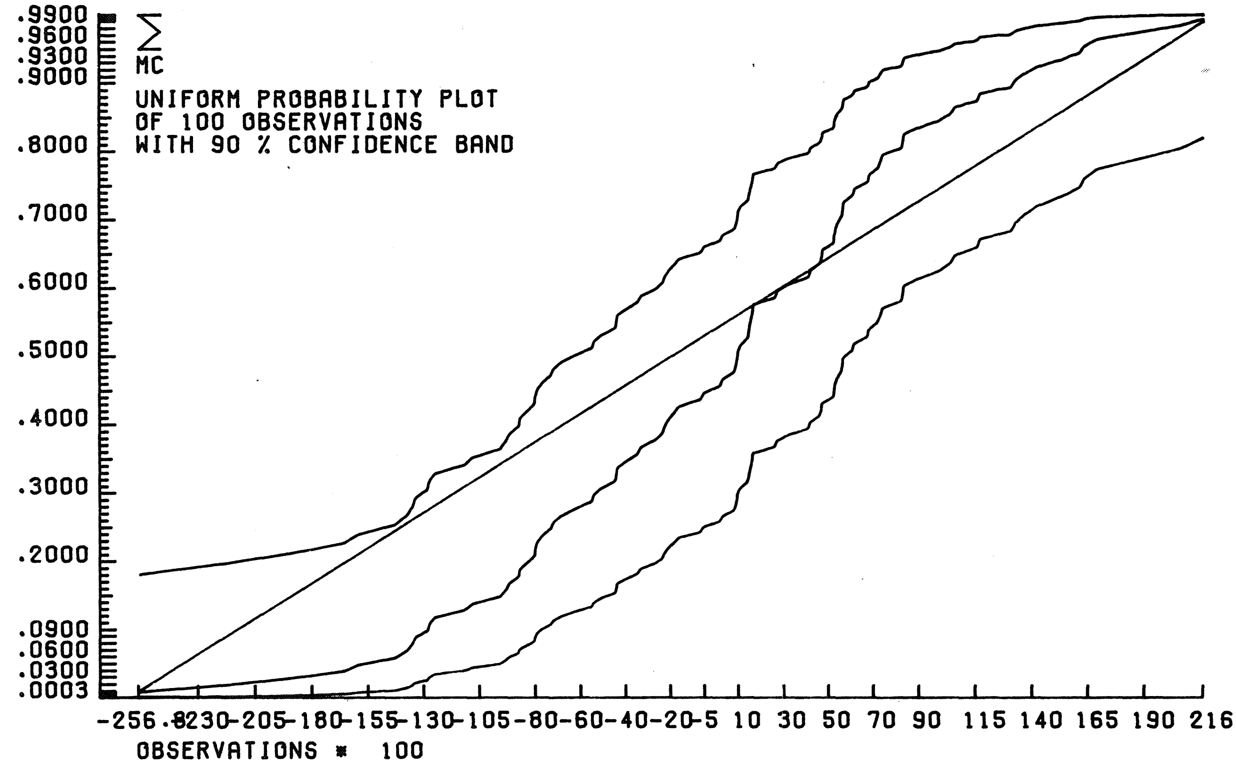


Figure 83. Sample 2

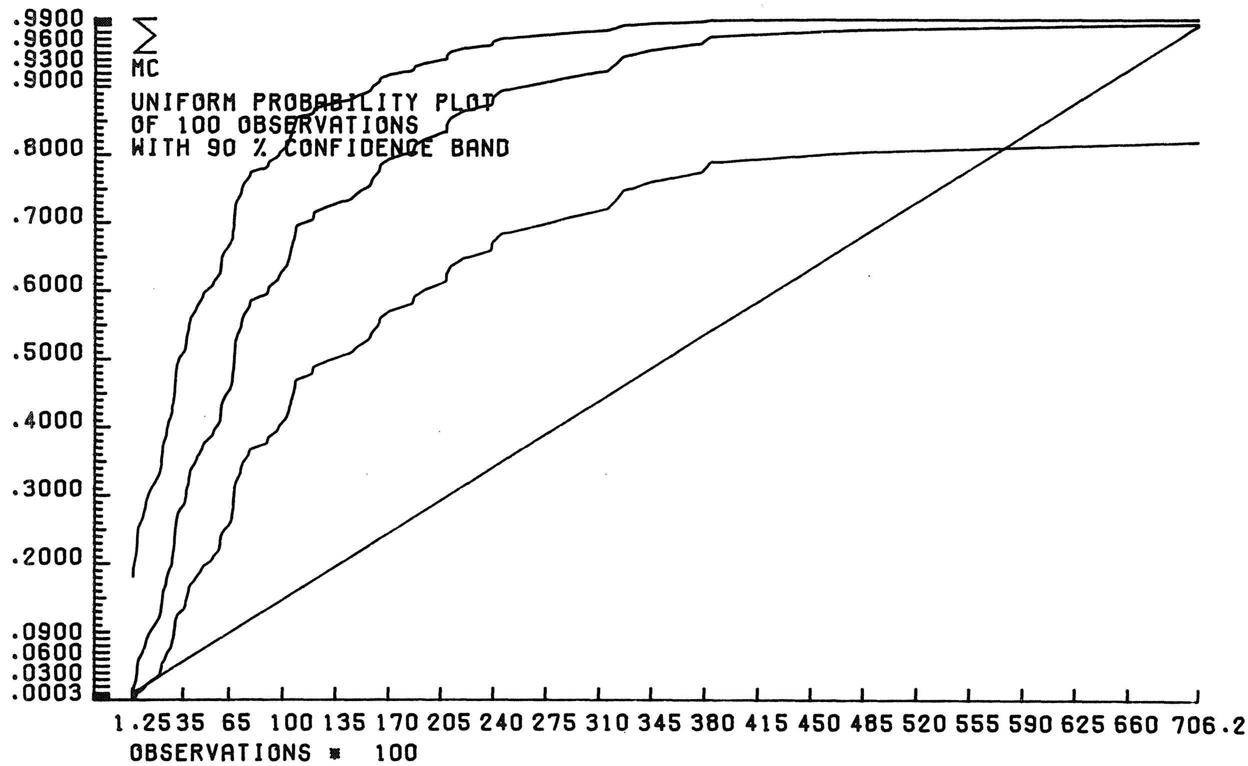


Figure 84. Sample 3

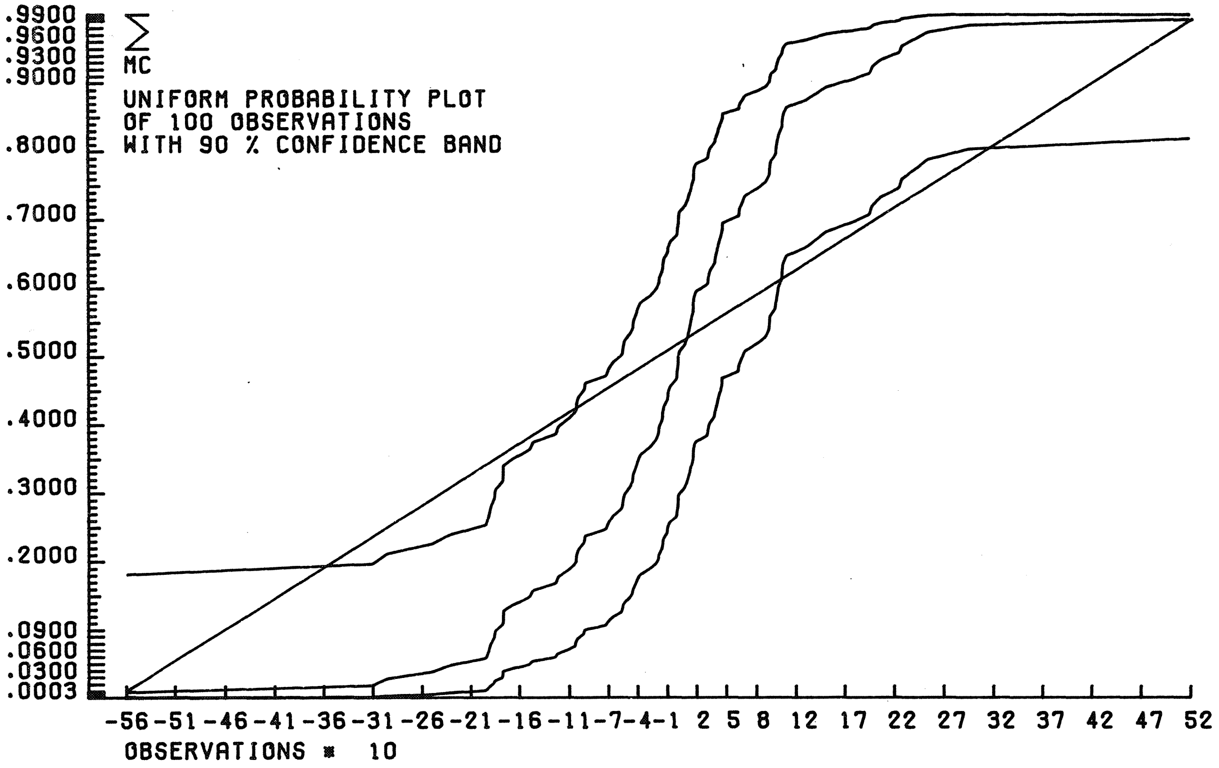


Figure 85. Sample 4

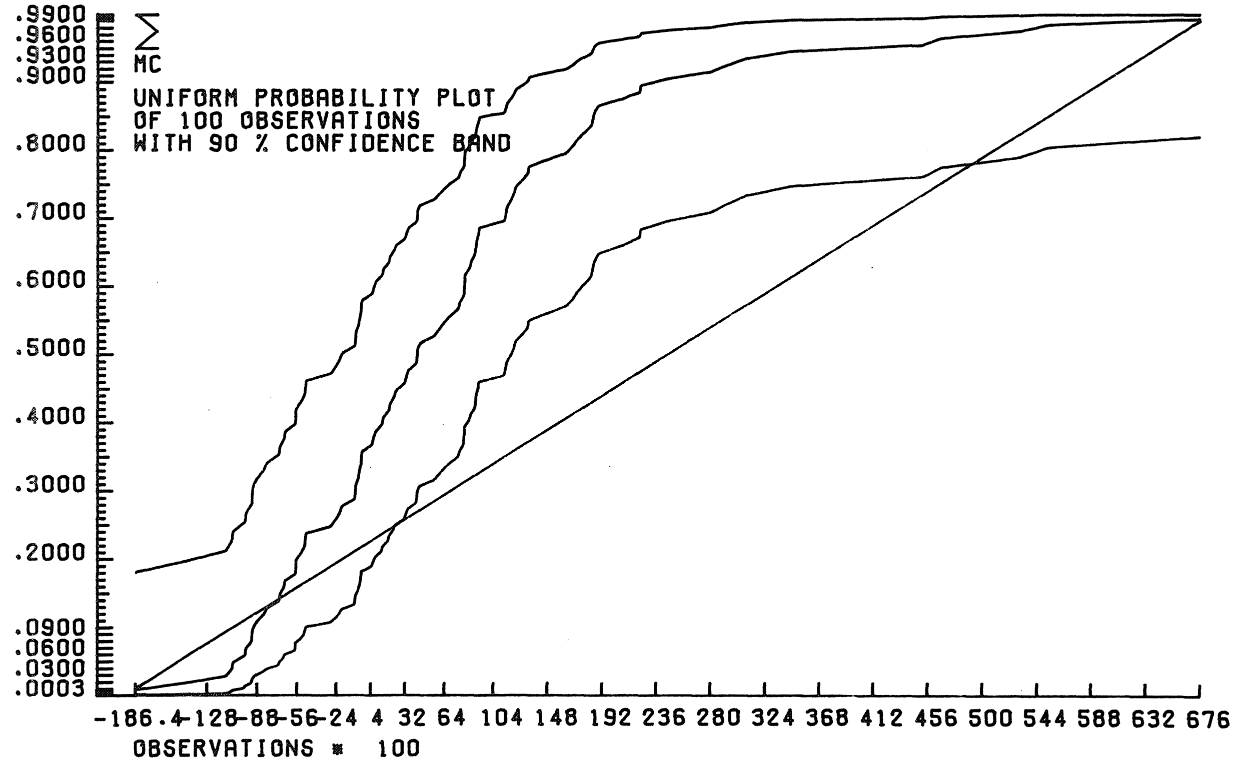


Figure 86. Sample 5

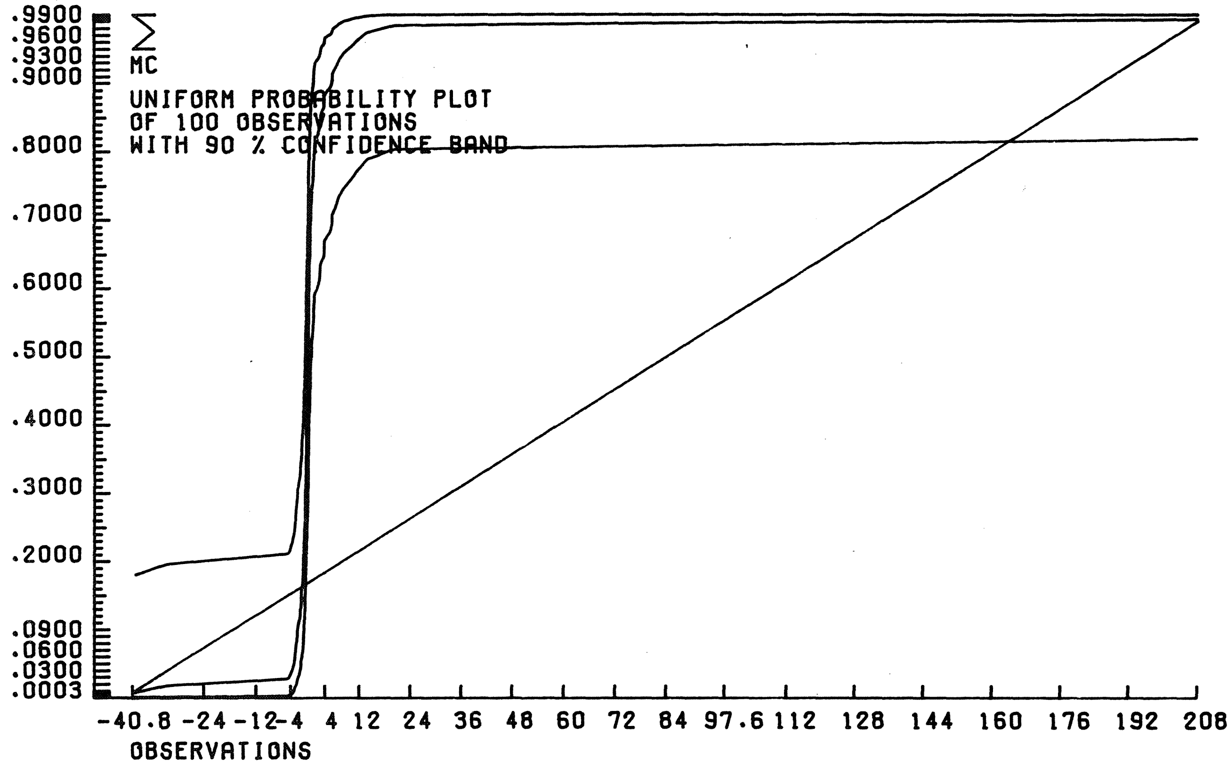


Figure 87. Sample 6

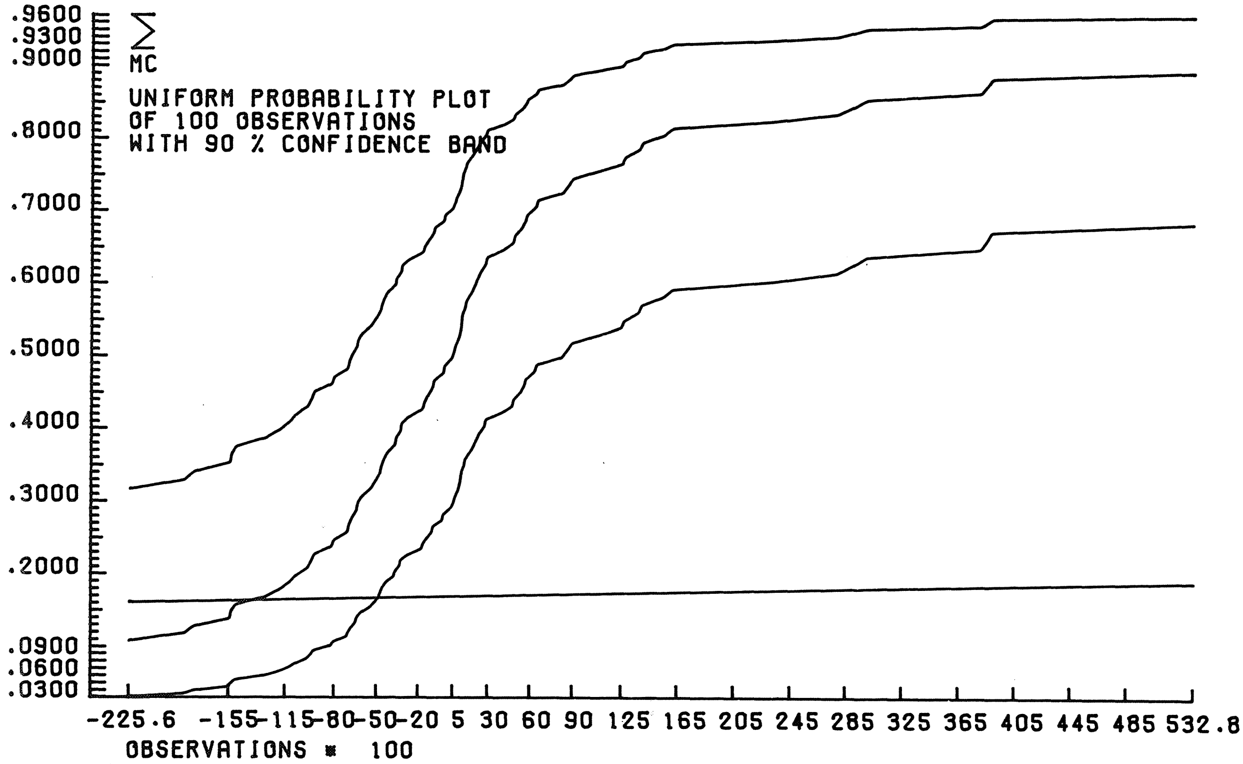


Figure 88. Sample 6, enlargement (LB = 0.1, UB = 0.9)

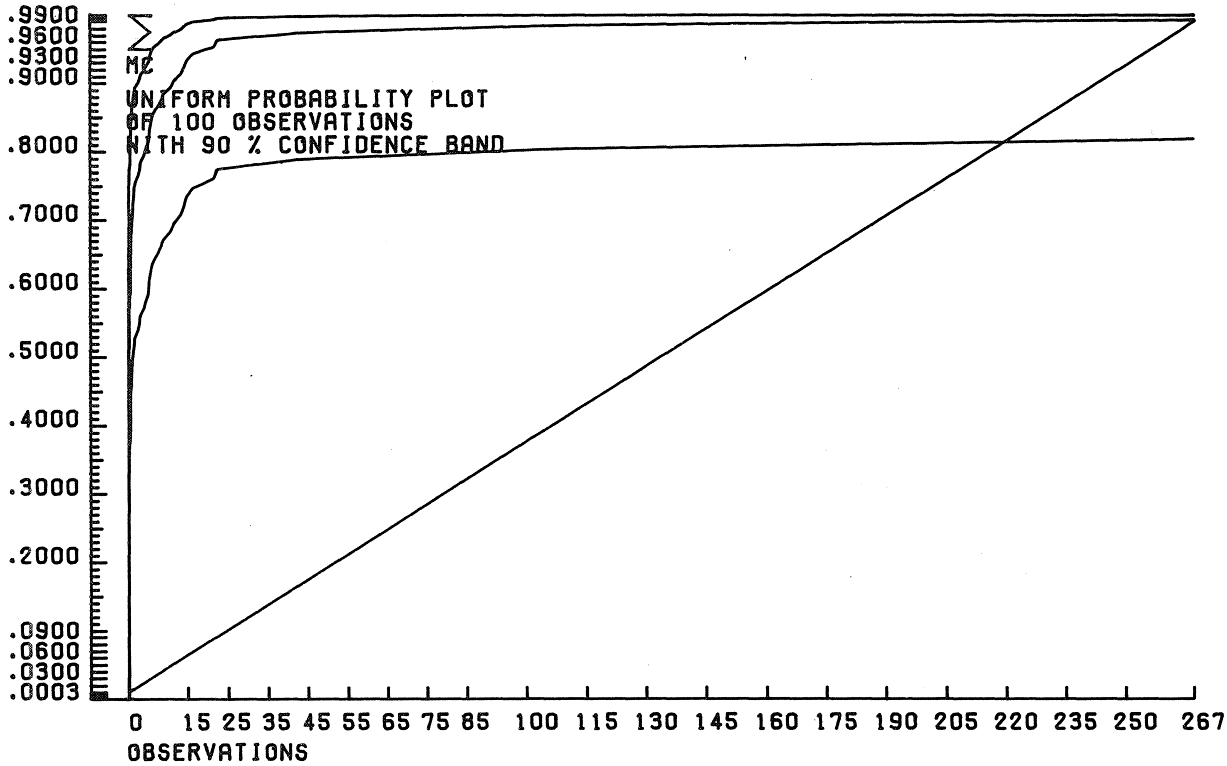
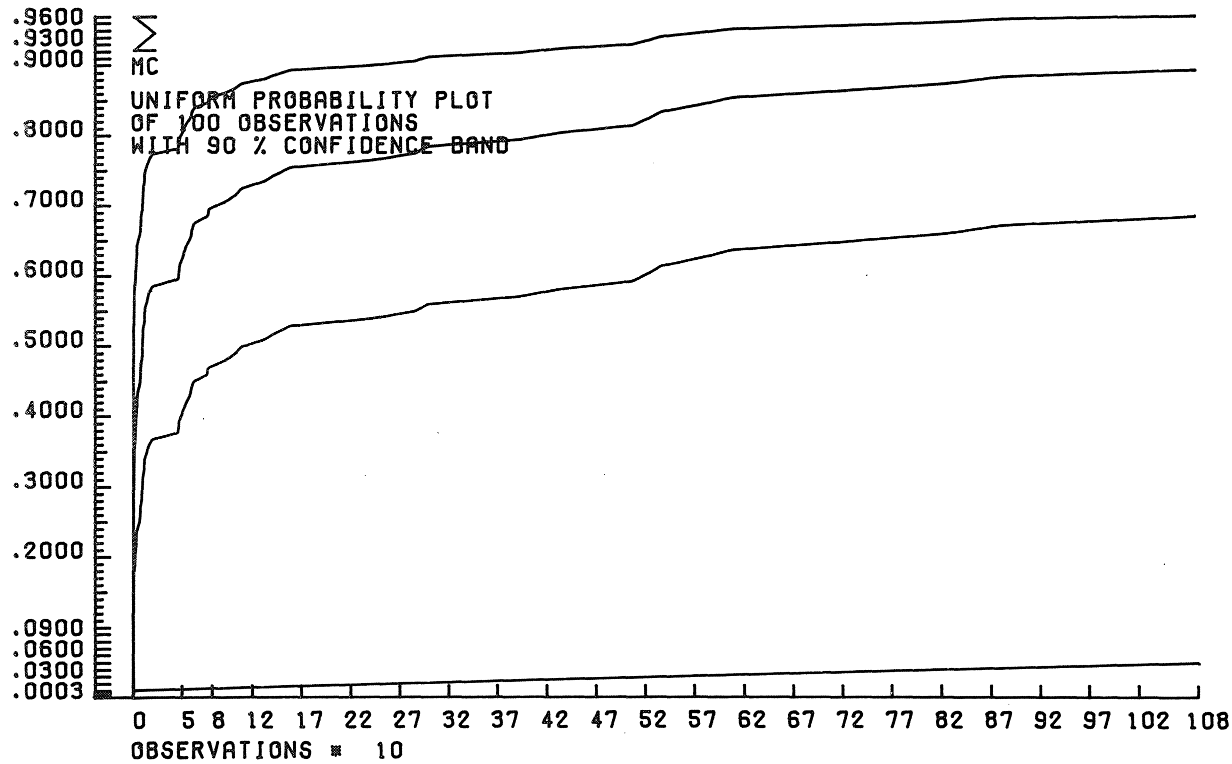


Figure 89. Sample 7 WEIBULL(0,1,0.25)



WEIBULL(0,1,0.25)

Figure 90. Sample 7, enlargement (UB = 0.9)

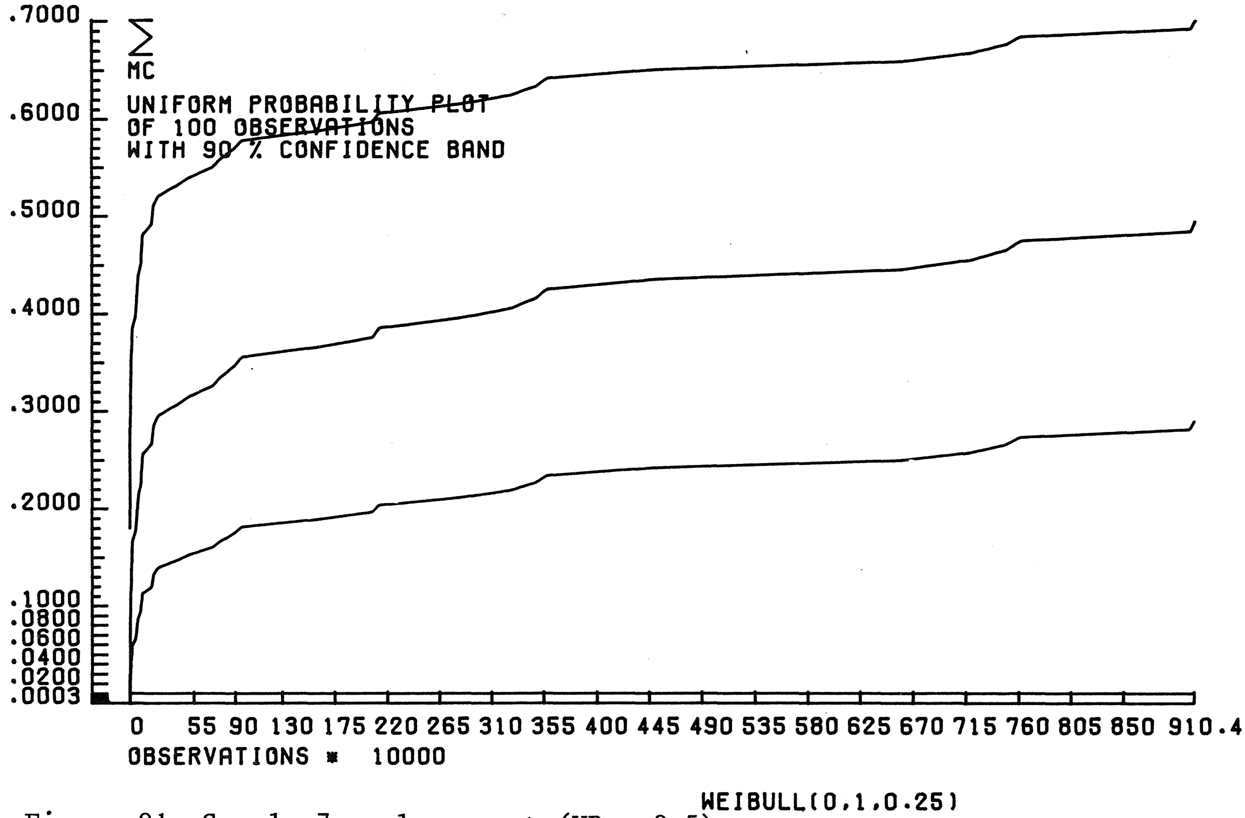


Figure 91. Sample 7, enlargement (UB = 0.5)

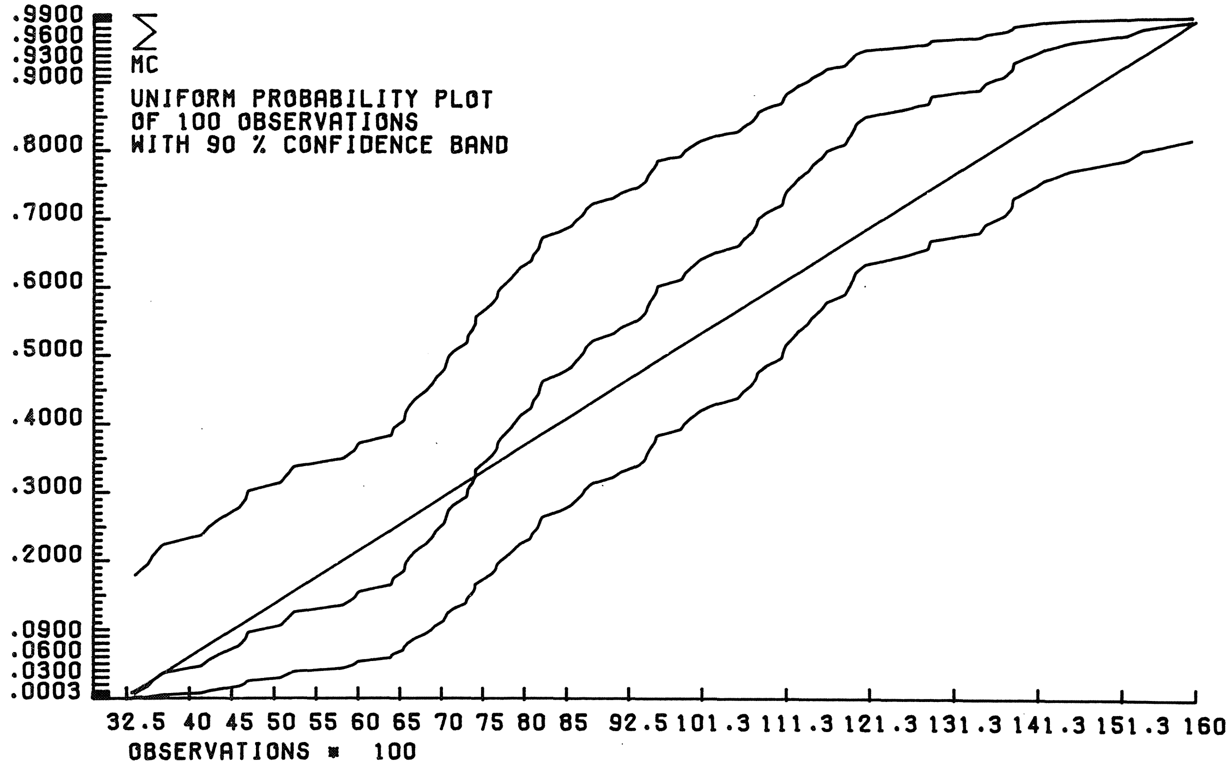


Figure 92. Sample 8

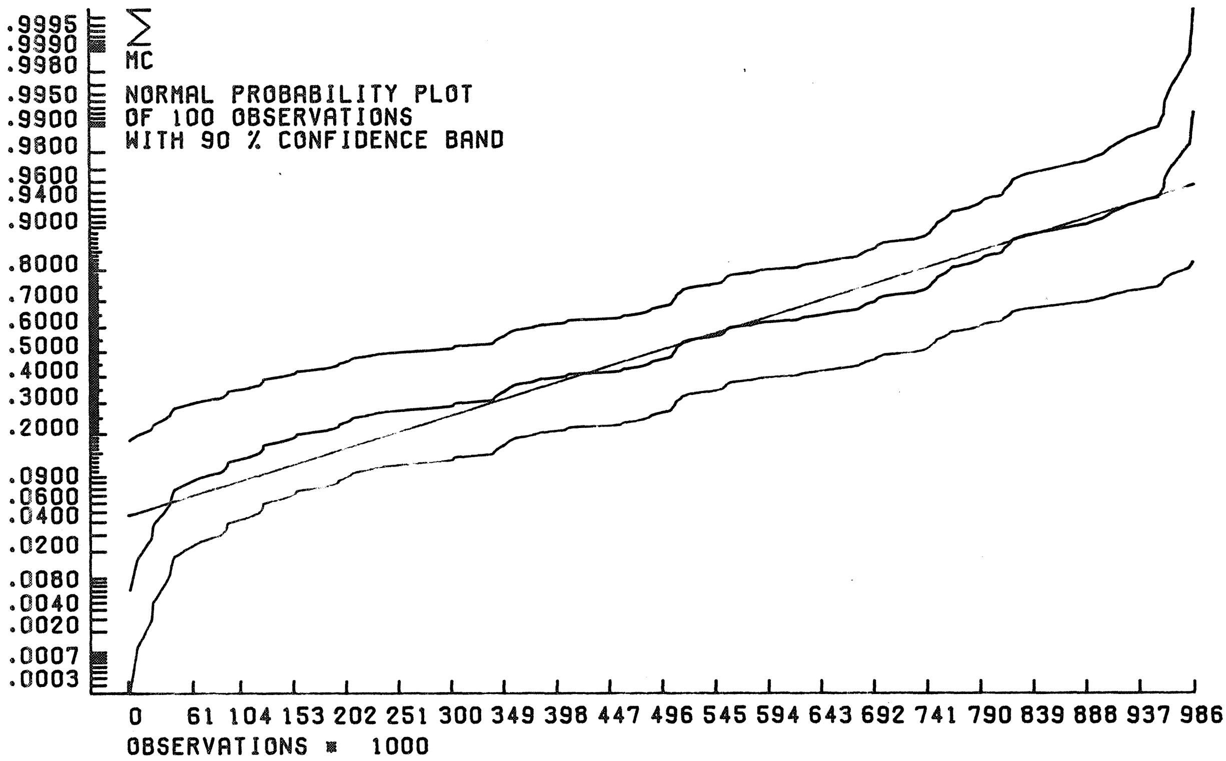


Figure 93. Sample 1

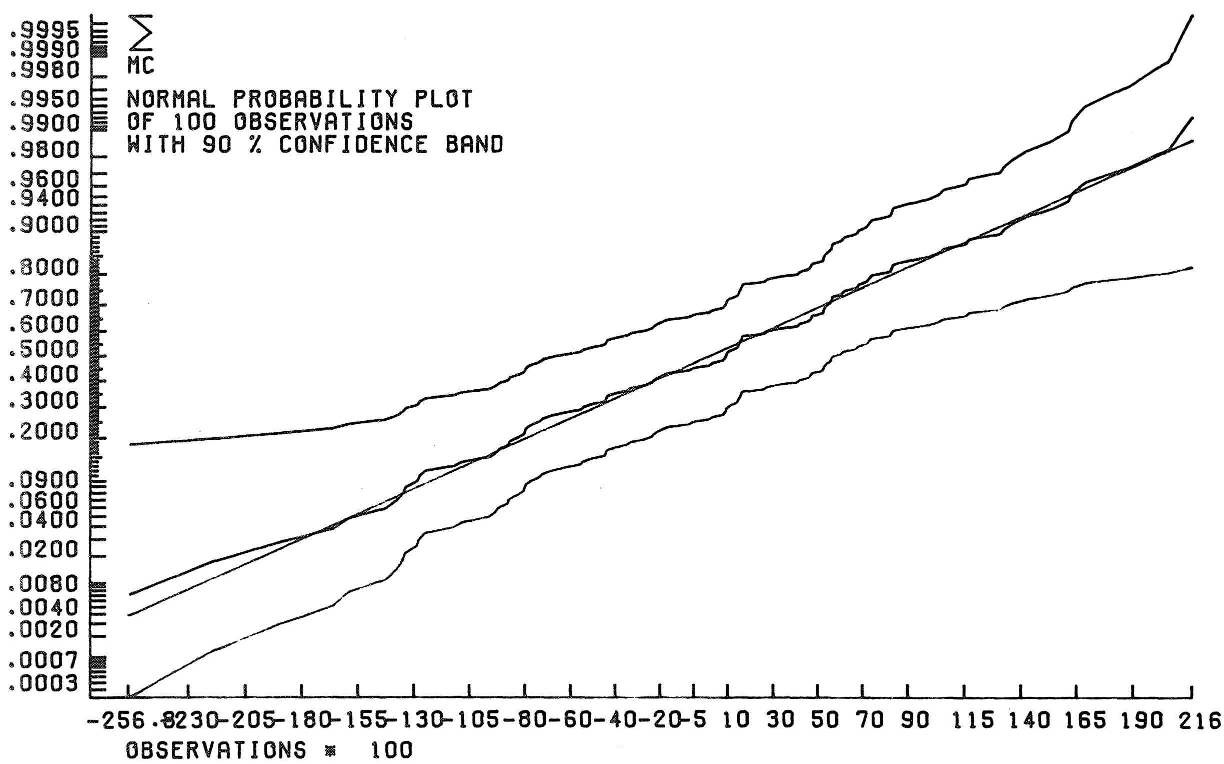


Figure 94. Sample 2

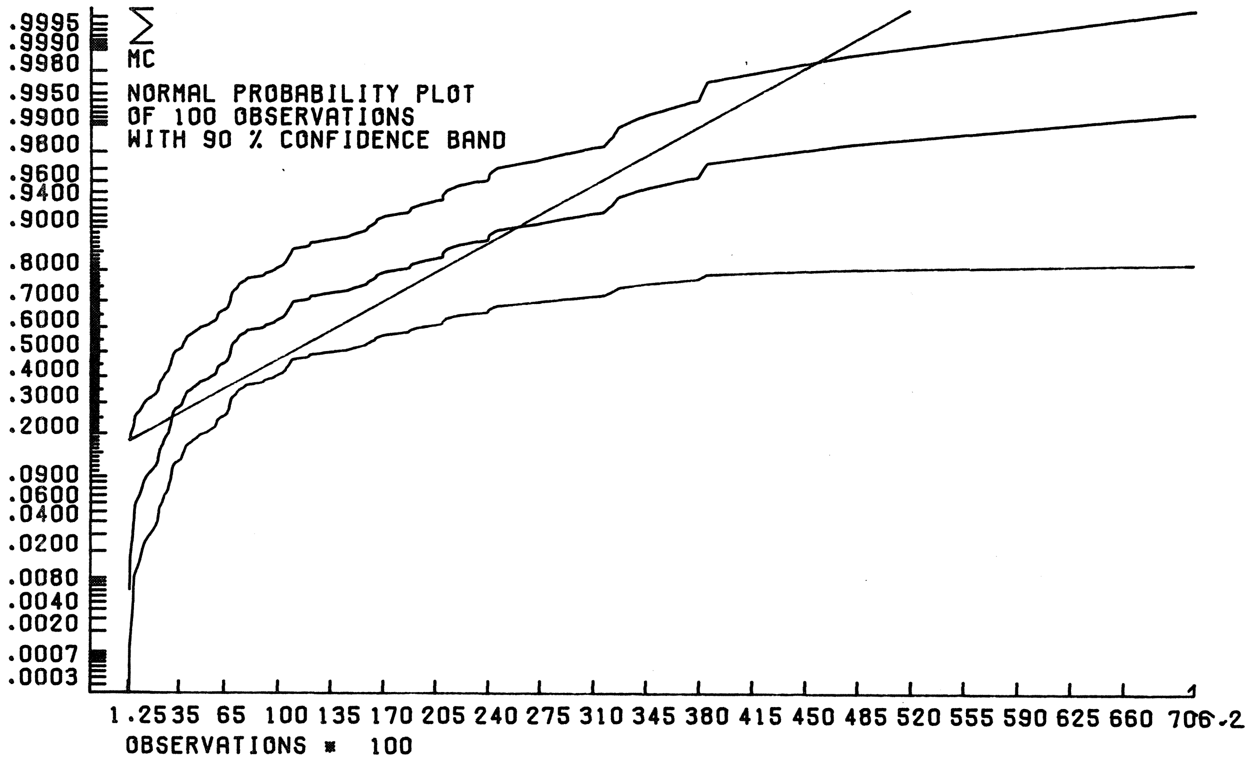


Figure 95. Sample 3

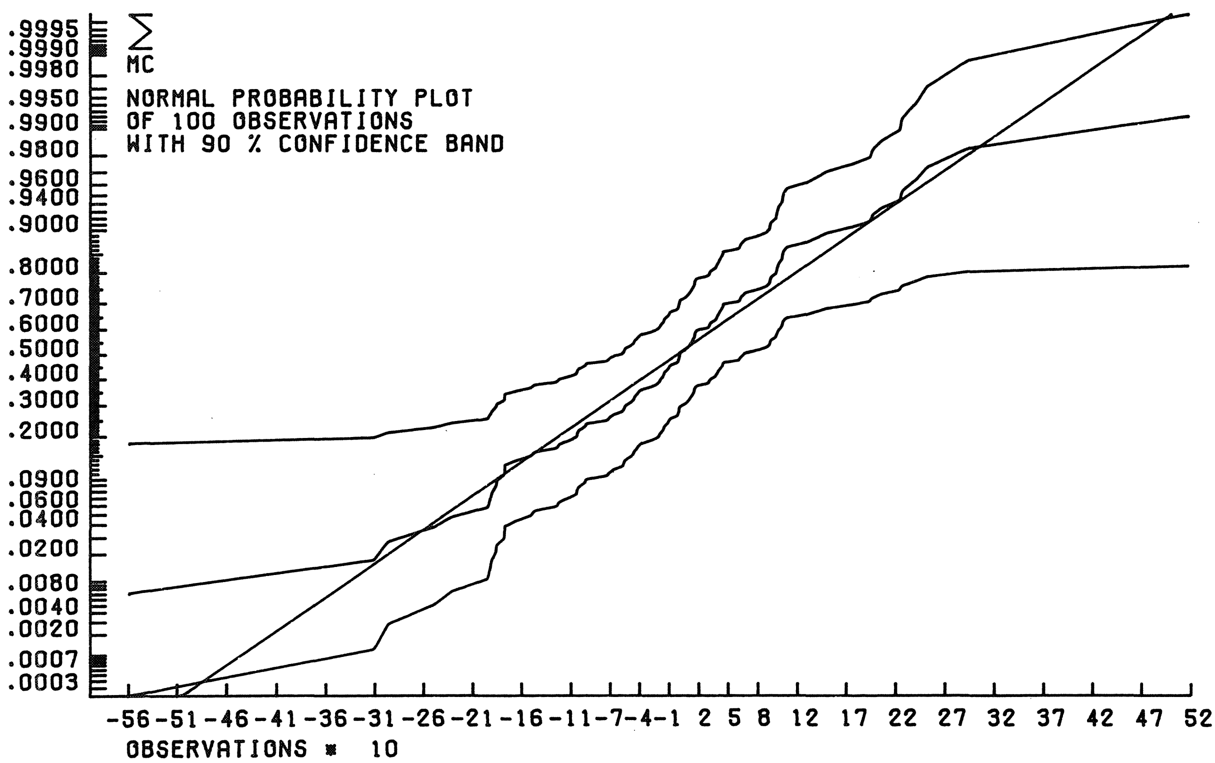


Figure 96. Sample 4

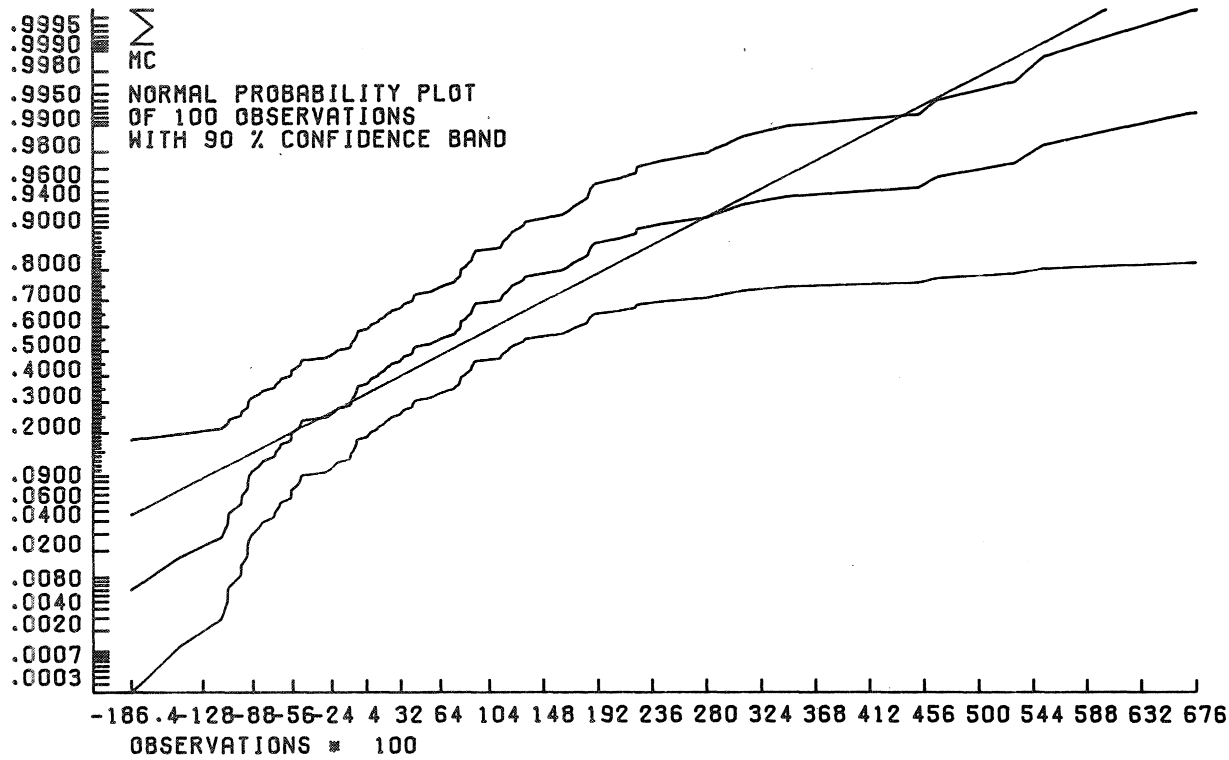


Figure 97. Sample 5

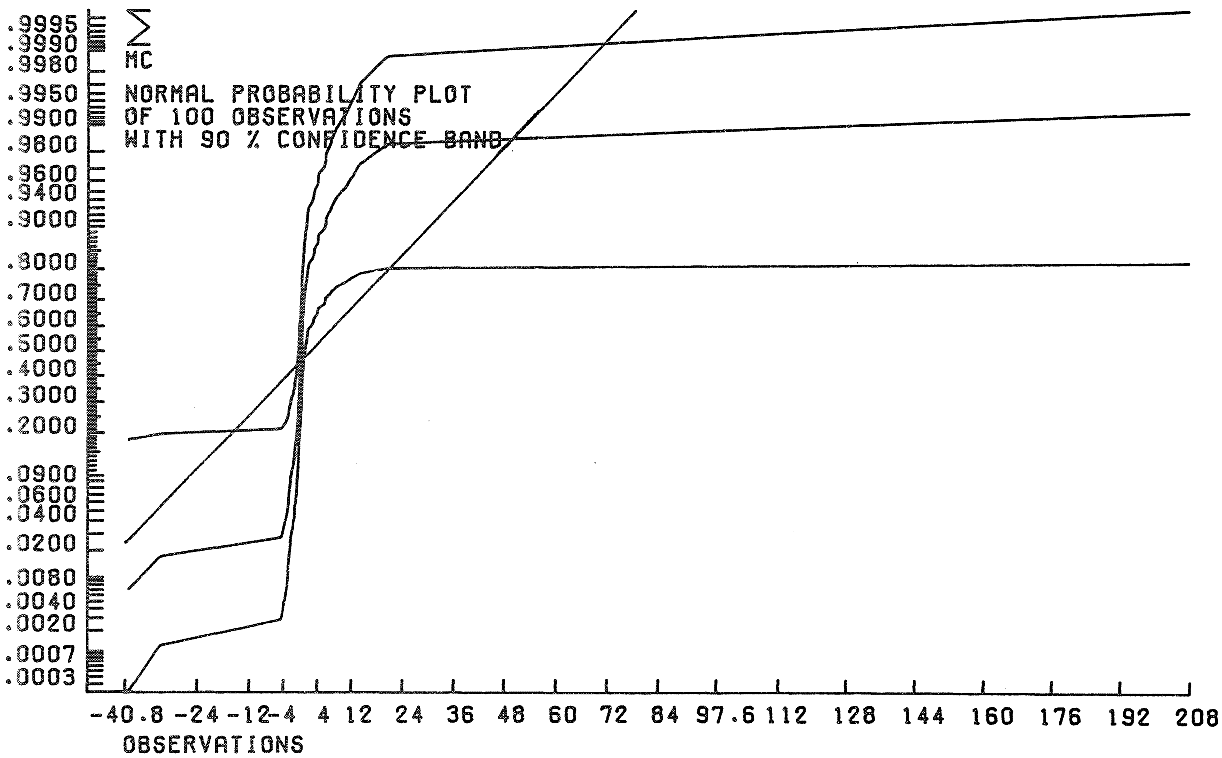


Figure 98. Sample 6

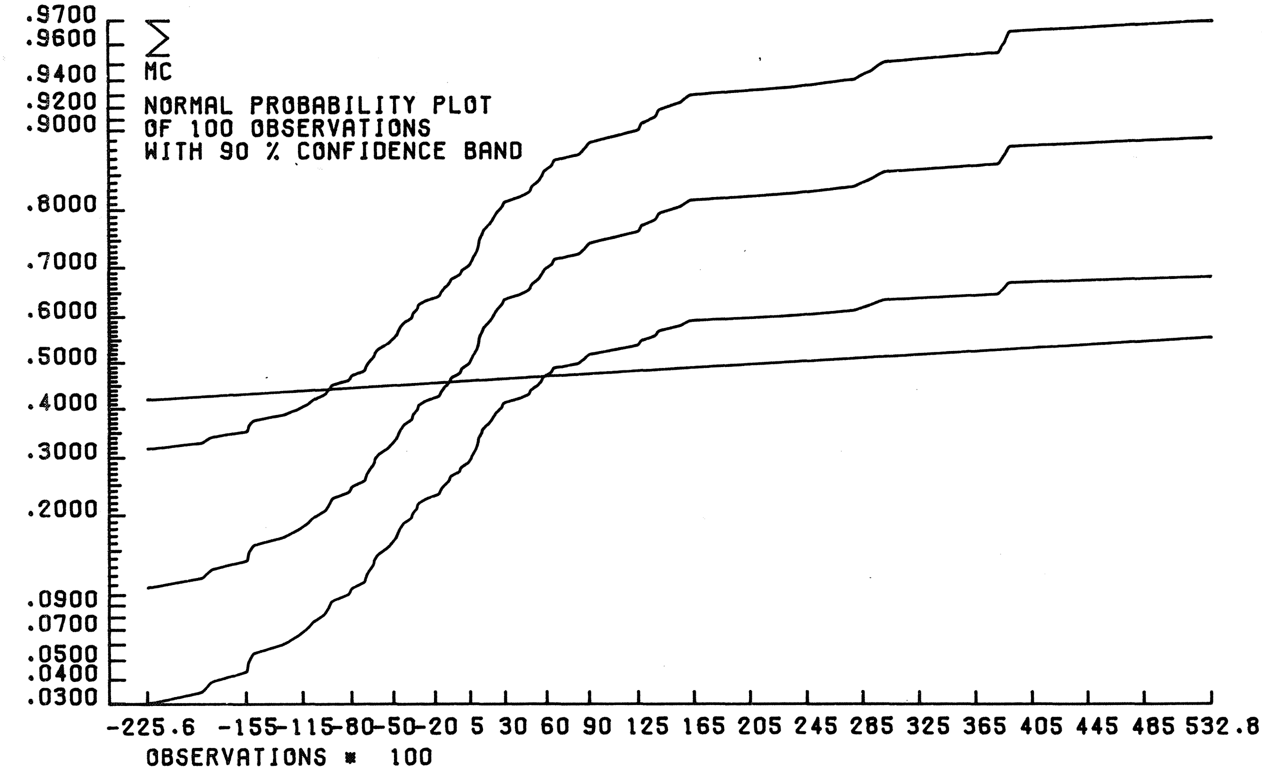


Figure 99. Sample 6, enlargement (LB = 0.1, UB = 0.9)

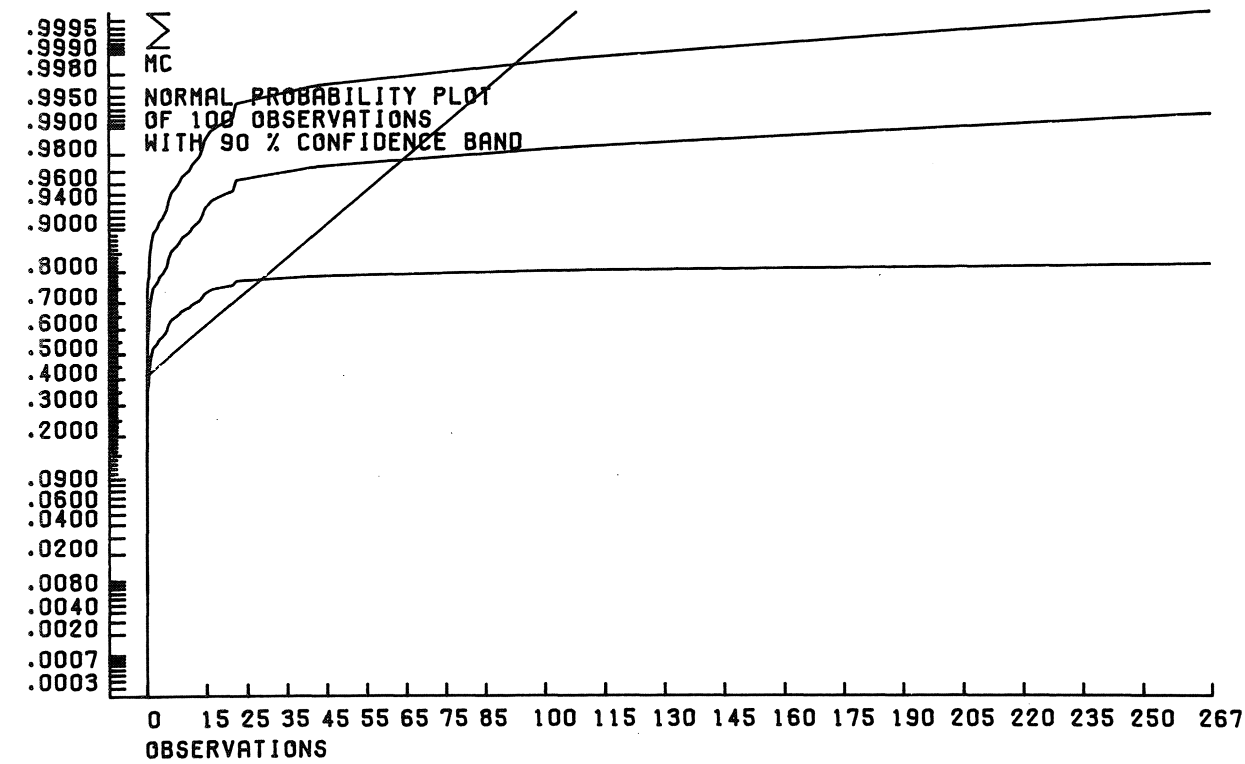


Figure 100. Sample 7

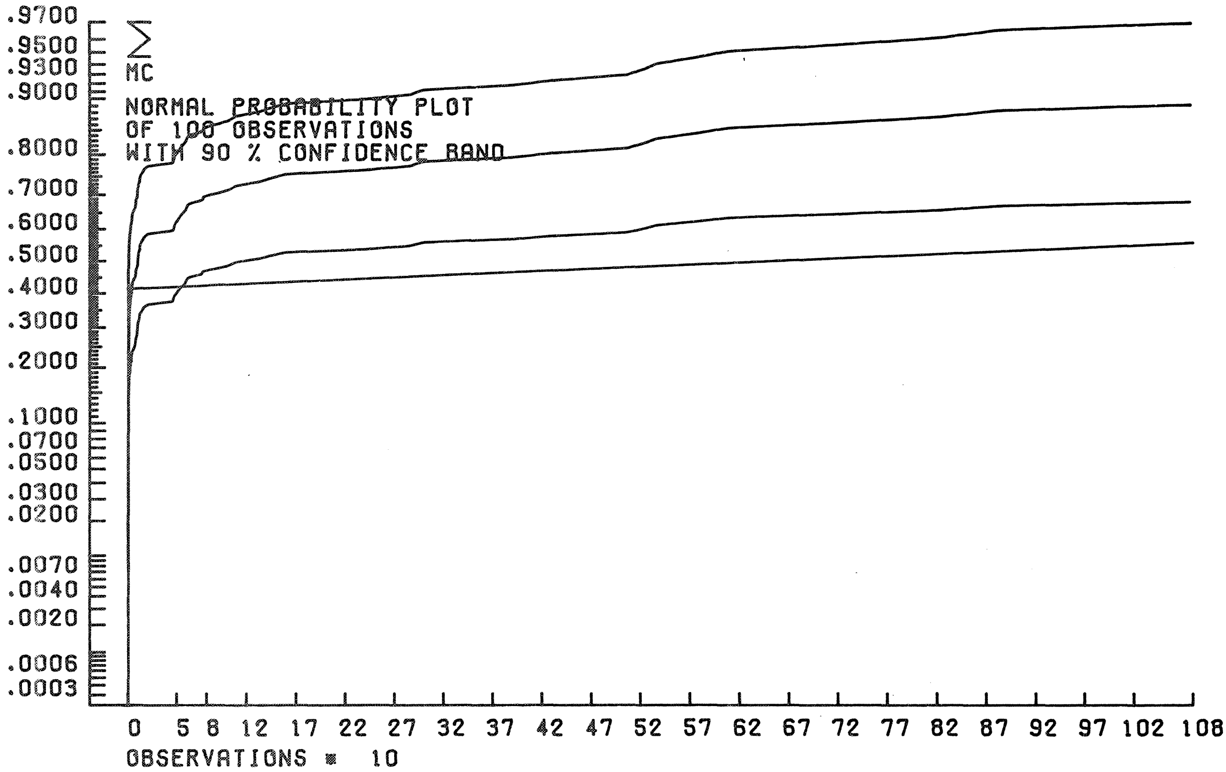


Figure 101. Sample 7, enlargement (UB = 0.9)

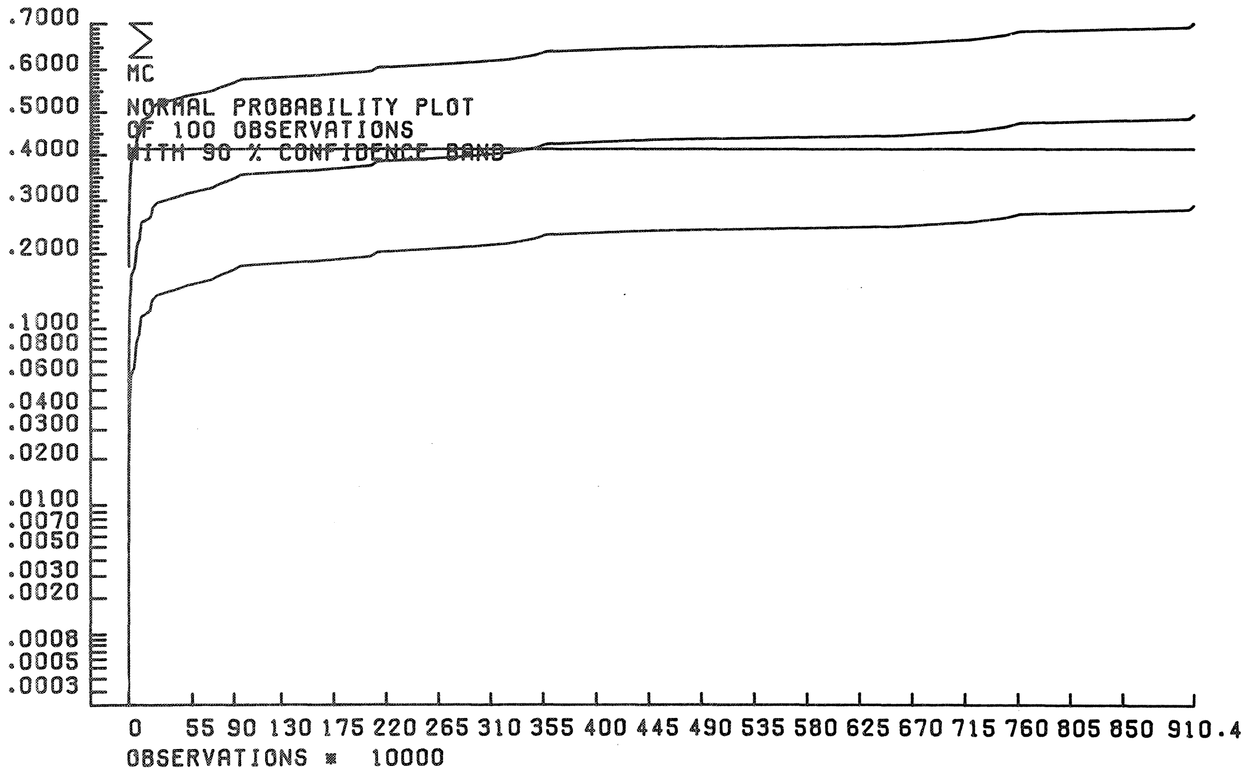


Figure 102. Sample 7, enlargement (UB = 0.5)

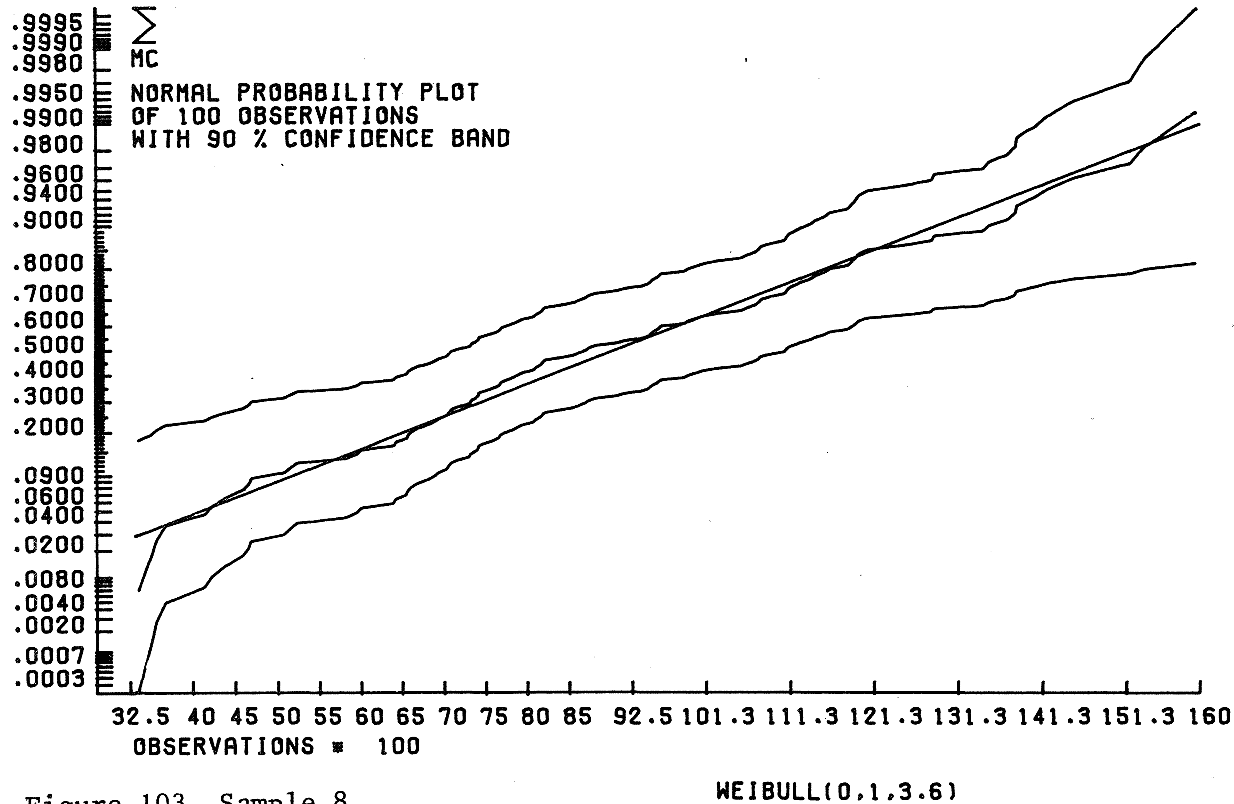


Figure 103. Sample 8

The skewness of the Weibull (0,1,3.6) distribution is approximately 0 and the kurtosis is 2.72 (JOHNSON & KOTZ [8]), explaining partially the close resemblance to a normal distribution.

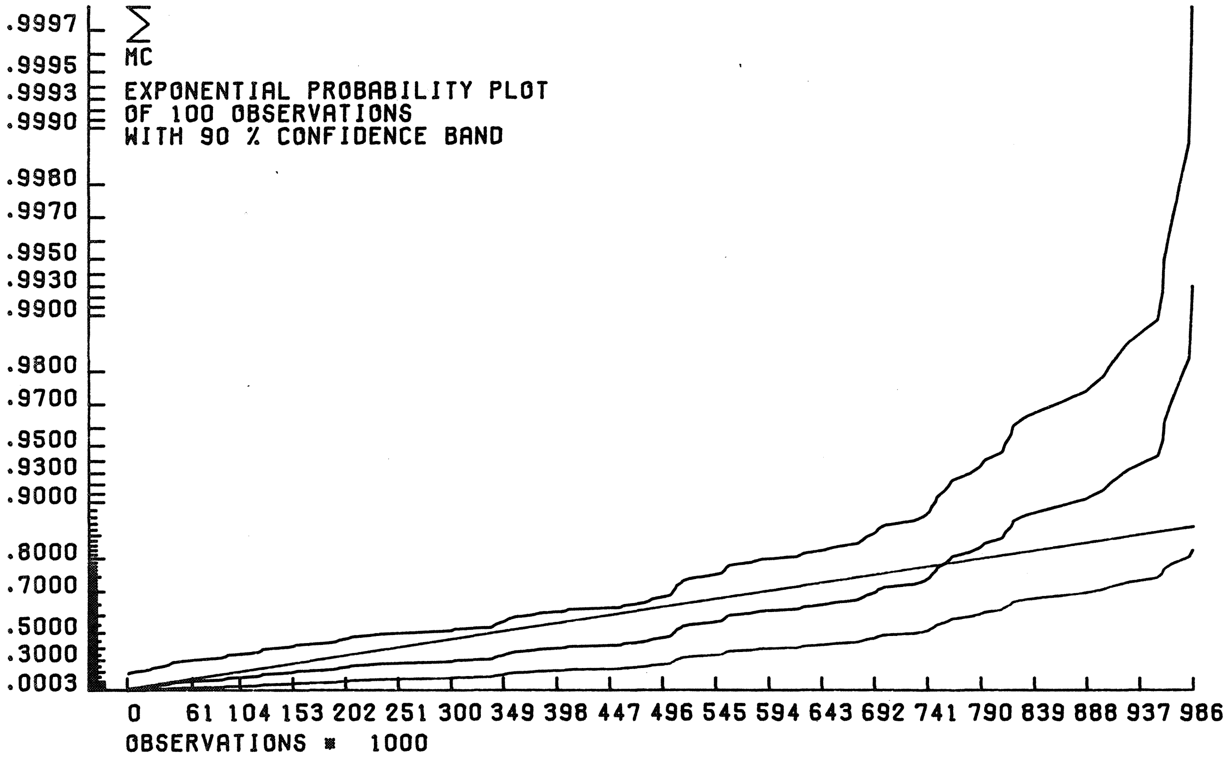


Figure 104. Sample 1

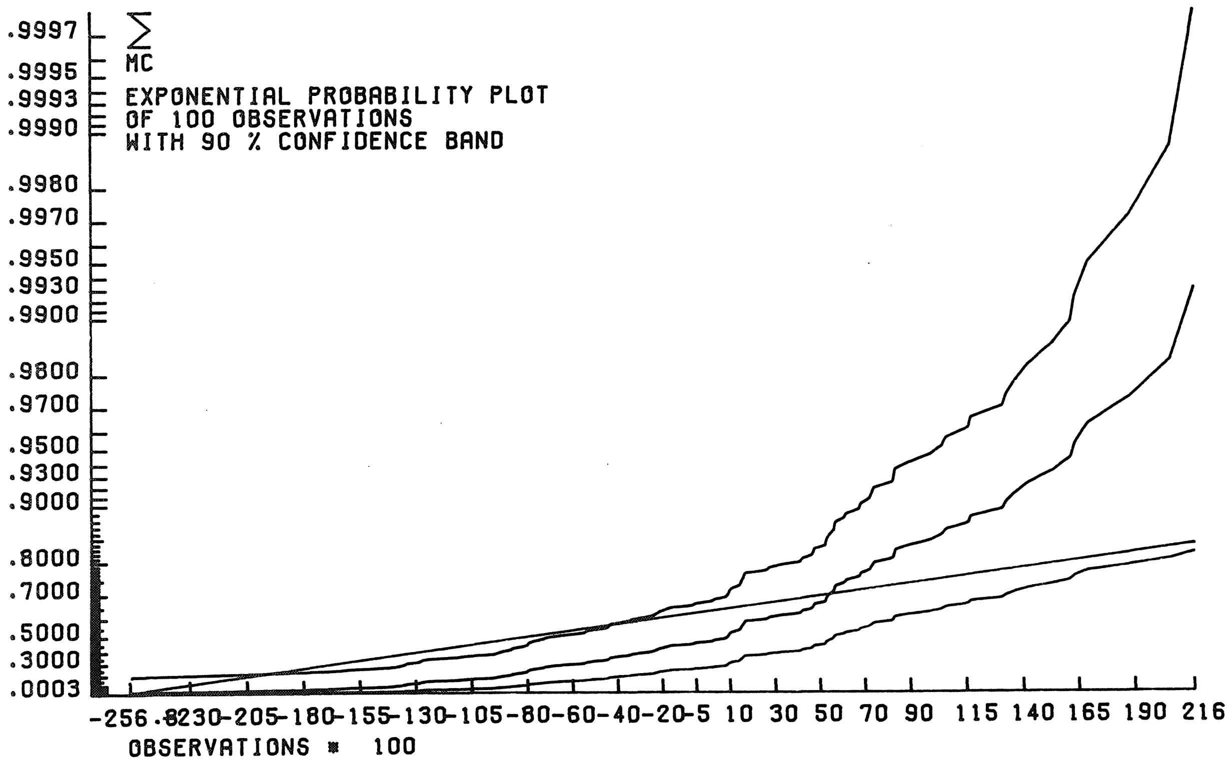


Figure 105. Sample 2

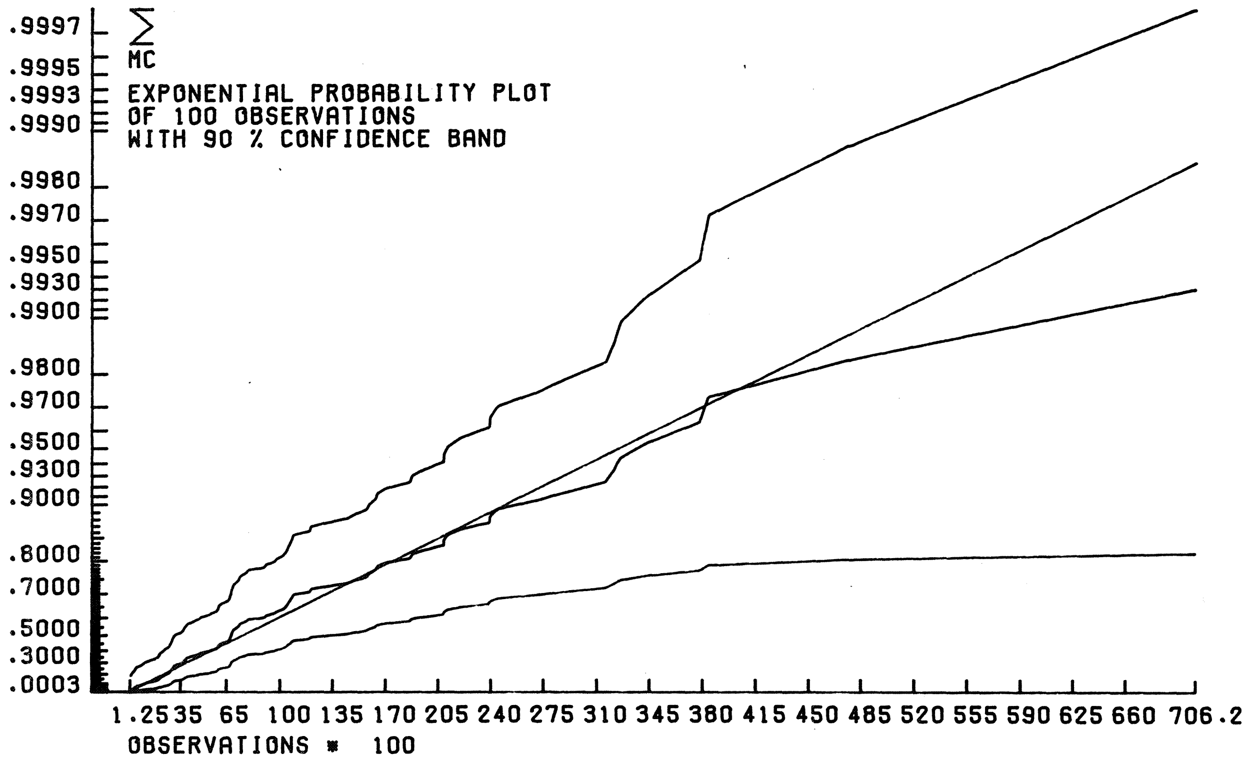


Figure 106. Sample 3

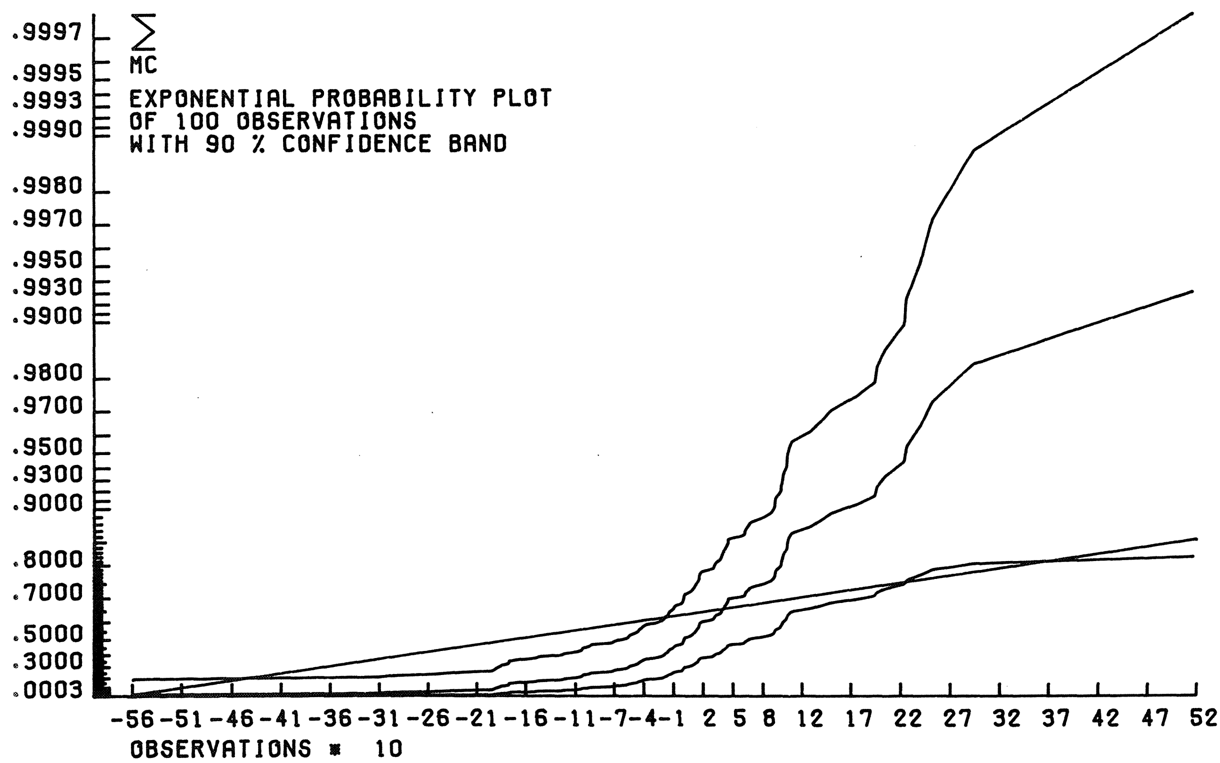


Figure 107. Sample 4

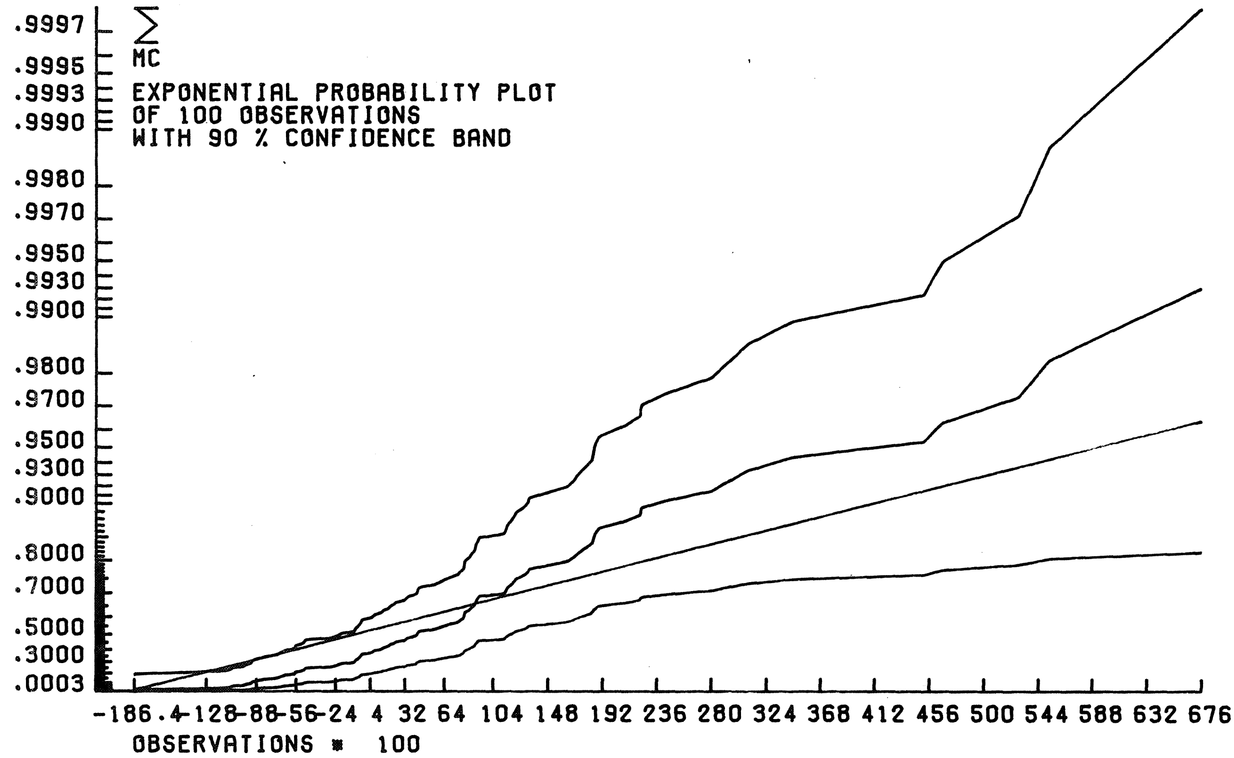


Figure 108. Sample 5

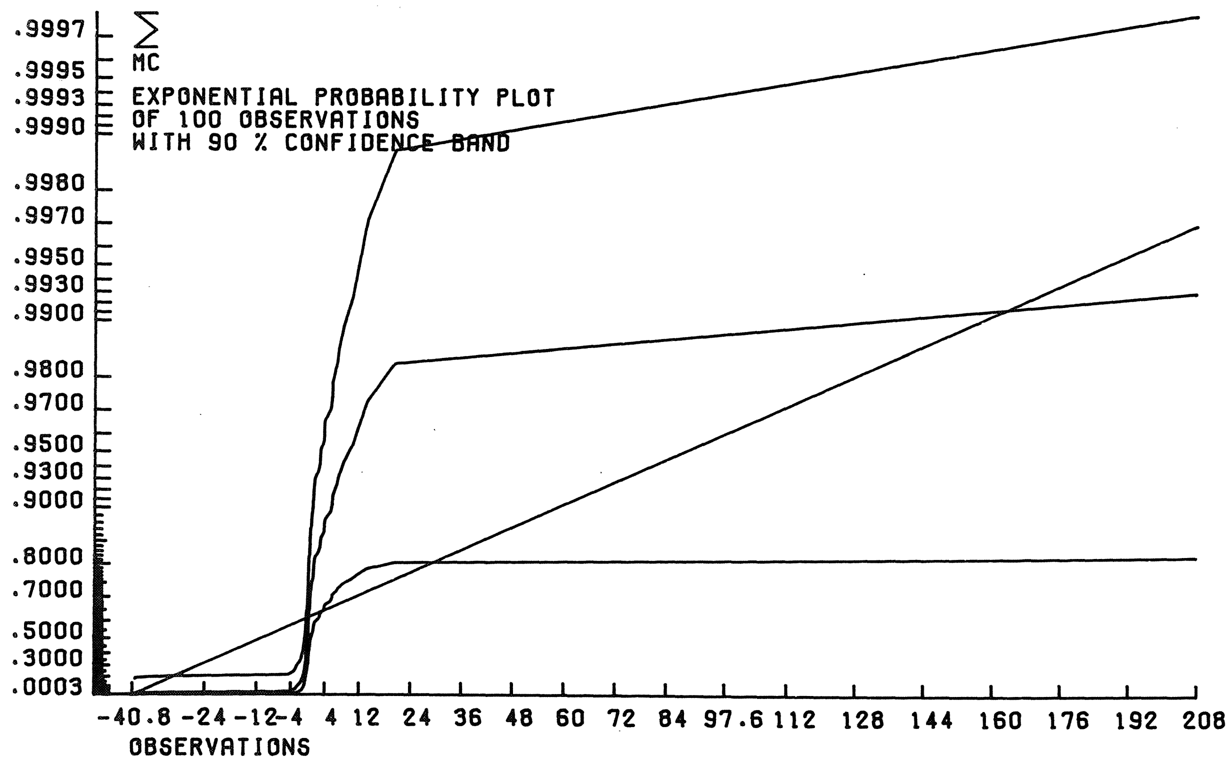


Figure 109. Sample 6

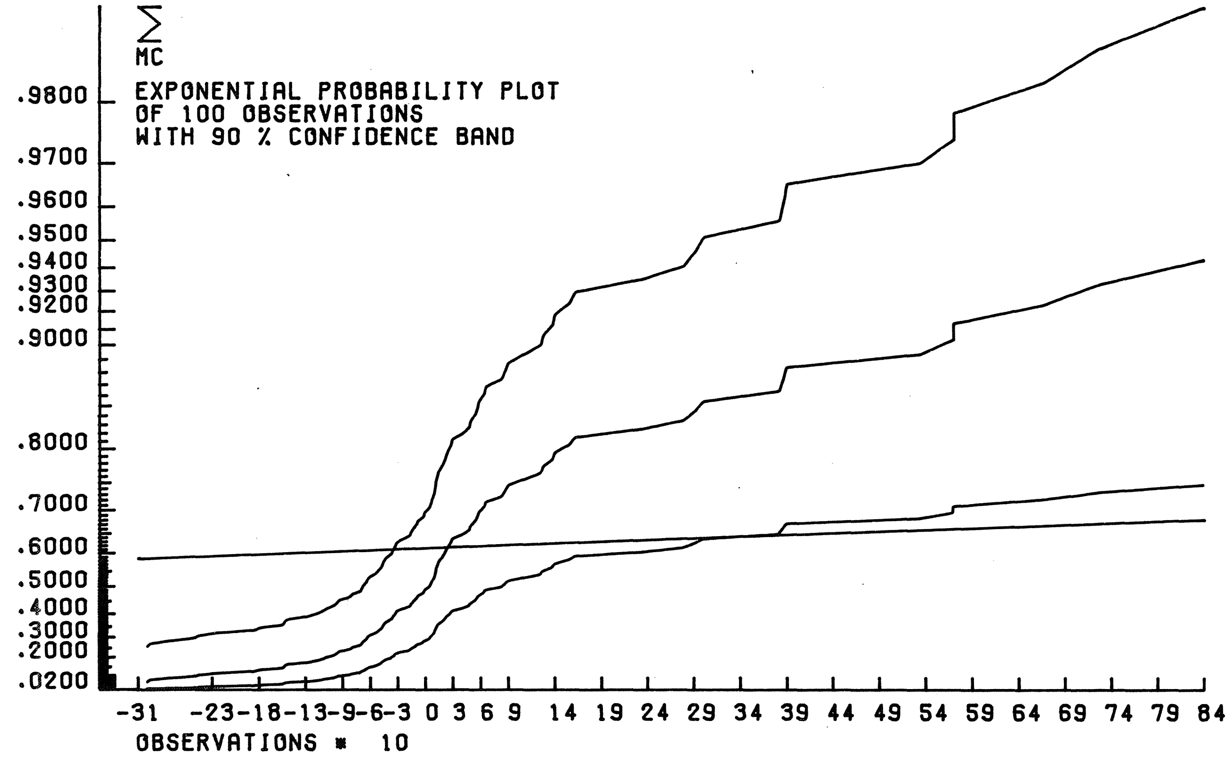


Figure 110. Sample 6, enlargement (LB = 0.05, UB = 0.95)

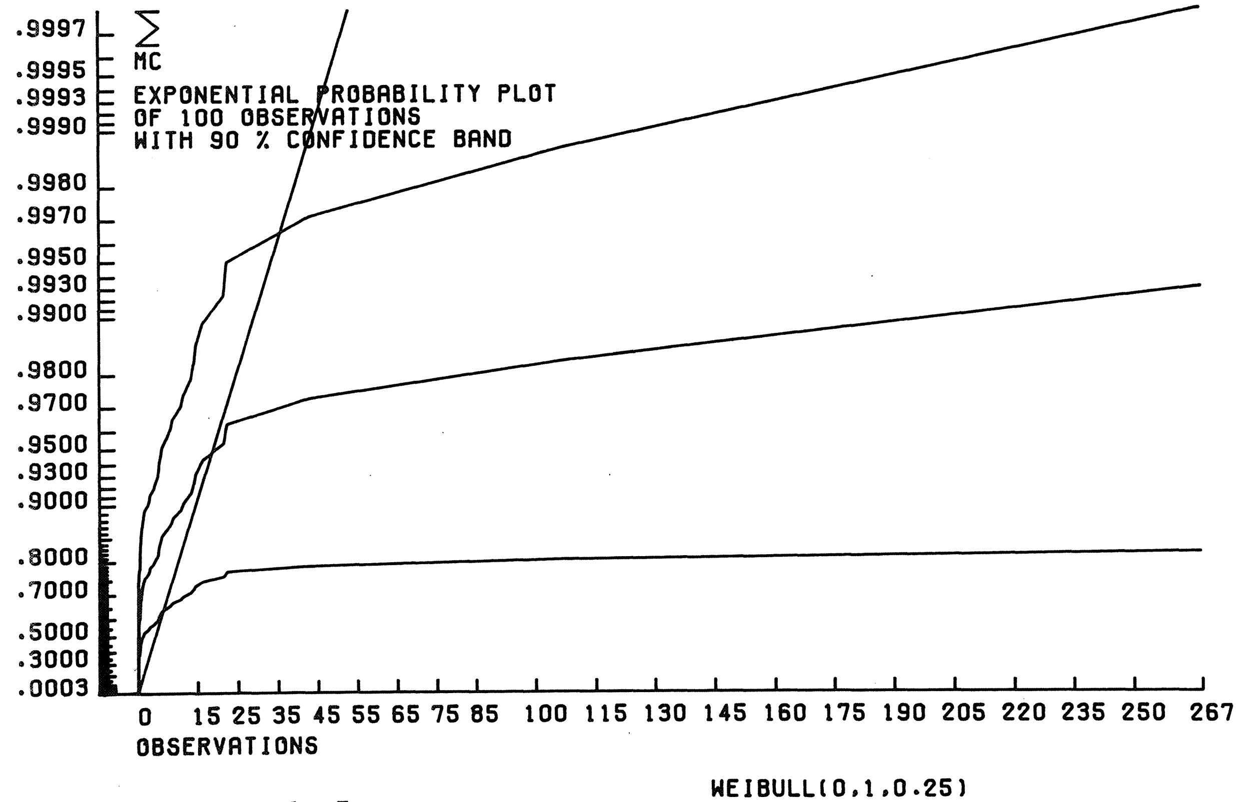


Figure 111. Sample 7

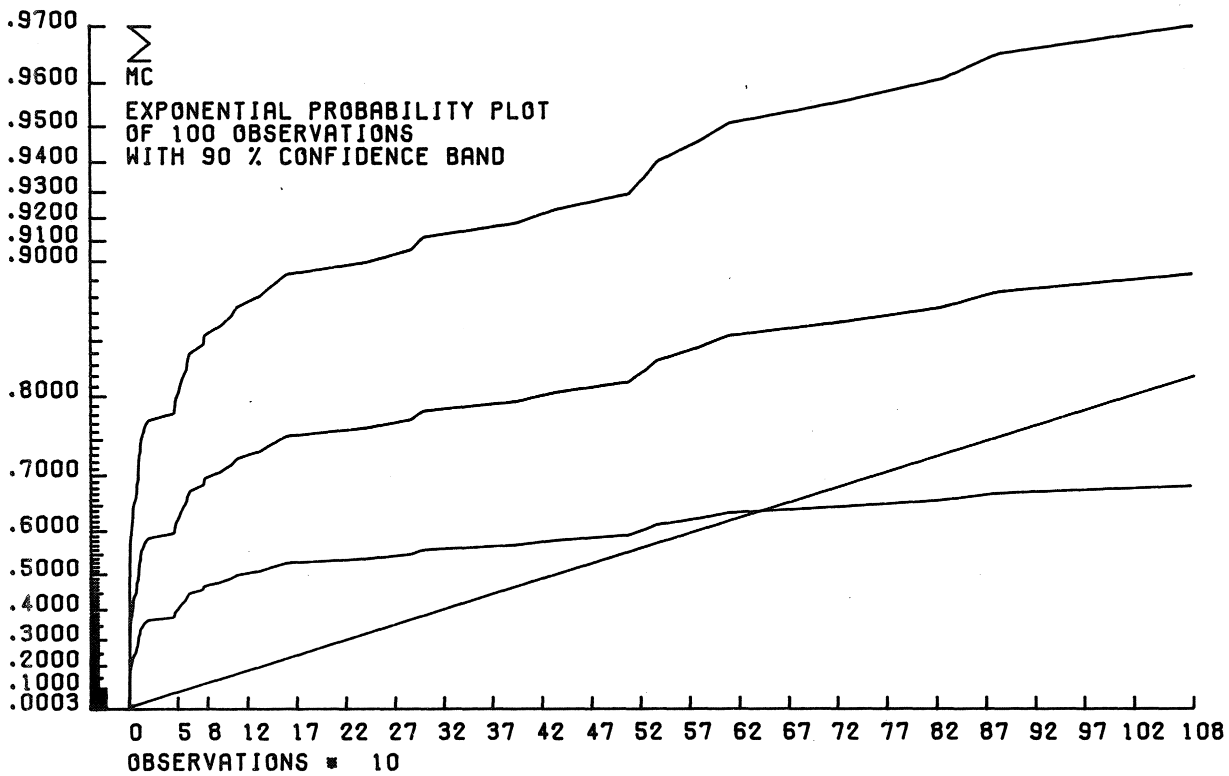


Figure 112. Sample 7, enlargement (UB = 0.9)

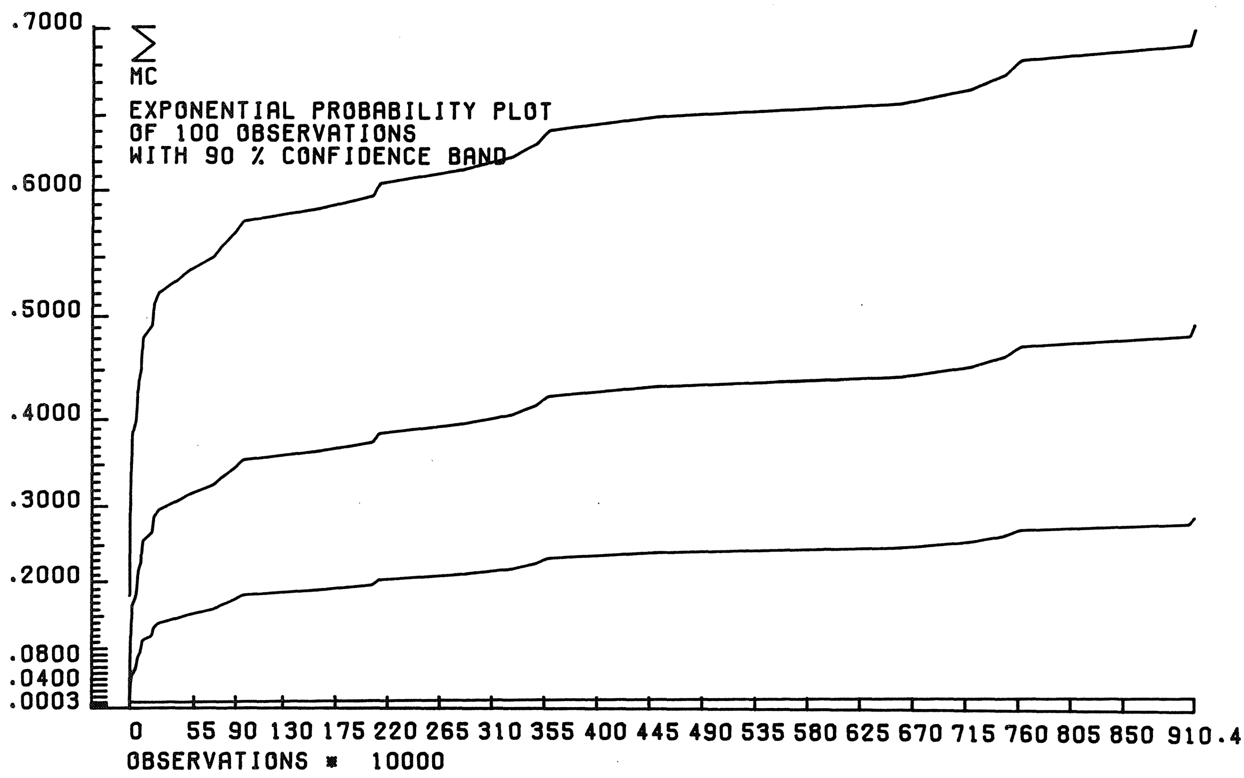


Figure 113. Sample 7, enlargement (UB = 0.5)

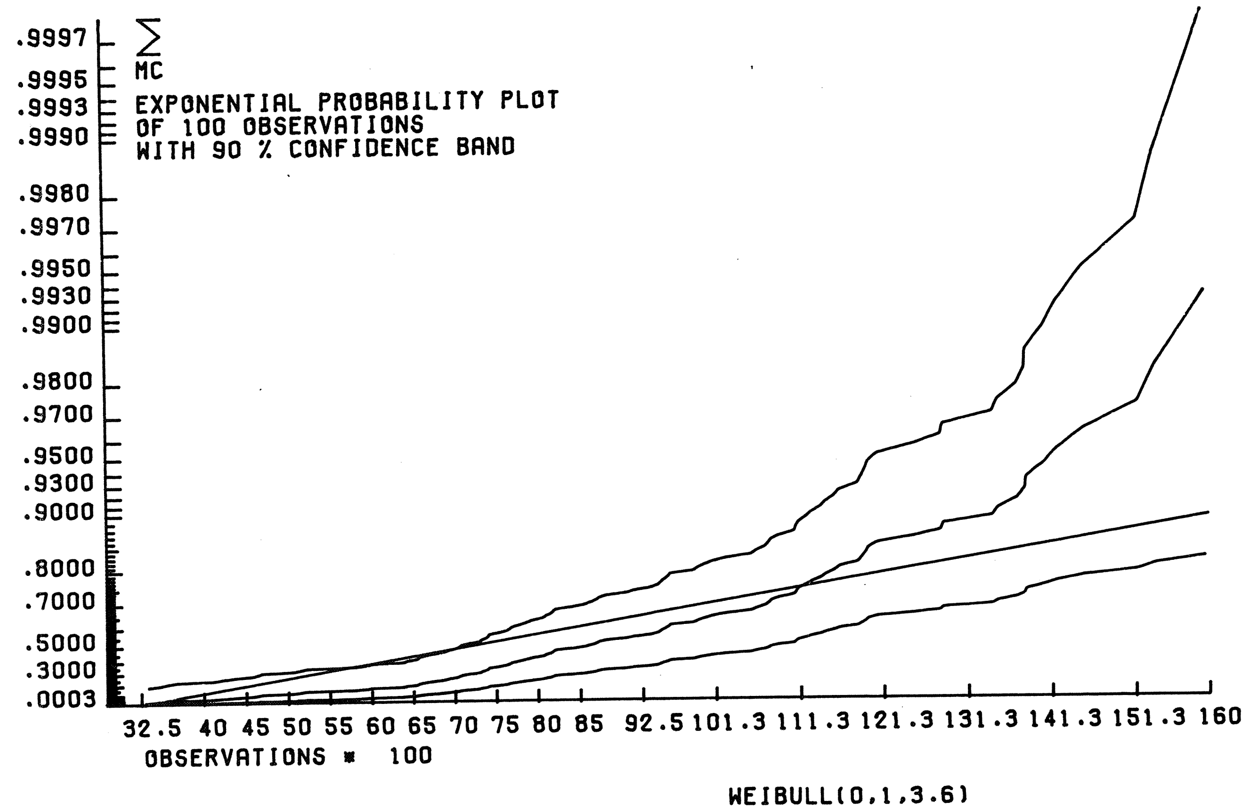


Figure 114. Sample 8

5.2.3.3. Different types of probability plots of uniformly, normally and exponentially distributed samples

The first sample is from the uniform (0,1) distribution, the second one is from the standard normal distribution and the third one is from the exponential (0,1) distribution.

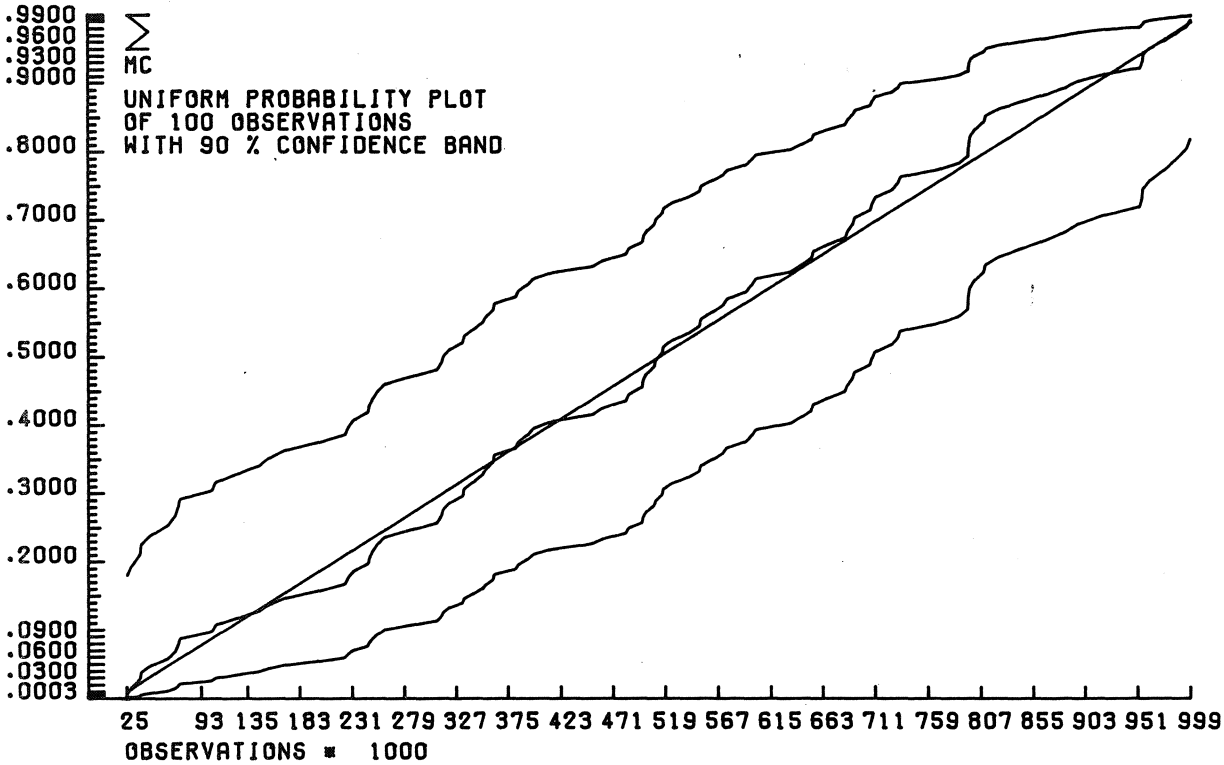


Figure 115. First sample

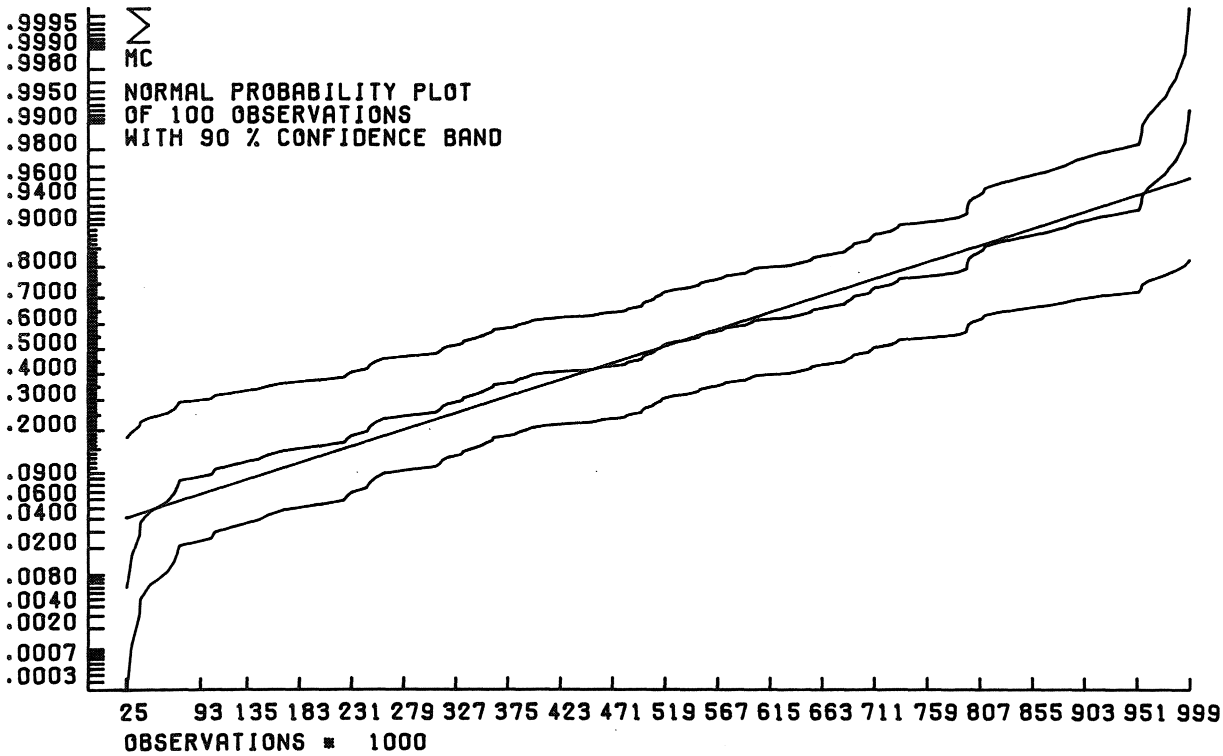


Figure 116. First sample

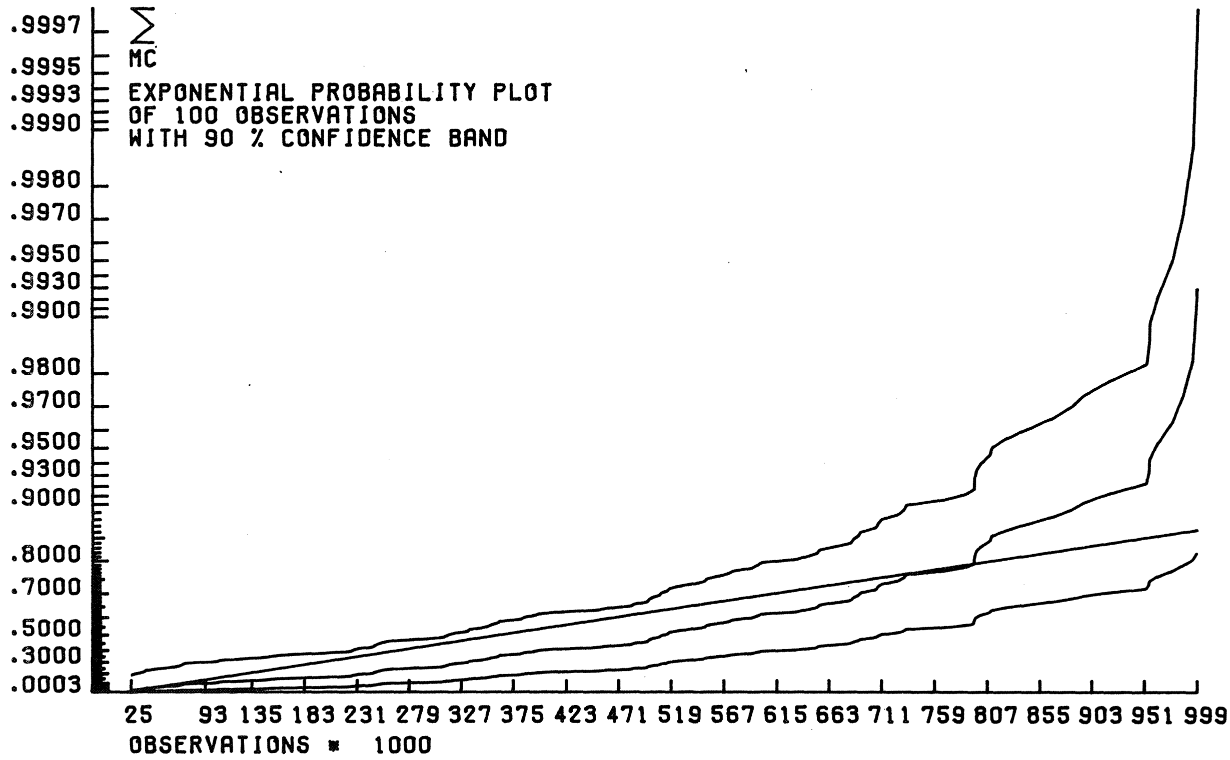


Figure 117. First sample

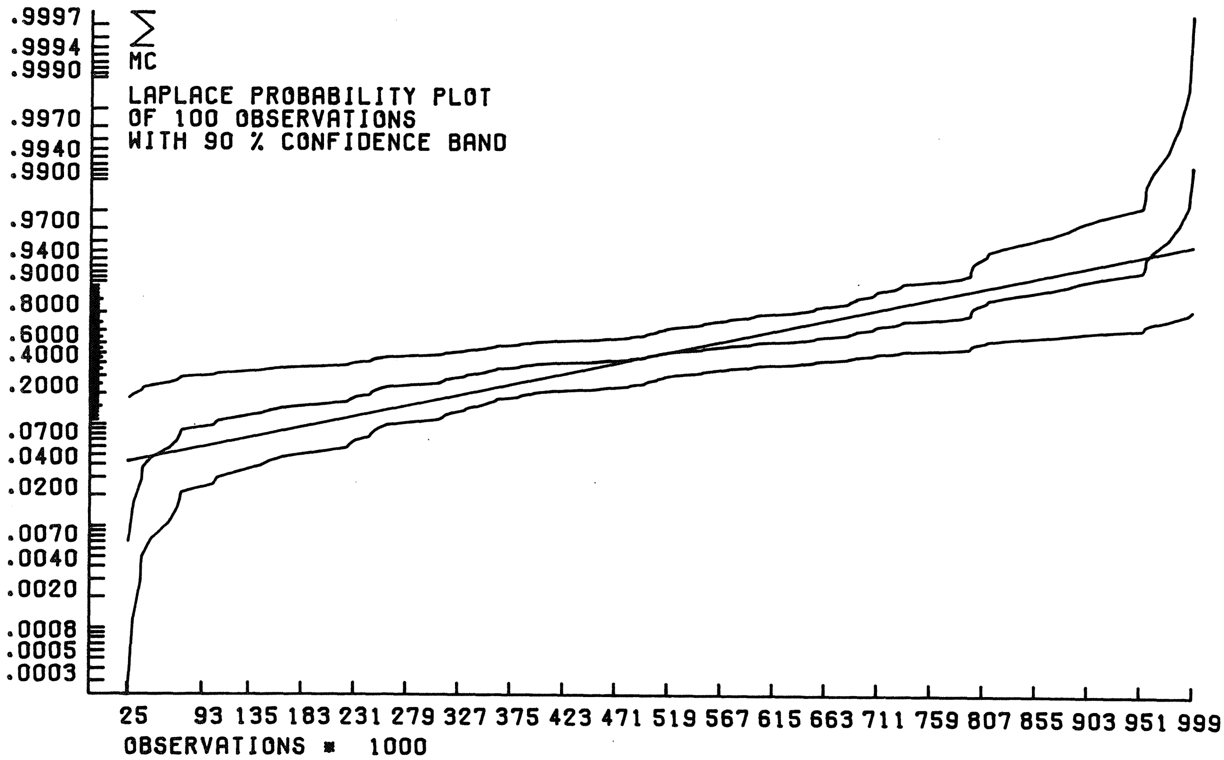


Figure 118. First sample

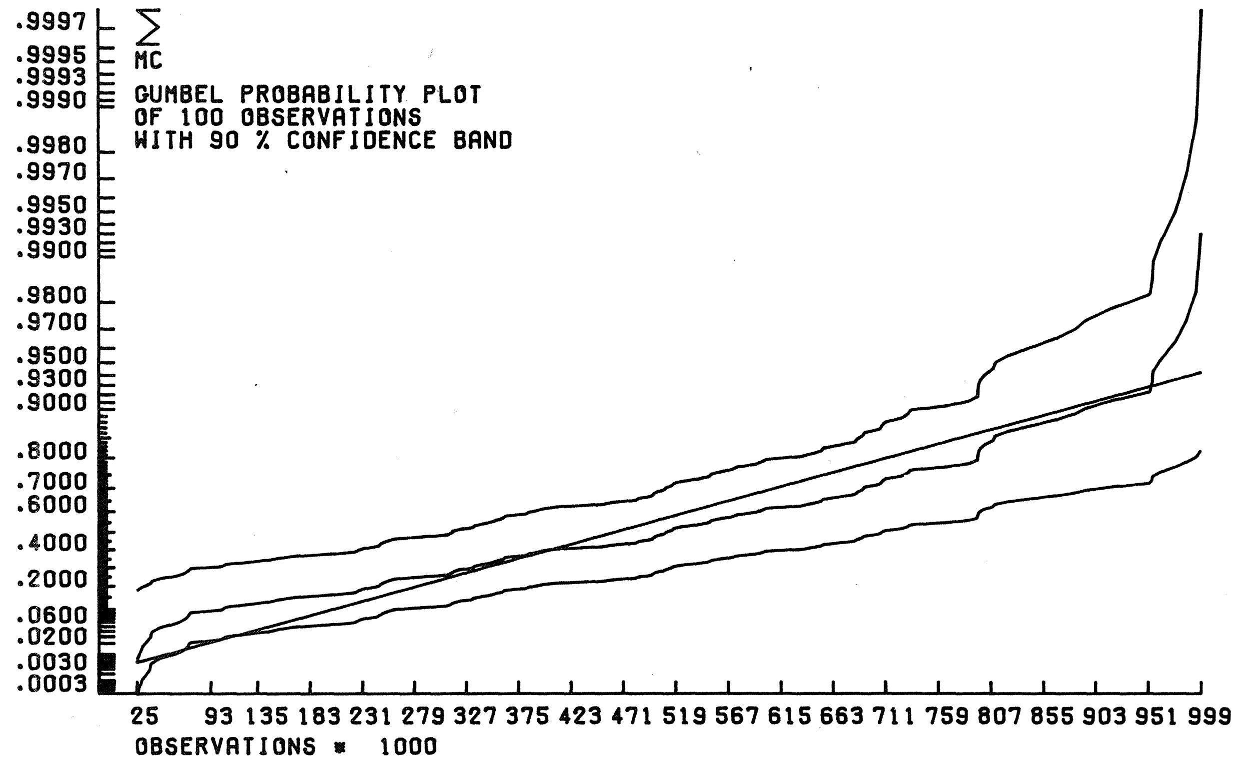


Figure 119. First sample

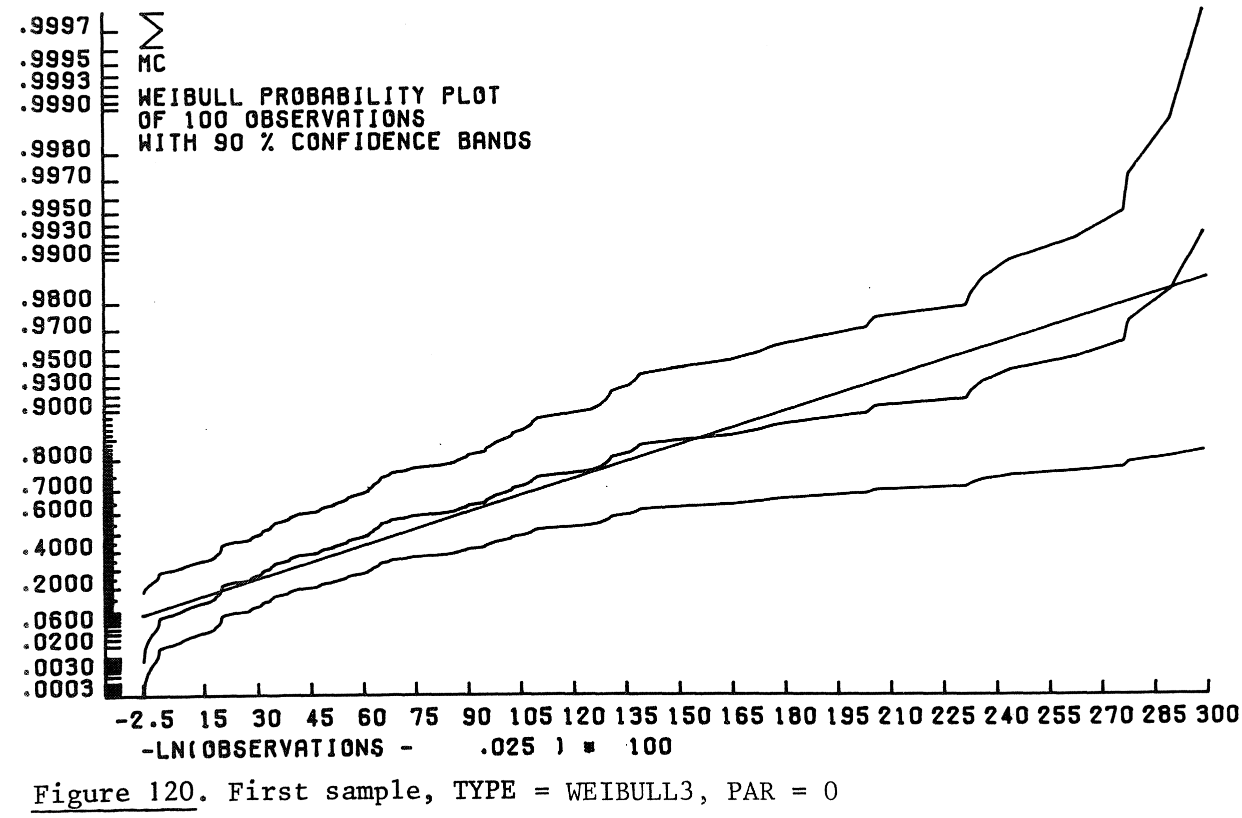


Figure 120. First sample, TYPE = WEIBULL3, PAR = 0

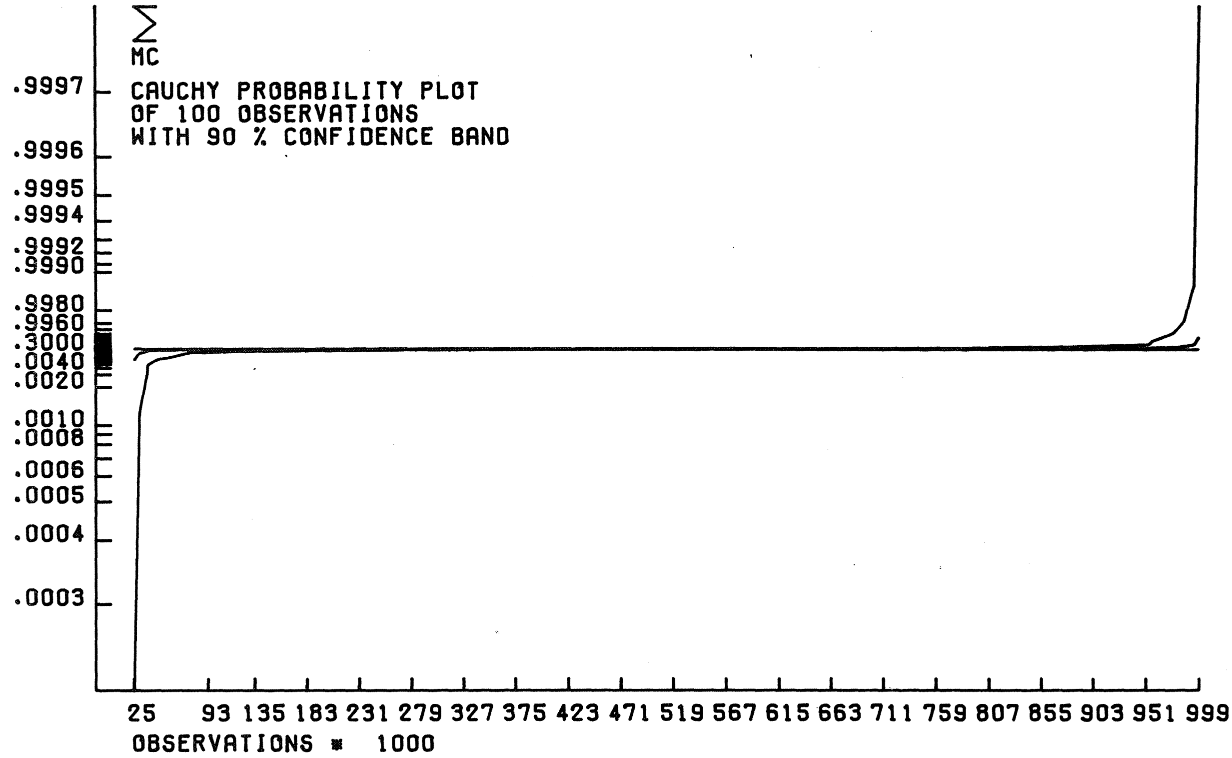


Figure 121. First sample

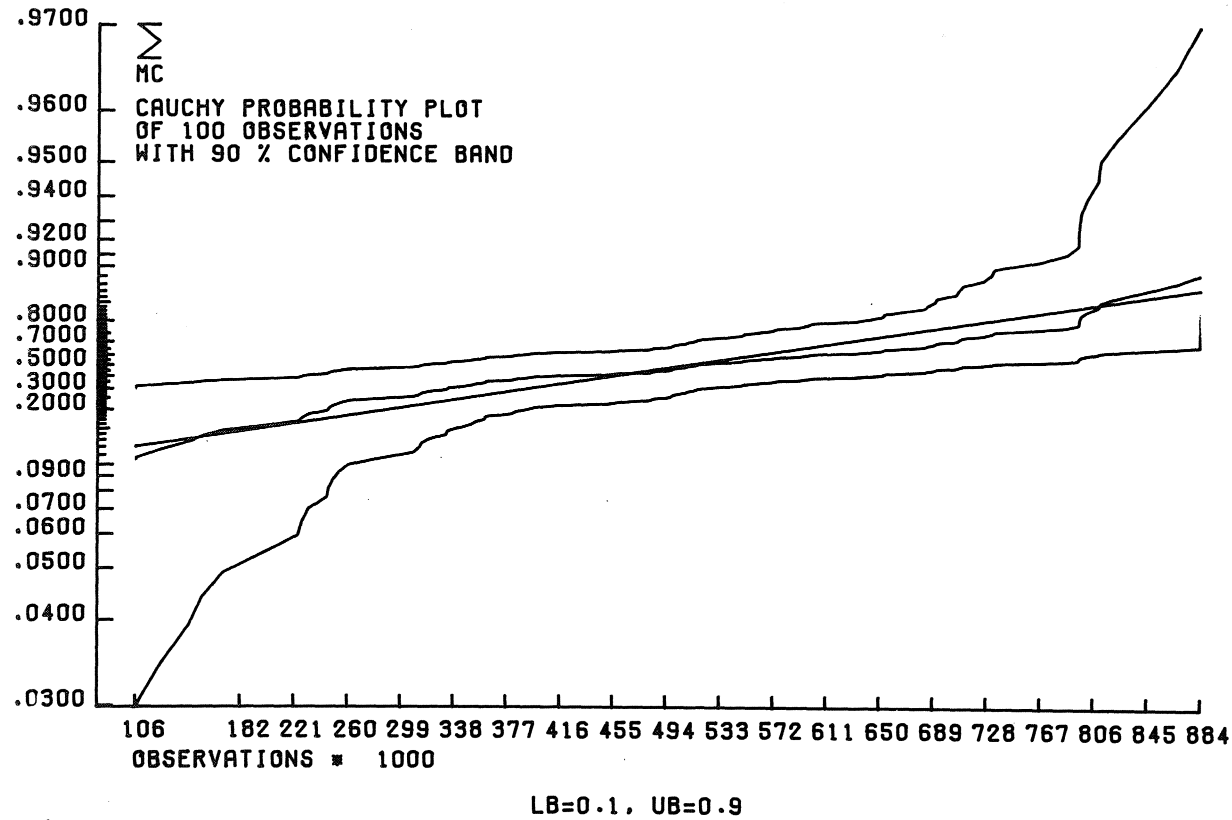


Figure 122. First sample, enlargement

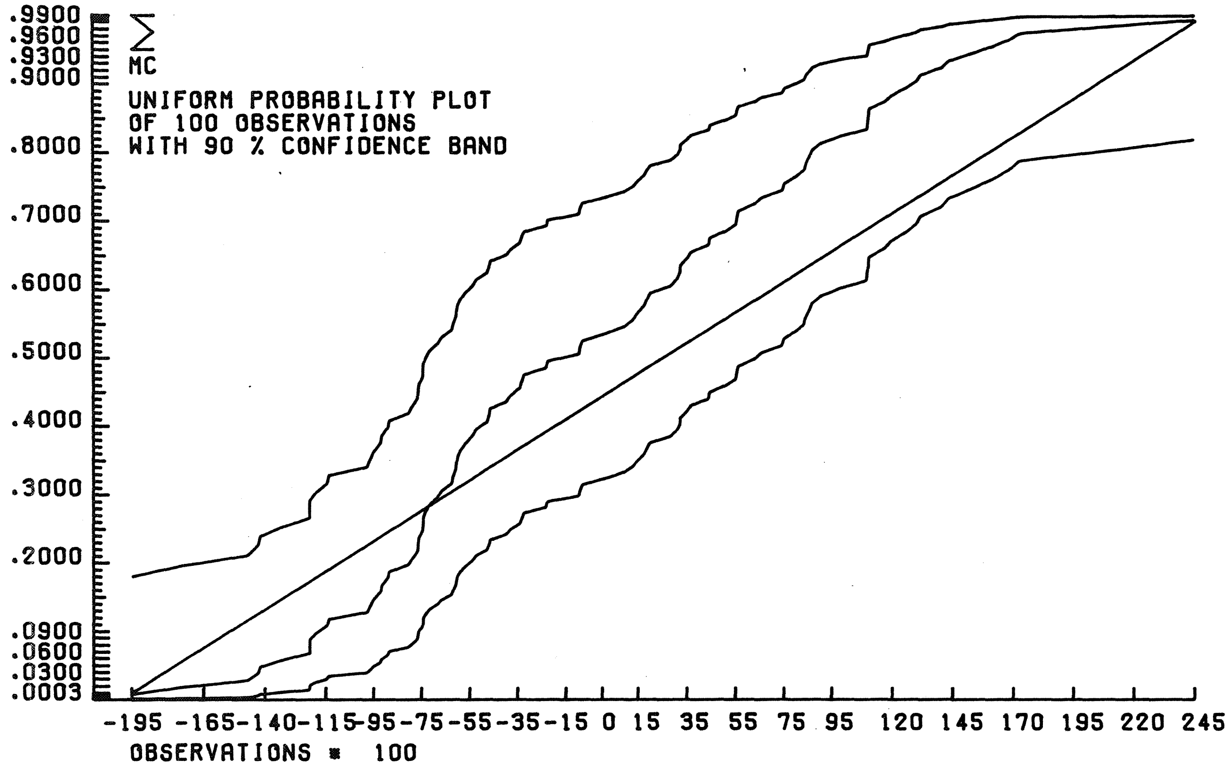


Figure 123. Second sample (from the normal (0,1) distribution)

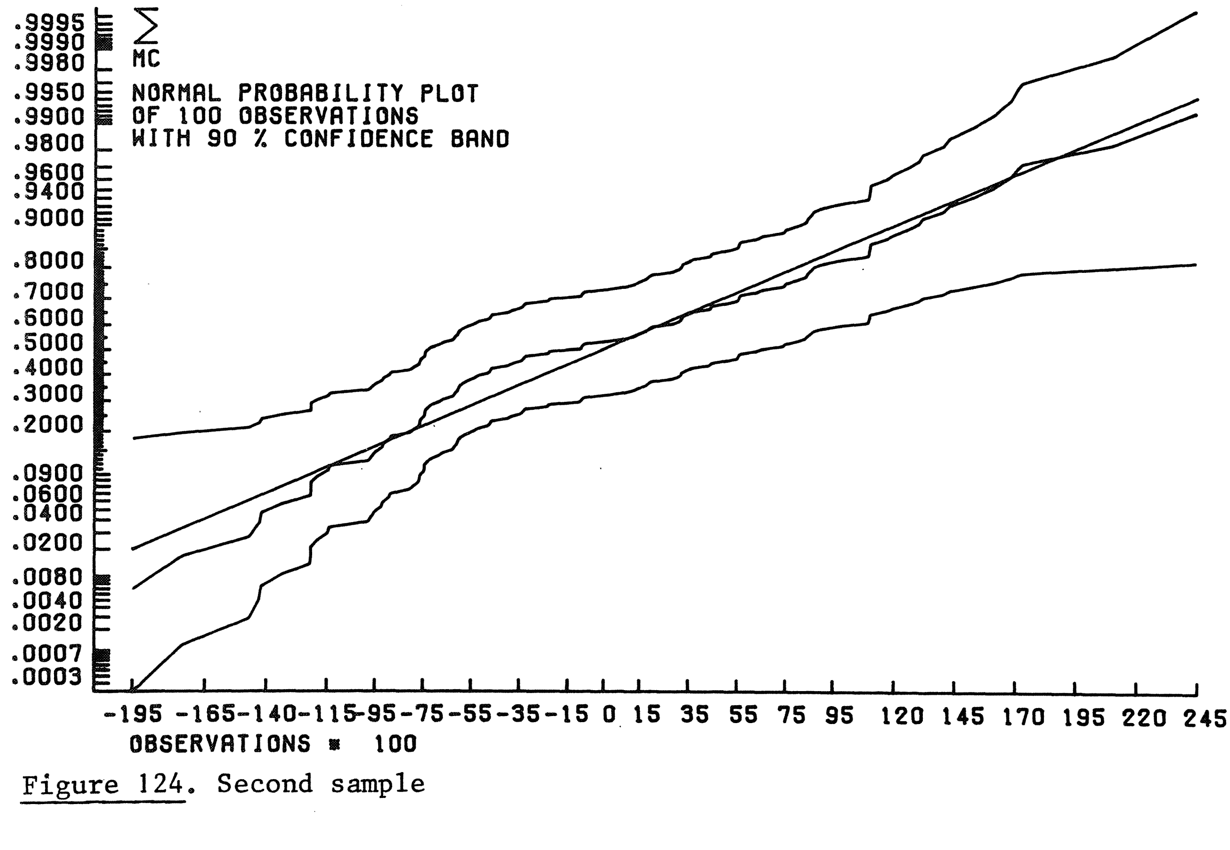


Figure 124. Second sample

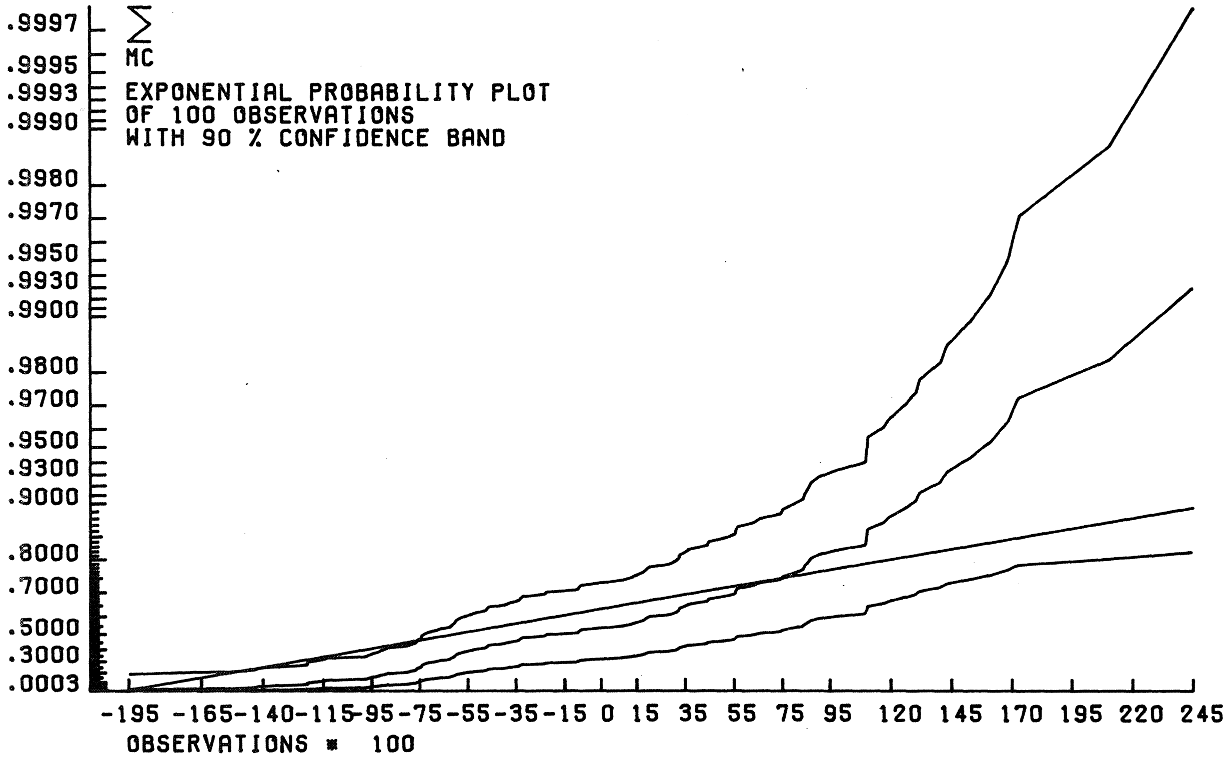


Figure 125. Second sample

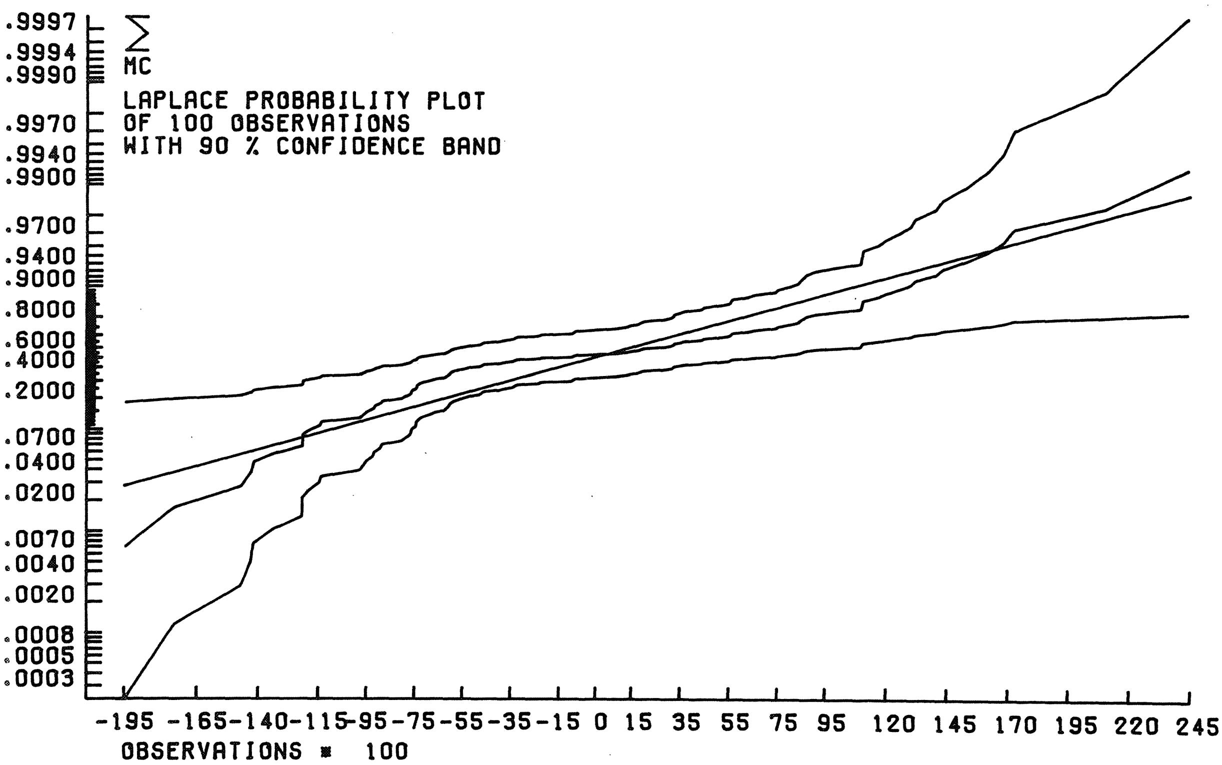


Figure 126. Second sample

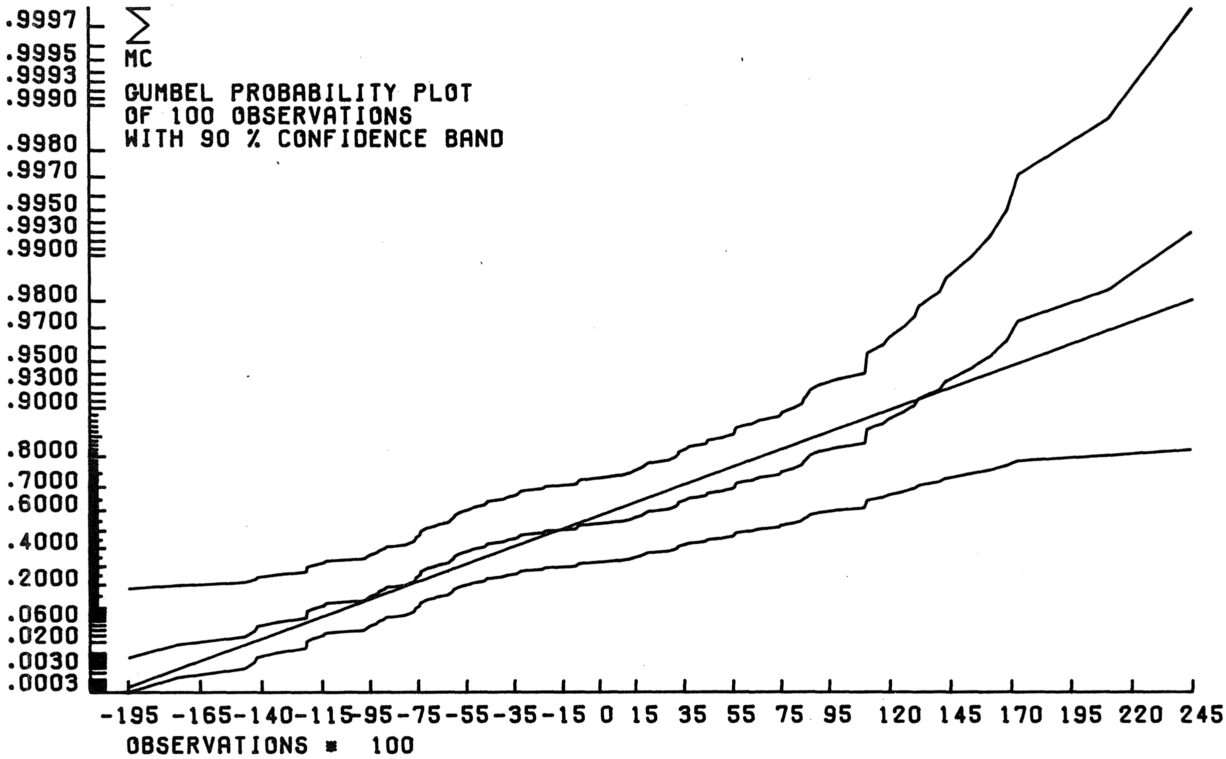


Figure 127. Second sample

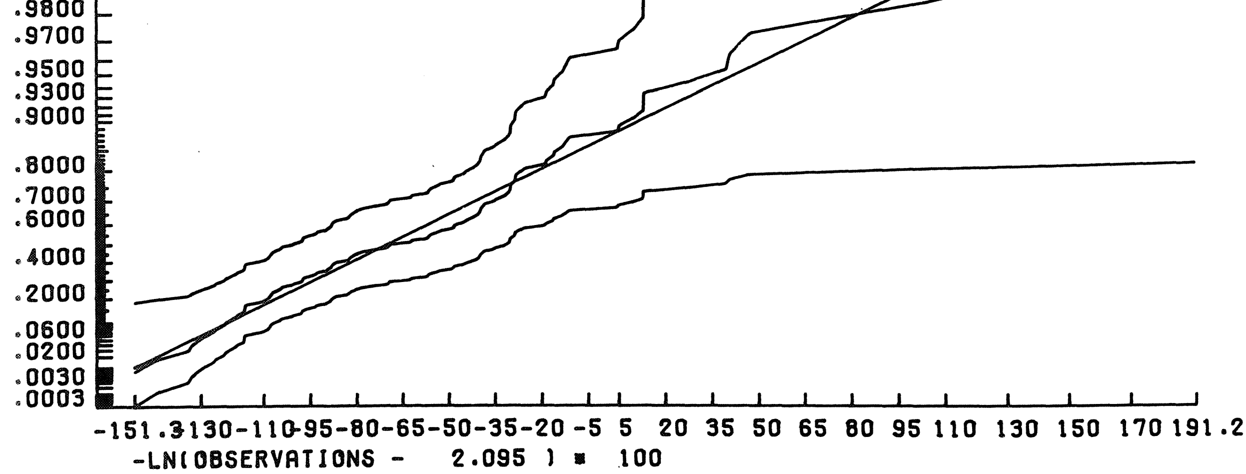


Figure 128. Second sample, TYPE = WEIBULL3, PAR = 0

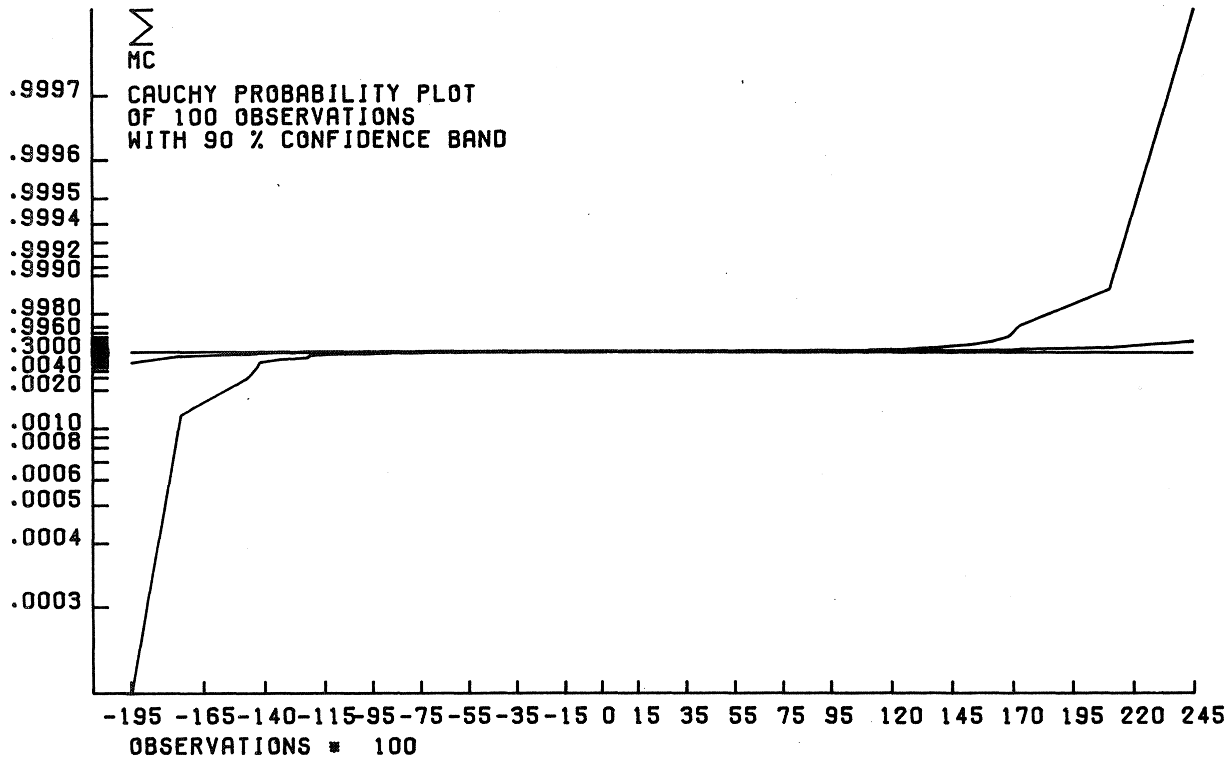


Figure 129. Second sample

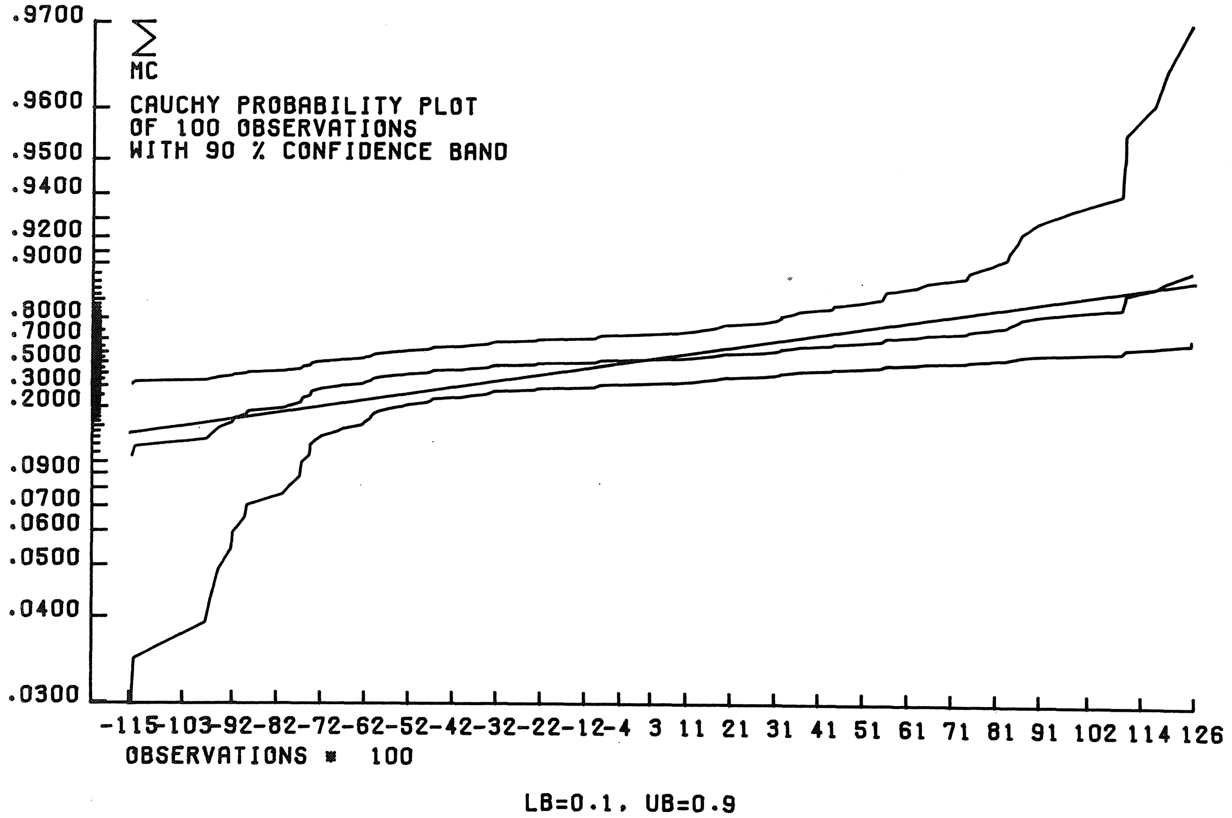


Figure 130. Second sample, enlargement

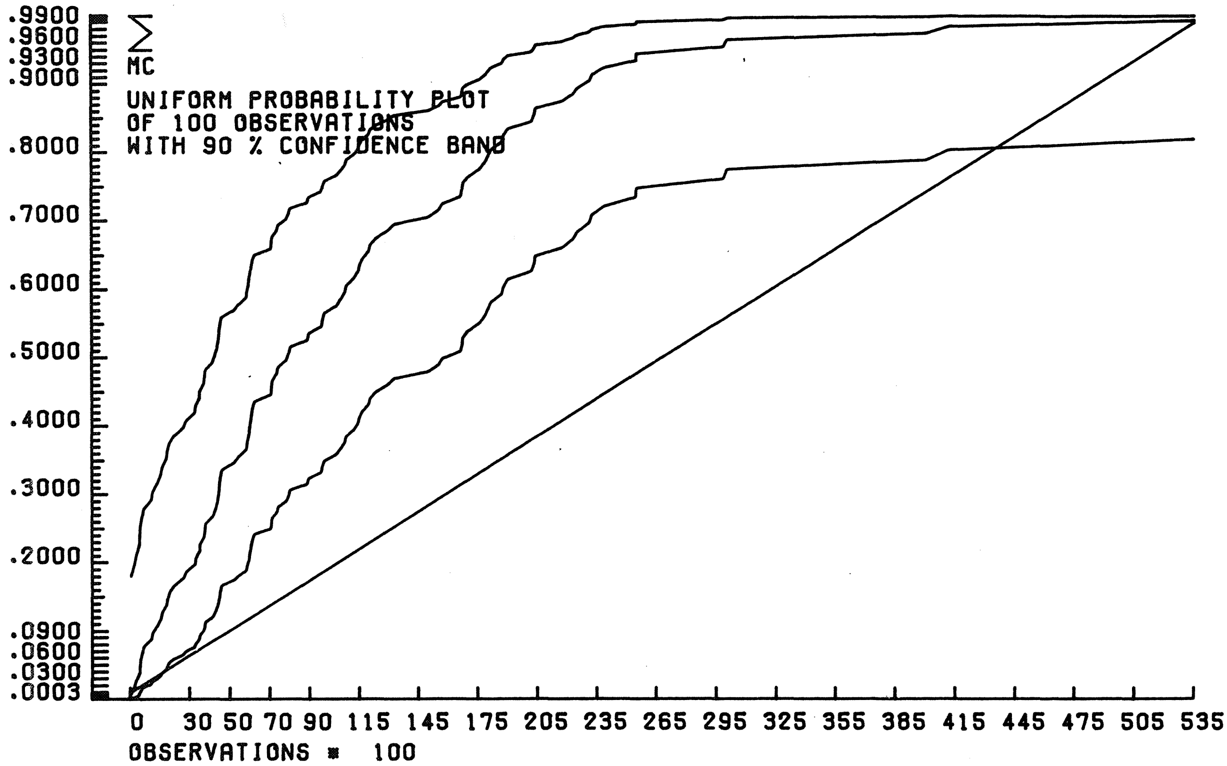


Figure 131. Third sample (from the exponential (0,1) distribution)

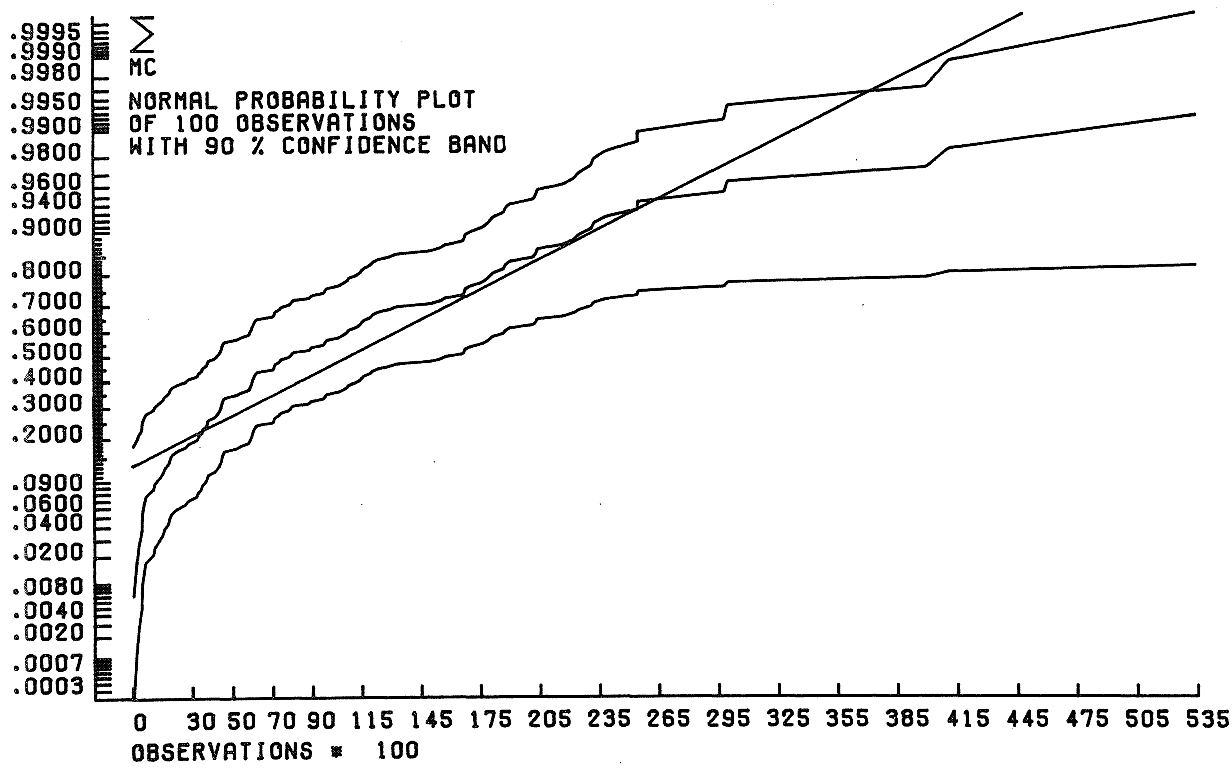


Figure 132. Third sample

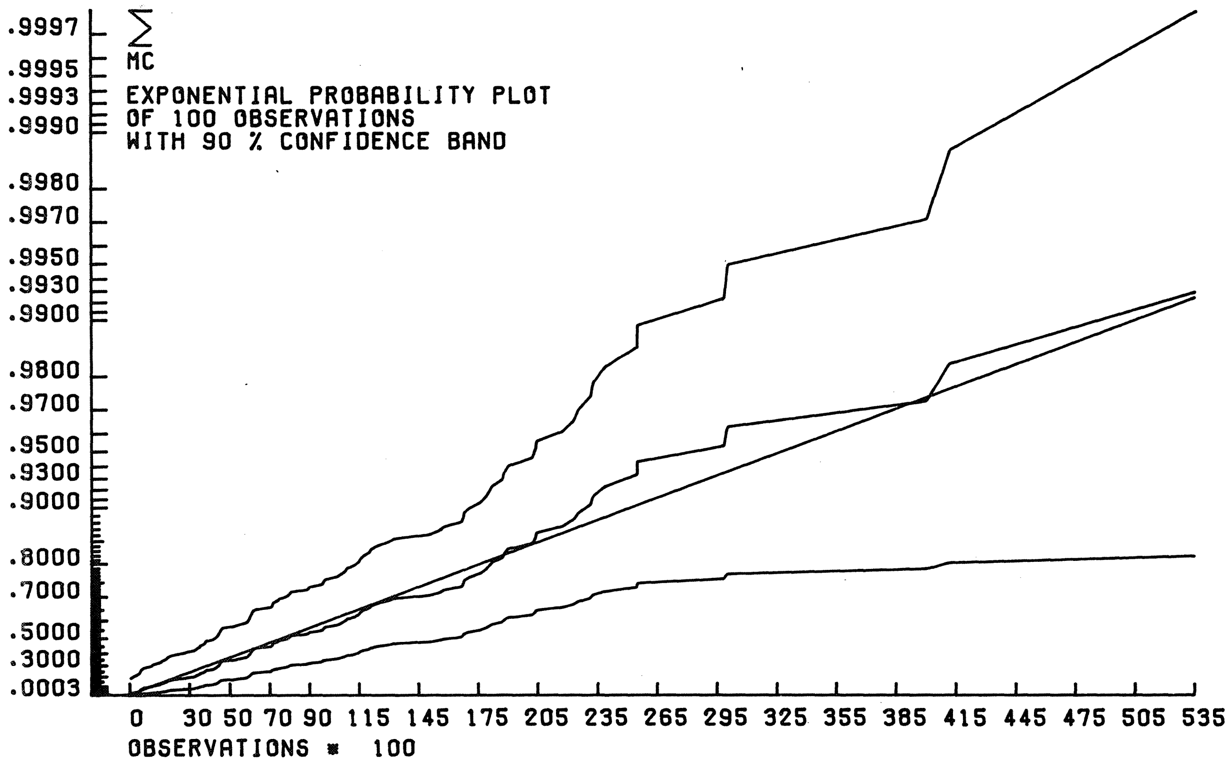


Figure 133. Third sample

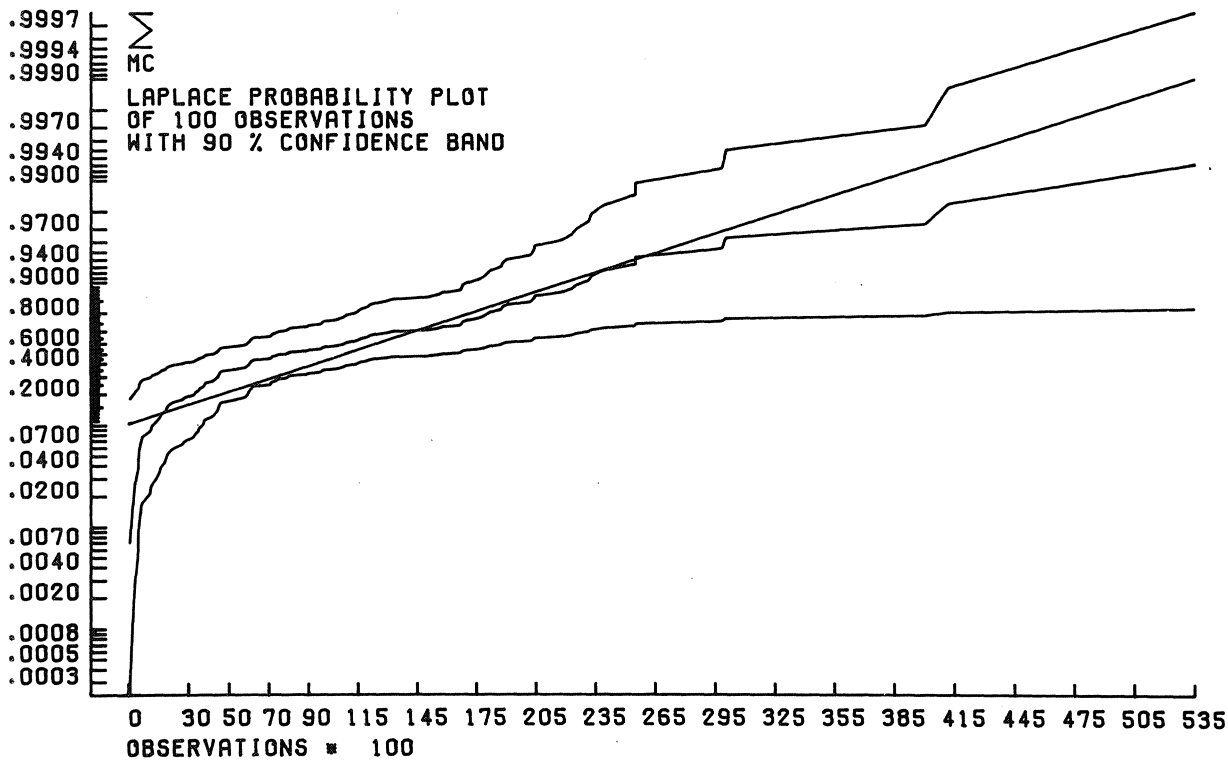


Figure 134. Third sample

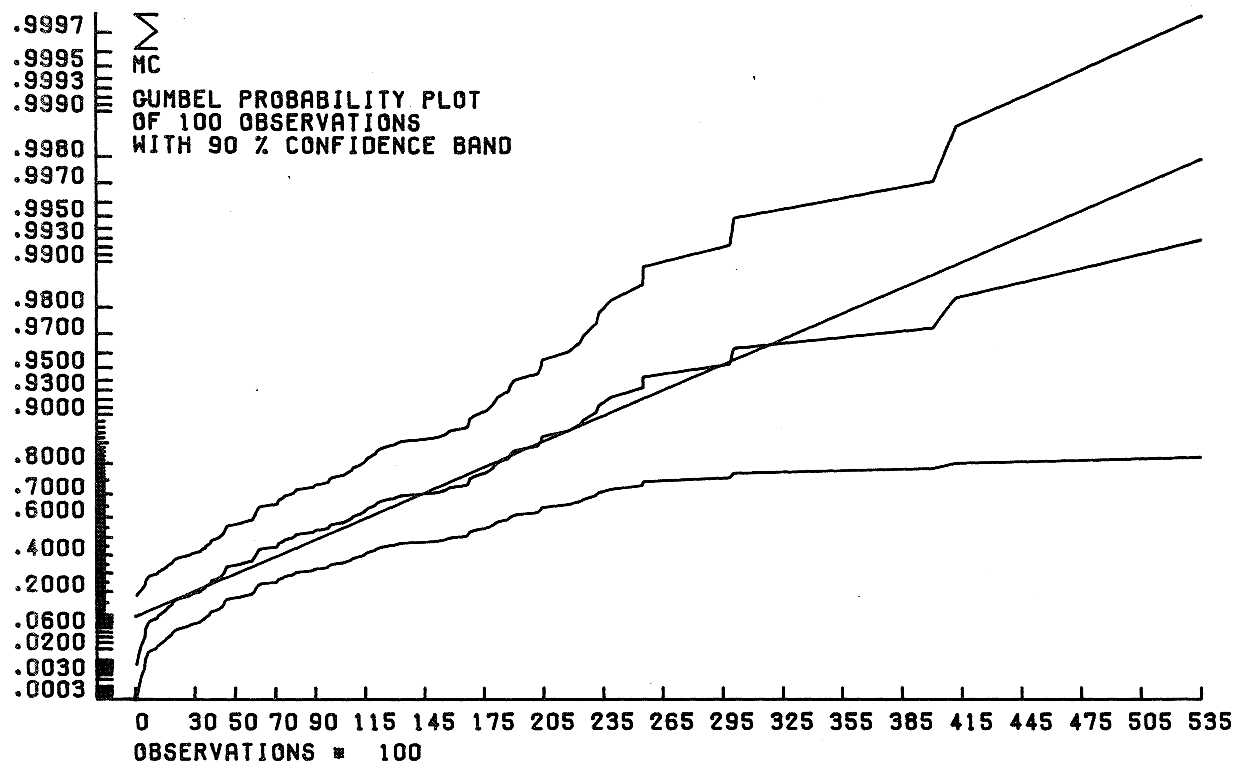


Figure 135. Third sample

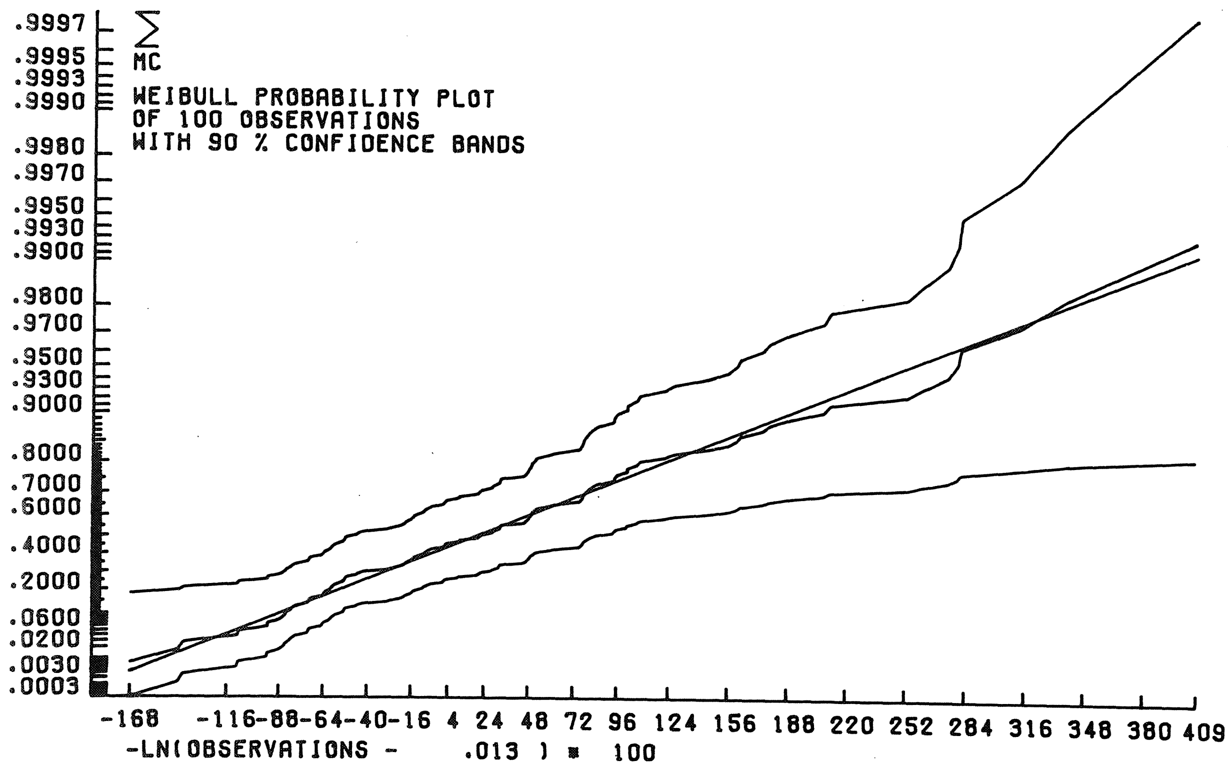


Figure 136. Third sample. Note that exponential (0,1) = Weibull (0,1,1)!

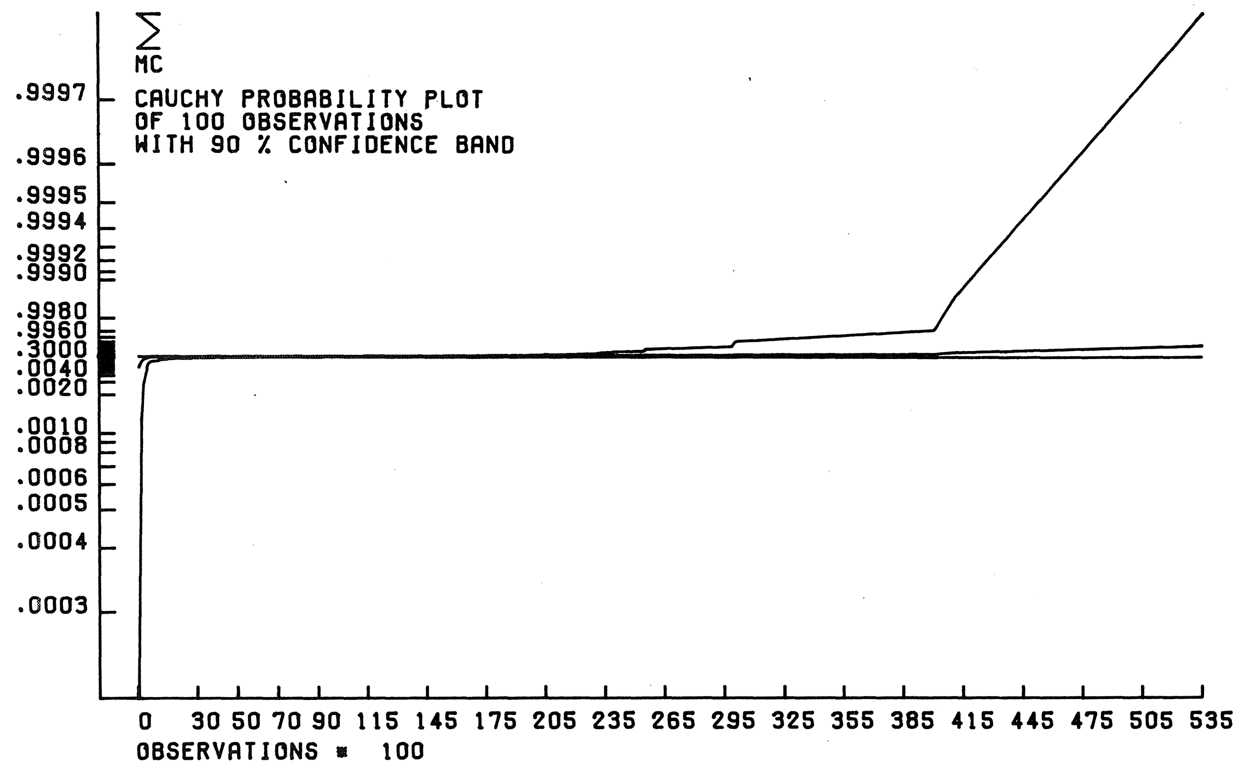


Figure 137. Third sample

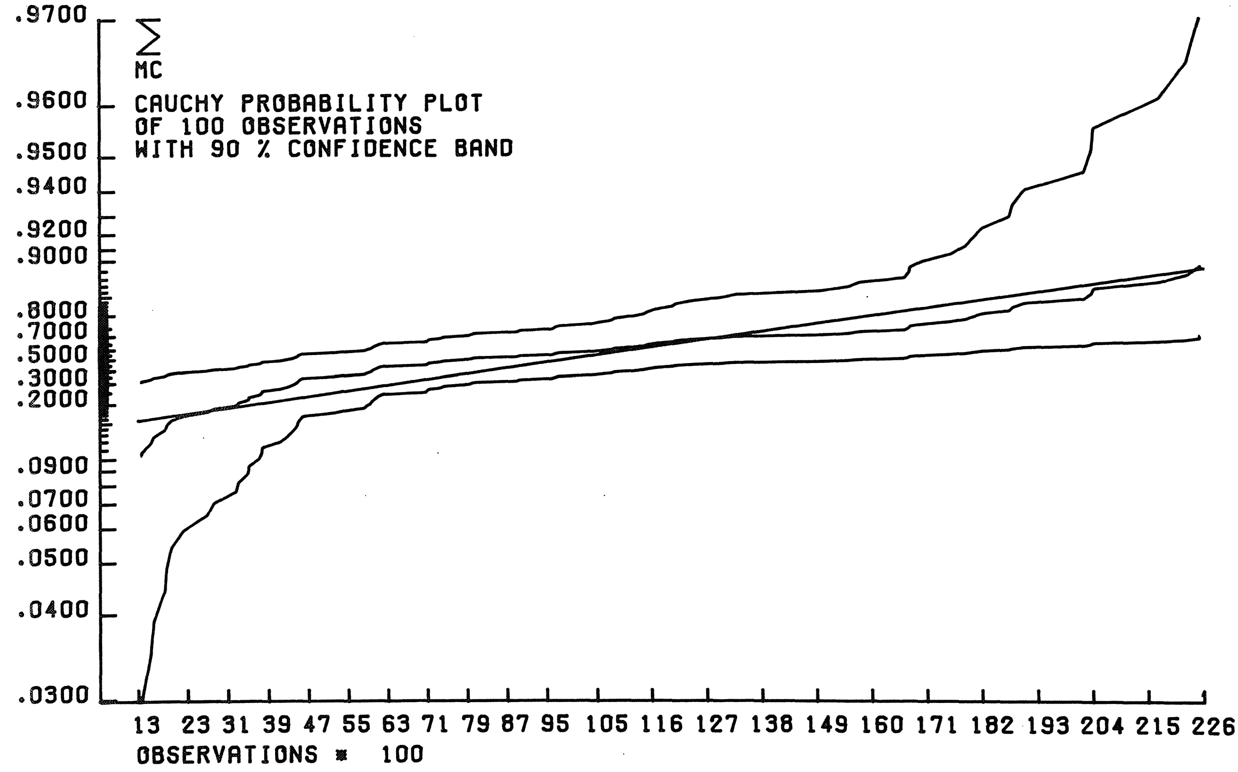


Figure 138. Third sample, enlargement

ACKNOWLEDGEMENTS

We thank Ingeborg van der Tweel for substantial contributions to the computerprogram, Jaap Wisse for producing the plots in chapter 5, Ed Opperdoes for adjusting the program to the new ALGOL compiler and Jon Nool for testing the adjusted version.

REFERENCES

[1] ANSCOMBE, F.J. & J.W. TUKEY (1963), *The Examination and Analysis of Residuals*, Technometrics 5, 141-160.

[2] BENARD, A. & E.C. BOS-LEVENBACH (1953), *Het uitzetten van waarnemingen op waarschijnlijkheidspapier*, Statistica Neerlandica 7, 163-173 (in Dutch).

[3] BICKEL, P.J. & K.A. DOKSUM (1977), *Mathematical Statistics: Basic Ideas and Selected Topics*, Holden-Day, San Francisco.

[4] BURY, K.V. (1975), *Statistical Models in Applied Science*, J. Wiley, New York.

[5] DOKSUM, K.A. (1977), *Some Graphical Methods in Statistics. A Review and Some Extensions*, Statistica Neerlandica 31, 53-68.

[6] ES, A.J. VAN & C. VAN PUTTEN (1979), *The Statal Random Number Generator*, Statal Report 1, Techn. Report SN 8/79, Mathematical Centre, Amsterdam.

[7] HEMELRIJK, J. & J. KRIENS (1976), *Leergang Besliskunde, Deel 3, Statistiek*, MC Syllabus 1.3, Mathematisch Centrum, Amsterdam, 1-15 (in Dutch).

[8] JOHNSON, N.L., & S. KOTZ (1970), *Distribution in Statistics, Continuous Univariate Distributions* vol. 1 & 2, Houghton Mifflin, Boston.

[9] LOÈVE, M. (1960), *Probability Theory*, D. van Nostrand, Princeton, 20.

[10] NUMAL, *A Library of Numerical Procedures in ALGOL 60*, Reference Manual, Mathematical Centre, Amsterdam, or
HEMKER, P.W. (Ed.) (1981), *Numal, Numerical Procedures in ALGOL 60*, MC Syllabus 47, Mathematical Centre, Amsterdam.

[11] PUTTEN, C. VAN & I. VAN DER TWEEL (1979), *On Generating Random Variables*, Statal Report 2, Techn. Report SN 9/79, Mathematical Centre, Amsterdam.

[12] QUADE, D. (1973), *The Pairchart*, Statistica Neerlandica 27, 29-45.

[13] STATAL, *A Library of Statistical Procedures and Programs*, Reference Manual, Mathematical Centre, Amsterdam.

[14] TUKEY, J.W. (1962), *The Future of Data Analysis*, A.M.S. 33, 1-67.

[15] WILK, M.B. & R. GNANADESIKAN (1968), *Probability Plotting Methods for the Analysis of Data*, Biometrika 55, 1-17.