AD-A056 095 GEORGE WASHINGTON UNIV WASHINGTON D C INST FOR MANAG--ETC F/G 14/4
THE GINI INDEX, THE LORENZ CURVE, AND THE TOTAL TIME ON TEST TR--ETC(U)
FEB 78 M CHANDRA, N D SINGPURWALLA N00014-77-C-0263
UNCLASSIFIED SERIAL-T-368 NL

END
DATE FILMED
8-78
DDC

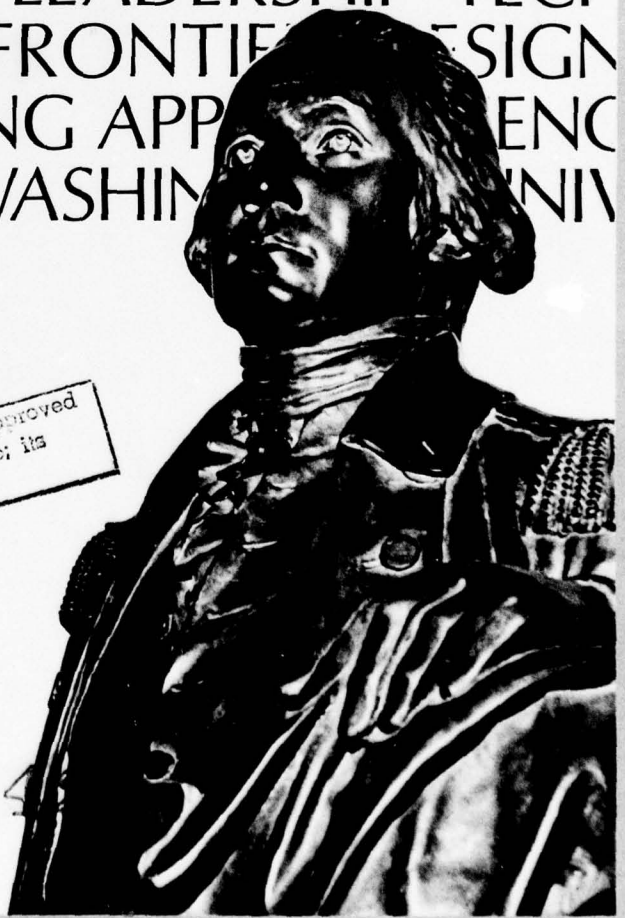
LEVEL II

12

B.S.

THE GEORGE WASHINGTON UNIVERSITY

STUDENTS FACULTY STUDY RESEARCH DEVELOPMENT FUTURE CAREER CREATIVITY COMMUNITY LEADERSHIP TECHNOLOGY FRONTIERS DESIGN ENGINEERING APPLIED SCIENCE GEORGE WASHINGTON UNIVERSITY



INSTITUTE FOR MANAGEMENT SCIENCE AND ENGINEERING
SCHOOL OF ENGINEERING AND APPLIED SCIENCE

LEVEL II

(6) THE GINI INDEX, THE LORENZ CURVE, AND THE TOTAL TIME ON TEST TRANSFORMS.

by

(14) Mahesh Chandra
Nozer D. Singpurwalla

(14) Serial-T-368
(11) 17 February 1978

(9) Scientific rept.,
(12) 30 p.

The George Washington University
School of Engineering and Applied Science
Institute for Management Science and Engineering

(15) Research Sponsored by
Contract N00014-77-C-0263
Project NR 042 372
Office of Naval Research

NONE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER T-368	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THE GINI INDEX, THE LORENZ CURVE, AND THE TOTAL TIME ON TEST TRANSFORMS		5. TYPE OF REPORT & PERIOD COVERED SCIENTIFIC 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) MAHESH CHANDRA NOZER D. SINGPURWALLA		8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0263
9. PERFORMING ORGANIZATION NAME AND ADDRESS THE GEORGE WASHINGTON UNIVERSITY INSTITUTE FOR MANAGEMENT SCIENCE & ENGINEERING WASHINGTON, D. C. 20052		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS OFFICE OF NAVAL RESEARCH CODE 436 ARLINGTON, VA 22217		12. REPORT DATE 17 February 1978 13. NUMBER OF PAGES 26
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) none 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

DISTRIBUTION OF THIS REPORT IS UNLIMITED.

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

18. SUPPLEMENTARY NOTES

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

LORENZ CURVE
GINI INDEX
TOTAL TIME ON TEST TRANSFORMS
MEAN RESIDUAL LIFE
LORENZ PROCESS

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this note we point out the connection between some well-known indicators used in economics and a central concept in reliability theory. In particular, we show that the "Lorenz curve" and the "Gini index" are shown to related to the "total time on test transform" and the "cumulative total time on test transform," respectively. Thus, the recently proposed tests for exponentiality based on the Gini statistic inherit the properties of → next page

(Continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

NONE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

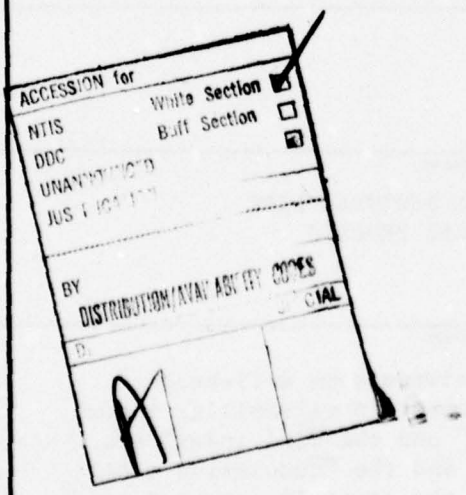
20. Abstract

the tests for exponentiality based on the cumulative total time on test statistic.

Analogous to the "total time on test process" we define the "Lorenz process," and show its weak convergence to functionals of a Brownian motion process. This provides a theory for developing goodness-of-fit tests for any general distribution using the Lorenz curve and the Gini statistic. In addition, we state some new results on the geometry of the Lorenz curve that follow from the geometry of the total time on test transform.

We show that there exists a relationship between the "mean residual life" and the Lorenz curve. This motivates us to propose that the Lorenz curve methods of economic theory also be considered for use in the analysis of failure data.

We hope that this note will help to consolidate and integrate statistical knowledge that has independently evolved in two different areas of application.



THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Institute for Management Science and Engineering

Abstract
of
Serial T-368
17 February 1978

THE GINI INDEX, THE LORENZ CURVE, AND THE TOTAL TIME ON TEST TRANSFORMS

by

Mahesh Chandra
Nozer D. Singpurwalla

In this note we point out the connection between some well-known indicators used in economics and a central concept in reliability theory. In particular, we show that the "Lorenz curve" and the "Gini index" are related to the "total time on test transform" and the "cumulative total time on test transform," respectively. Thus, the recently proposed tests for exponentiality based on the Gini statistic inherit the properties of the tests for exponentiality based on the cumulative total time on test statistic.

Analogous to the "total time on test process," we define the "Lorenz process," and show its weak convergence to functionals of a Brownian motion process. This provides a theory for developing goodness-of-fit tests for any general distribution using the Lorenz curve and the Gini statistic. In addition, we state some new results on the geometry of the Lorenz curve that follow from the geometry of the total time on test transform.

We show that there exists a relationship between the "mean residual life" and the Lorenz curve. This motivates us to propose that the Lorenz curve methods of economic theory also be considered for use in the analysis of failure data.

We hope that this note will help to consolidate and integrate statistical knowledge that has independently evolved in two different areas of application.

THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Institute for Management Science and Engineering

THE GINI INDEX, THE LORENZ CURVE, AND THE TOTAL TIME ON TEST TRANSFORMS

by

Mahesh Chandra
Nozer D. Singpurwalla

1. Introduction

A unifying concept in the statistical theory of reliability and life testing is the "total time on test transform," first discussed by Marshall and Proschan in 1965. Barlow (1968) and Barlow and Doksum (1972) have introduced and studied a scale-free test for exponentiality based on the "cumulative total time on test statistic," which is derived from the total time on test transform. Barlow and Campo (1975), and Barlow (1977) have studied the geometry of the total time on test transform, and have also used it for a graphical analysis of failure data. Langberg, Lêon, and Proschan (1978) provide characterizations of the total time on test transform.

Measures of income inequality used by econometricians are the Lorenz curve and the Gini index (which is derived from the Lorenz curve). The Lorenz curve plots the percentage of total income earned by various portions of the population when the members are ordered by the size of their incomes. Gastwirth (1972) has studied the various properties of the Lorenz curve and the Gini index. Recently, in a series of papers, Gail and Gastwirth (1977a, 1977b) proposed scale-free tests for exponentiality based on the Lorenz curve and the Gini statistic. Among other things, they have shown that tests for exponentiality based on the Lorenz statistic and the Gini statistic are powerful against a variety of alternatives.

Our objective is to demonstrate that there exists a relationship between the total time on test transform and the Lorenz curve, and between the cumulative total time on test transform and the Gini index. Thus the tests for exponentiality proposed by Gail and Gastwirth (1977a, 1977b) inherit some very general properties of the tests for exponentiality based on the cumulative total time on test statistic given by Barlow and Doksum (1972). The relationship mentioned above also prompts us to define what we call the "Lorenz process" and discuss the weak convergence of this process to functionals of a Brownian motion process. Such a result increases interest in the Lorenz curve and the Gini index, since it provides a theory for developing goodness of fit tests for any general distribution using the Lorenz curve and the Gini index.

Gastwirth (1972) has given some properties of the geometry of the Lorenz curve that are of interest to an economist. In this note we present some additional properties of the geometry of the Lorenz curve, and generalize some of Gastwirth's results.

Bryson and Siddiqui (1969) and Hollander and Proschan (1975) have pointed out that the notion of a "mean residual lifetime" is useful for the analysis of biological data. We show that there exists a relationship between the mean residual lifetime and the Lorenz curve. Such a relationship enables us to extend the use of Lorenz curve methods for the analysis and interpretation of failure data. We illustrate our ideas by plotting and interpreting the Lorenz curves of two sets of failure data.

2. Definitions and Notation

Let X be a random variable with distribution F, and let μ be the mean of F; let $F(0^-) \equiv 0$. Then, the *total time on test transform* is defined as

Definition 2.1:

$$H_F^{-1}(t) \stackrel{\text{def}}{=\!=} \int_0^{F^{-1}(t)} \bar{F}(u)du , \quad 0 \leq t \leq 1 ,$$

where $\bar{F}(u) = 1-F(u)$ and $F^{-1}(t)$, the inverse of $F(t)$, is defined by

$$F^{-1}(t) = \inf_x \{x : F(x) \geq t\} .$$

It is easy to verify that $H_F^{-1}(1) = \mu$.

The *scaled total time on test transform* is defined in Definition 2.2:

$$W_F(t) \overset{\text{def}}{=} \frac{H_F^{-1}(t)}{H_F^{-1}(1)} , \quad 0 \leq t \leq 1 .$$

In Figure 2.1, we show a plot of the scaled total time on test transform for a gamma distribution with shape parameters $\alpha = 1$ and 2 , respectively. Other properties of the scaled total time on test transform are discussed by Barlow and Campo (1975).

The *cumulative total time on test transform* is defined in Definition 2.3:

$$V_F \overset{\text{def}}{=} \int_0^1 W_F(u)du = \frac{1}{\mu} \int_0^1 H_F^{-1}(u)du .$$

Thus, the cumulative total time on test transform is simply the area under the scaled total time on test transform.

Gastwirth (1971) has defined the *Lorenz curve* corresponding to a random variable X with distribution F , $F(0^-) = 0$, and mean μ as Definition 2.4:

$$L_F(p) \overset{\text{def}}{=} \frac{1}{\mu} \int_0^p F^{-1}(u)du , \quad 0 \leq p \leq 1 .$$

In econometrics, $L_F(p)$ denotes the fraction of total income that the holders of the lowest pth fraction of incomes possess. In Figure 2.2, we show a plot of the Lorenz curve for a gamma distribution with shape parameters $\alpha = 1$ and 2 , respectively. It is easy to verify that the Lorenz curve is always a convex function of p .

Analogous to Definition 2.3, we define the *cumulative lorenz curve* in Definition 2.5:

$$(CL)_F \overset{\text{def}}{=} \int_0^1 L_F(p)dp = \frac{1}{\mu}\int_0^1 \int_0^p F^{-1}(u)du\, dp \ .$$

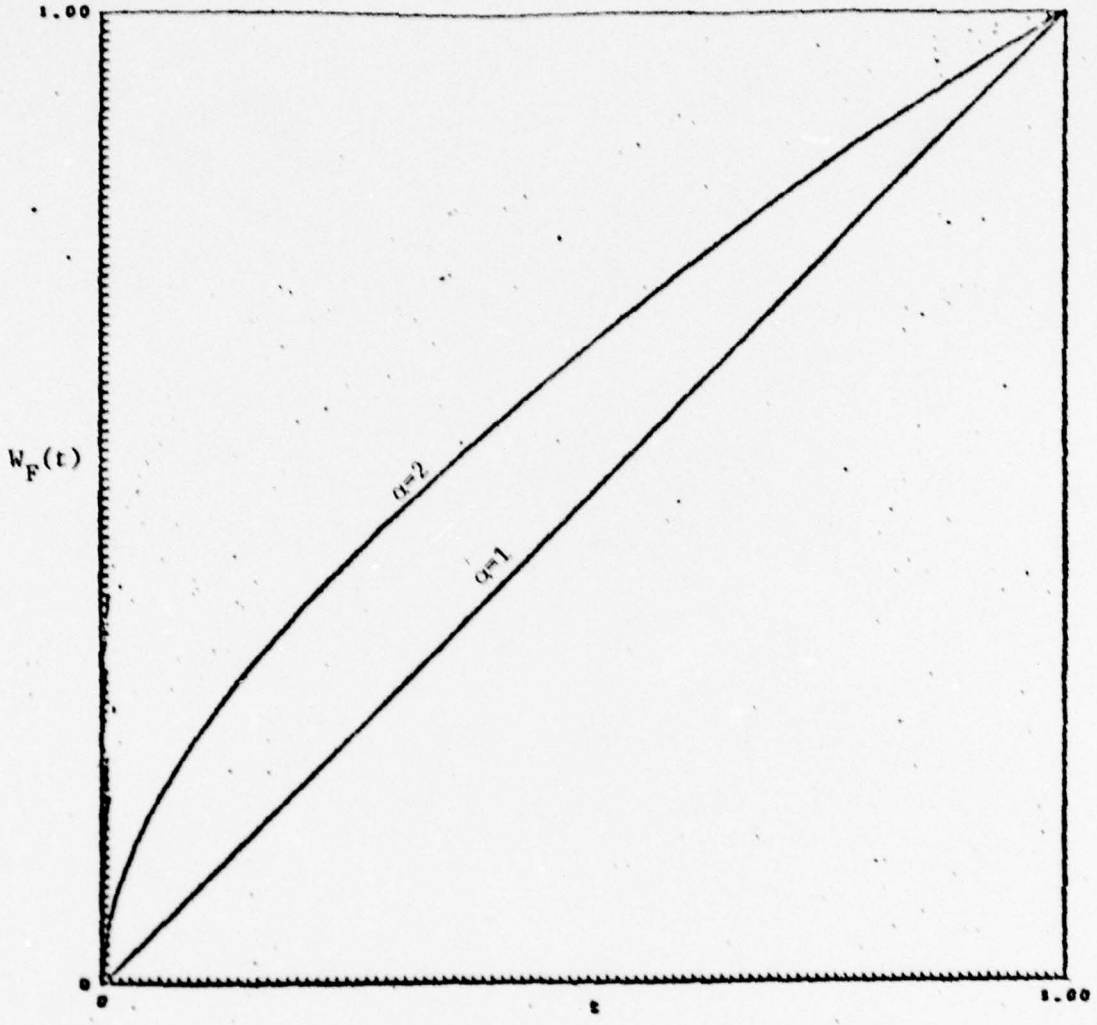


Figure 2.1--Total time on test transforms for gamma distribution

$$F(x) = \int_0^x \frac{\lambda^\alpha u^{\alpha-1} e^{-\lambda u}}{\Gamma\alpha}\, du \quad \text{for} \quad \alpha = 1,2 \ .$$

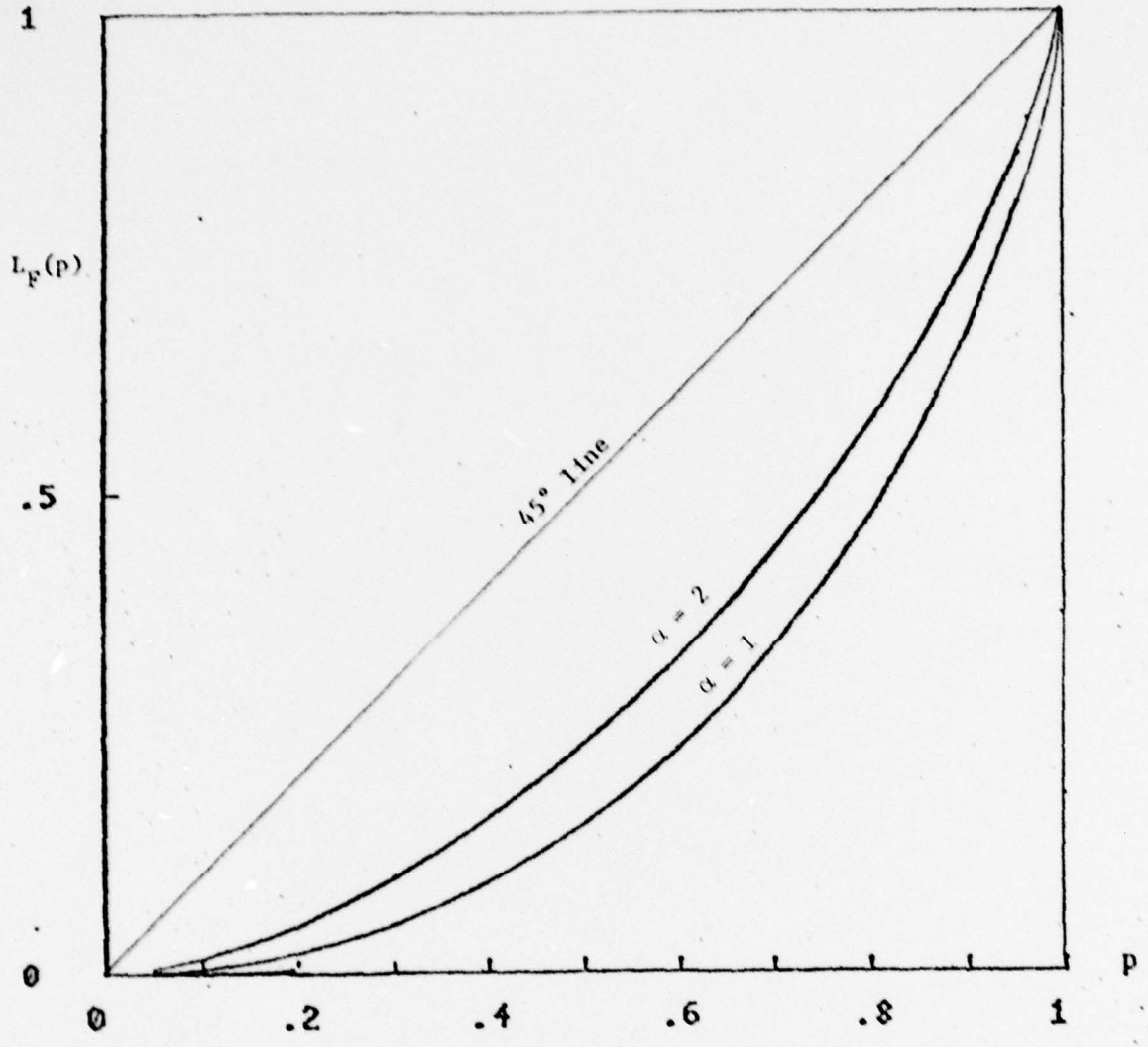


Figure 2.2--Lorenz curves for gamma distribution

$$F(x) = \int_0^x \frac{\lambda^\alpha u^{\alpha-1} e^{-\lambda u}}{\Gamma\alpha}\, du \quad \text{for} \quad \alpha = 1,2 \ .$$

The most common measure of income inequality is the *Gini index*, defined in

Definition 2.6:

$$G_F = \frac{\frac{1}{2} - \int_0^1 L_F(p)dp}{(1/2)} .$$

That is, the Gini index is the ratio of the area between the Lorenz curve $L_F(p)$ and the 45° line, to the area under the 45° line (which is 1/2). The area between the 45° line and $L_F(p)$ is called the *area of concentration*.

3. Some Relationships Among the Concepts of Section 2

We now establish some relationships that exist among some of the concepts introduced in Section 2.

Integrating $W_F(t) = \frac{1}{\mu} \int_0^{F^{-1}(t)} \overline{F}(u)du$ by parts, we have

$$W_F(t) = \frac{1}{\mu} (1-t)F^{-1}(t) + \frac{1}{\mu} \int_0^t F^{-1}(u)du ,$$

or

$$W_F(t) = \frac{1}{\mu} (1-t)F^{-1}(t) + L_F(t) , \quad 0 \le t \le 1 . \tag{3.1}$$

Thus, the scaled total time on test transform is related to the Lorenz curve as shown in (3.1).

Since $V_F = \int_0^1 W_F(t)dt$,

$$V_F = \frac{1}{\mu} \int_0^1 (1-t)F^{-1}(t)dt + \frac{1}{\mu} \int_0^1 \int_0^t F^{-1}(u)du\, dt .$$

Integrating by parts, we can verify that

$$\int_0^1 \int_0^t F^{-1}(u)\,du\,dt = \int_0^1 (1-t)F^{-1}(t)\,dt .$$

Thus

$$V_F = \frac{2}{\mu}\int_0^1 (1-t)F^{-1}(t)\,dt = \frac{2}{\mu}\int_0^1 \int_0^t F^{-1}(u)\,du\,dt ,$$

or

$$V_F = 2(CL)_F . \tag{3.2}$$

Thus the cumulative total time on test transform is twice the cumulative Lorenz curve.

In order to see the relationship between the Gini index and the cumulative total time on test transform, we note that

$$G_F = 2\left[\frac{1}{2} - \int_0^1 L_F(p)\,dp\right]$$

$$= 1 - 2\int_0^1 \frac{1}{\mu}\int_0^p F^{-1}(u)\,du\,dp ,$$

or

$$G_F = 1 - V_F . \tag{3.3}$$

Thus the Gini index is simply one minus the cumulative total time on test transform.

Relationships (3.1), (3.2), and (3.3) now enable us to state some other results for the Lorenz curve and the Gini index.

4. Some Properties of the Lorenz Curve and the Gini Index

Gastwirth (1972) has given several properties of the Lorenz curve and the Gini index that are of interest. We give here some additional properties which follow naturally from the results of the previous section.

Remark 4.1: L_F^{-1}, the inverse of L_F, is a distribution function with support on $[0,1]$; also L_F^{-1} is concave.

Proof: The conclusion follows from the fact that

$$L_F(1) = \frac{1}{\mu}\int_0^1 F^{-1}(p)dp = 1$$

when $F(0^-) \equiv 0$, and that $L_F^{-1}(p)$ increases in $p \in [0,1]$. Since L_F is convex, L_F^{-1} is concave.

We shall make use of Remark 4.1 in Theorem 4.6.

Definition 4.2: Let F be the class of continuous distributions on $[0,\infty)$, and let $\{deg\}$ be the class of degenerate distributions. Then F_1 *is star ordered with respect to* F_2, denoted by $F_1 \lesssim_* F_2$, if $F_1, F_2 \in F \cup \{deg\}$, and $\dfrac{F_2^{-1}F_1(x)}{x}$ is nondecreasing in x for $0 \leq x \leq F_1^{-1}(1)$.

We shall now state and prove

Theorem 4.3: If $F_1 \lesssim_* F_2$, and if $\int_0^\infty x dF_1(x) = \int_0^\infty x dF_2(x) = \mu$, then

(a) $L_{F_1}(p) \geq L_{F_2}(p)$

(b) $(CL)_{F_1} \geq (CL)_{F_2}$, and

(c) $G_{F_1} \leq G_{F_2}$.

Proof: Consider $L_{F_1}(p) - L_{F_2}(p) = \int_0^p \frac{1}{\mu}\left(F_1^{-1}(u) - F_2^{-1}(u)\right)du$; let $h(u) = F_1^{-1}(u) - F_2^{-1}(u)$, and note that $\int_0^1 h(u)du = 0$. Since $F_1 \lesssim_* F_2$, by the "single crossing property" of star ordered distributions [cf. Barlow and Proschan (1975), p. 107], it follows that $h(u)$ changes sign exactly once, and from positive to negative values. Thus,

$$\frac{1}{\mu}\int_0^p h(u)du \geq 0 ,$$

and this completes part (a) of the theorem. Proof of parts (b) and (c) follow from the above result and the definitions of $(CL)_F$ and G_F.

If $F_1 \underset{*}{\leq} F_2$, and if F_2 is taken to be an exponential distribution, then F_1 belongs to the class of distributions which have "increasing failure rate average" [cf. Barlow and Proschan (1975)]. Theorem 7 of Gastwirth (1972) is analogous to Theorem 4.3 of this paper. However, our theorem is more general than that of Gastwirth, since it applies to a much larger class of distributions.

Definition 4.4: Let F be the class of continuous distributions on $[0,\infty)$, and let $\{deg\}$ be the class of degenerate distributions. Then F_1 is *convex ordered with respect to* F_2, denoted by $F_1 \underset{c}{\leq} F_2$, if $F_1, F_2 \in F \cup \{deg\}$, and $F_2^{-1}F_1(x)$ is convex in x for $0 \leq x \leq F_1^{-1}(1)$.

Remark 4.5: $F_1 \underset{c}{\leq} F_2$ implies $F_1 \underset{*}{\leq} F_2$ [cf. Barlow and Proschan (1975), p. 107].

In the following theorem we shall show that the convex ordering property is preserved by L_F^{-1}, the inverse of L_F. If $F_1 \underset{c}{\leq} F_2$, and if F_2 is taken to be an exponential distribution, then F_1 belongs to the class of distributions which has an "increasing failure rate" [cf. Barlow and Proschan (1975)].

Theorem 4.6: If $F_1 \underset{c}{\leq} F_2$ then $L_{F_1}^{-1} \underset{c}{\leq} L_{F_2}^{-1}$.

Proof: We wish to show that $L_{F_2} L_{F_1}^{-1}(x)$ is convex in $0 \leq x \leq F_1^{-1}(1)$. We shall assume that F_1 and F_2 are absolutely continuous. Then we need only show that $\frac{d}{dx} L_{F_2}\left[L_{F_1}^{-1}(x)\right]$ is nondecreasing in $0 \leq x \leq F_1^{-1}(1)$.

Let μ_1 and μ_2 be the means of F_1 and F_2, respectively. Then

$$\frac{d}{dx} L_{F_2}\left[L_{F_1}^{-1}(x)\right] = \frac{d}{dx}\frac{1}{\mu_2}\int_0^{L_{F_1}^{-1}(x)} F_2^{-1}(u)\,du$$

$$= \frac{F_2^{-1}\left(L_{F_1}^{-1}(x)\right)}{\mu_2}\,\frac{dL_{F_1}^{-1}(x)}{dx}\,.$$

Let

$$x = L_{F_1}(p)$$

$$\frac{dx}{dp} = \frac{F_1^{-1}(p)}{\mu_1}$$

$$\frac{dp}{dx} = \frac{\mu_1}{F_1^{-1}(p)}\,\Bigg|\, p = L_{F_1}^{-1}(x) = \frac{\mu_1}{F_1^{-1}\left[L_{F_1}^{-1}(x)\right]}\,.$$

Hence

$$\frac{dL_{F_1}^{-1}(x)}{dx} = \frac{\mu_1}{F_1^{-1}\left[L_{F_1}^{-1}(x)\right]}$$

and

$$\frac{d}{dx}\frac{1}{\mu_2}\int_0^{L_{F_1}^{-1}(x)} F_2^{-1}(u)\,du = \frac{\mu_1 F_2^{-1}\left[L_{F_1}^{-1}(x)\right]}{\mu_2 F_1^{-1}\left[L_{F_1}^{-1}(x)\right]}\,.$$

Since $F_1 \underset{c}{\leq} F_2$ implies that $\dfrac{F_2^{-1}F_1(x)}{x}$ is nondecreasing in $0 \leq x \leq F_1^{-1}(1)$, and since $F_1^{-1}L_{F_1}^{-1}(x) = t$ is nondecreasing in $0 \leq x \leq F_1^{-1}(1)$, a change of

variable shows that $\frac{d}{dx}\frac{1}{\mu_2}\int_0^{L_{F_1}^{-1}(x)} F_2^{-1}(u)du$ is nondecreasing in $0 \leq x \leq F^{-1}(1)$. Since continuous distributions can be approximated arbitrarily closely by absolutely continuous distributions, the proof of the theorem is completed.

5. Some Statistics of Interest

Let $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ denote the order statistics corresponding to a random sample of size n from a distribution F, where $F(0^-) = 0$. The *total time on test statistic* to the ith failure, $T\left(X_{(i)}\right)$, is defined by

$$T\left(X_{(i)}\right) \stackrel{\text{def}}{=} \sum_{j=1}^{i} (n-j+1)\left(X_{(j)} - X_{(j-1)}\right) . \tag{5.1}$$

Barlow and Campo (1975) have used the *scaled total time on test statistic*, $W\left(\frac{i}{n}\right)$, defined as

$$W\left(\frac{i}{n}\right) \stackrel{\text{def}}{=} \frac{\sum_{j=1}^{i} (n-j+1)\left(X_{(j)} - X_{(j-1)}\right)}{\sum_{j=1}^{n} X_{(j)}} \tag{5.2}$$

for analyzing failure data.

The *cumulative total time on test statistic*, V_n, defined as

$$V_n \stackrel{\text{def}}{=} \frac{1}{n-1}\sum_{i=1}^{n-1} W\left(\frac{i}{n}\right) ,$$

has been used by Barlow and Doksum (1972) for testing for exponentiality. They show that a test based on V_n is asymptotically minimax against a class of alternatives defined by the Kolmogorov distance.

Gail and Gastwirth (1977a) have proposed a test for exponentiality based on the *Lorenz statistic*, $L_n(p)$, defined as

$$L_n(p) \overset{\text{def}}{=} \sum_{i=1}^{[np]} X_{(i)} \Big/ \sum_{i=1}^{n} X_{(i)} ,$$

where $0 < p < 1$, and $[np]$ is the largest integer in np. The statistic $L_n(\cdot 5)$ is shown to have good power against a range of alternatives; this is based on a Monte Carlo investigation.

Recently, Gail and Gastwirth (1977) proposed another test for exponentiality based on the *Gini statistic*, G_n, defined as

$$G_n \overset{\text{def}}{=} \frac{\sum_{i=1}^{n-1} i(n-i)\left(X_{(i+1)} - X_{(i)}\right)}{(n-1)\sum_{i=1}^{n} X_{(i)}} . \tag{5.5}$$

Based upon Monte Carlo studies, Gail and Gastwirth (1977b) have concluded that G_n is *more powerful* than $L_n(\cdot 5)$ for $n=20$, against most of the alternatives that are studied.

We can easily verify the following relationships between the various test statistics that we have discussed thus far:

$$W\left(\frac{i}{n}\right) = L_n\left(\frac{i}{n}\right) + \frac{(n-i)X_{(i)}}{\sum_{j=1}^{n} X_{(j)}} \tag{5.6}$$

and

$$V_n = 1 - G_n . \tag{5.7}$$

In view of (5.7) above, *the test for exponentiality based on the Gini statistic is identical to the test for exponentiality based on the cumulative total time on test statistic.* Thus, we can say that the test for exponentiality based on the Gini statistic is asymptotically minimax against some restricted alternatives.

The exact distribution of G_n under exponentiality follows from Theorem 6.2 of Barlow, Bartholomew, Bremner and Brunk (1972), and from Equation (5.7) above. Gail and Gastwirth (1977b) have also derived the exact distribution of G_n, but by using a different argument.

6. The Lorenz Process and its Weak Convergence

Using previous notation, we define the *Lorenz process*, $\{\mathcal{L}_n(t) ; 0 \leq t \leq 1\}$, as

$$\mathcal{L}_n(t) = \sqrt{n}\left\{L_n\left(\frac{i}{n}\right) - L_F(t)\right\}, \quad \frac{i-1}{n} < t \leq \frac{i}{n}$$

$$1 \leq i \leq n ; \qquad (6.1)$$

$$\mathcal{L}_n(0) = 0 .$$

We are interested in the asymptotic behavior of this process for F in general.

Following the notation and terminology of Barlow and Campo (1975), we shall say that a stochastic process $\{W(t) ; t \geq 0\}$ is a "Brownian motion process" with drift coefficient equal to 0 if:

(i) $W(0) = 0$;

(ii) $\{W(t), t \geq 0\}$ has stationary independent increments;

(iii) $W(t)$ is normally distributed with mean 0 and variance t, for all $t > 0$.

A process $\{U(t) ; 0 \leq t \leq 1\}$ is called a "Brownian Bridge" on $[0,1]$ when $U(t) = W(t) - tW(1)$, $0 \leq t \leq 1$. Such a process is normal, has all sample paths continuous, $E(U(t)) = 0$ for $0 \leq t \leq 1$, and has covariance $s(1-t)$ for $0 \leq s \leq t \leq 1$. Note that $V(t) = -U(t)$ is also a Brownian Bridge on $[0,1]$.

6.1 Weak convergence of the Lorenz process

Let $\nu_n(u)$ be a discrete measure putting mass $1/n$ at $u = i/n$, $i=1,2,\ldots,n$. Then

$$L_n\left(\frac{i}{n}\right) = \frac{\frac{1}{n}\sum_{j=1}^{i} X_{(j)}}{\frac{1}{n}\sum_{j=1}^{n} X_{(j)}} = \int_0^{i/n} \frac{X_{[nu]}}{\bar{X}}\, d\nu_n(u) .$$

where $\bar{X} \overset{\text{def}}{=} \sum_{j=1}^{n} X_{(j)}/n$, and $[nu]$ is the greatest integer in nu .

Since $L_F(t) = \frac{1}{\mu}\int_0^t F^{-1}(u)du$, substitution in (6.1) gives

$$\mathcal{L}_n\left(\frac{i}{n}\right) = \int_0^{i/n} \sqrt{n}\left\{\frac{X_{[nu]}}{\bar{X}} - \frac{F^{-1}(u)}{\mu}\right\} d\nu_n(u) + \int_0^{i/n} \sqrt{n}\,\frac{F^{-1}(u)}{\mu}\, d\left[\nu_n(u) - u\right] . \tag{6.2}$$

If we assume that $\int_0^\infty x\,dF(x) \leq \infty$, and if $g = F^{-1}$ has a nonzero continuous derivative g' on $(0,1)$, then by Shorack (1972),

$$\sqrt{n}\left\{\frac{X_{[nu]}}{\bar{X}} - \frac{F^{-1}(t)}{\mu}\right\} \underset{\substack{n\to\infty \\ [nu]\to t}}{\xrightarrow{P}} -\frac{g'(t)}{\mu}U(t) - \frac{g(t)}{\mu^2}Z .$$

In the expression above "$\xrightarrow{P}$" denotes convergence in probability, U is the Brownian Bridge process on $(0,1)$, and $Z = \int_0^\infty U[F(x)]dx$ is normal with mean 0 and variance σ_F^2 , where σ_F^2 is the variance of F .

Since the second term of (6.2) converges deterministically to zero, it follows that

$$\mathcal{L}_n(t) \underset{n\to\infty}{\xrightarrow{P}} \int_0^t -\left(\frac{g'(u)}{\mu}U(u) + \frac{g(u)}{\mu^2}Z\right)du \overset{\text{def}}{=} \mathcal{L}(t) .$$

We can also express $\mathcal{L}(t)$ as

$$\mathcal{L}(t) = -\frac{1}{\mu}\int_0^{F^{-1}(t)} U[F(x)]dx - \frac{L_F(t)}{\mu}\int_0^{\infty} U[F(x)]dx .$$

By a direct but tedious calculation [cf. Gail (1977)], it can be shown that under exponentiality

$$\mathrm{Var}\big(\mathcal{L}(t)\big) = 2(1-t)\ell n(1-t) + t + t(1-t) - \big(t + (1-t)\ell n(1-t)\big)^2 .$$

Thus, in contrast to an analogous result based on the convergence of the total time on test process [see Barlow and Campo (1975)], under exponentiality $\{\mathcal{L}(t) ; 0 \le t \le 1\}$ is *not* the Brownian Bridge.

6.2 Uses of the Lorenz process

The Lorenz process can be used to find the asymptotic distribution of

$$\sup_{1\le i\le n} \sqrt{n}\left|L_n\left(\frac{i}{n}\right) - L_F\left(\frac{i}{n}\right)\right| ,$$

which by the invariance principle of Billingsley (1968) is the same as that of

$$\sup_{0\le t\le 1} |\mathcal{L}(t)| .$$

This statistic can be used to test the hypothesis that the given data has distribution F versus the general alternative that it does not. As seen at the end of Section 6.1, under exponentiality it is *not* the Kolmogorov-Smirnov statistic, and this is not very pleasing.

Another statistic that can be used for the same purpose is the area between the curve of $L_n\left(\frac{i}{n}\right)$ and the curve of $L_F(t)$. A consideration of this area leads us to Theorem 6.2, which follows from Theorem 6.6 of Barlow et al. (1972).

Theorem 6.1 [Barlow, Bartholomew, Bremner and Brunk (1972)]: If

$$\int_0^\infty x dF(x) < \infty \quad \text{and} \quad \sigma^2(F) < \infty ,$$

where

$$\sigma^2(F) = 2 \int\int_{s<t} \{2[1-F(s)] - V_F\} \times \{2[1-F(t)] - V_F\} F(s)\big(1-F(t)\big)\, ds\, dt ,$$

then

$$\sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^{n-1} W\left(\frac{i}{n}\right) - V_F \right\} \xrightarrow[n\to\infty]{\mathcal{D}} N\left(0, \frac{\sigma^2(F)}{\mu^2}\right) .$$

where "$\xrightarrow{\mathcal{D}}$" denotes convergence in distribution.

Using the fact that $G_F = 1-V_F$, and $G_n = 1-V_n$, we are now in a position to obtain the limiting distribution of the Gini statistic. We have

Theorem 6.2: Under the conditions of Theorem 6.1,

$$\sqrt{n}\left(G_n - G_F\right) \xrightarrow[n\to\infty]{\mathcal{D}} N\left(0, \frac{\sigma^2(F)}{\mu^2}\right) .$$

In the case of F exponential, $G_F = \frac{1}{2}$ and $\sigma^2(F) = \frac{1}{12}$. Thus

$$\sqrt{12n}\left(G_n - \frac{1}{2}\right) \xrightarrow[n\to\infty]{\mathcal{D}} N(0,1) ,$$

a result also obtained by Gail and Gastwirth (1977b) using some arguments due to Hoeffding (1949).

7. The Lorenz Curve and the Mean Residual Lifetime

Bryson and Siddiqui (1969) and also Hollander and Proschan (1975) have pointed out that the notion of "mean residual lifetime" is especially useful for the analysis of biological data. In this section we point out the relationship between the Lorenz curve and the mean residual lifetime.

Such a relationship suggests to us that the Lorenz curve methods, which have so far been mainly used in the social sciences, could also be used in the biological sciences. This possibility has also been hinted at by Thompson (1976).

The *mean residual lifetime* corresponding to a random variable X with distribution F, $F(0^-) \equiv 0$, is defined in

Definition 7.1:

$$\varepsilon_F(x) = \frac{\int_x^\infty \bar{F}(u)du}{\bar{F}(x)} .$$

We say that a distribution F has a *decreasing (increasing) mean residual lifetime* if $\varepsilon_F(x)$ is decreasing (increasing) in x for all $x \geq 0$.

Bryson and Siddiqui (1969) have used the decreasing mean residual lifetime property to interpret some survival data on patients suffering from leukemia.

If we denote the mean of F by μ, then we can write $\varepsilon_F(x)$ as

$$\varepsilon_F(x) = \frac{\mu\left[1 - \frac{1}{\mu}\int_0^x \bar{F}(u)du\right]}{\bar{F}(x)} .$$

From Definition 2.2, it follows that

$$\frac{1}{\mu}\int_0^x \bar{F}(u)du = W_F(F(x)) ;$$

thus

$$\varepsilon_F(x) = \frac{\mu\left[1 - W_F(F(x))\right]}{\bar{F}(x)} .$$

The above expression when used with Equation (3.1) gives us a relationship between $\varepsilon_F(x)$ and $L_F(\cdot)$; specifically, we have

$$L_F(F(x)) = 1 - \frac{\overline{F}(x)}{\mu} [\varepsilon_F(x) + x] . \qquad (7.1)$$

In order to demonstrate the use of the Lorenz curve for biological applications, we shall consider the data given by Bryson and Siddiqui (1969). These data pertain to survival times (in days), from the time of diagnosis, of patients suffering from chronic granulocytic leukemia. The ordered 43 survival times in days are: 7, 47, 58, 74, 177, 232, 273, 285, 317, 429, 440, 445, 455, 468, 495, 497, 532, 571, 579, 581, 650, 702, 715, 779, 881, 900, 930, 968, 1077, 1109, 1314, 1334, 1367, 1534, 1712, 1784, 1877, 1886, 2045, 2056, 2260, 2429, 2509.

If we denote the number of survivors at time x by S, and if the size of the initial population is denoted by n, then Bryson and Siddiqui estimate the mean residual life at time x by

$$\hat{\varepsilon}(x) = S^{-1} \Sigma (X_j - x) ,$$

where X_j denotes the survival time of the jth element and the sum is for those having survived up to time x.

In Figure 7.1 we show a plot of $\hat{\varepsilon}(x)$ versus the time x, for the data in question. Thus, the distribution of survival times has a decreasing mean residual life; this conclusion is based upon an inspection of Figure 7.1.

In Figure 7.2 we give a plot of the sample Lorenz curve for these data. The sample Lorenz curve is simply a plot of the Lorenz statistic $L_n(p)$ (defined in Section 5) versus p, $0 < p < 1$. The sample Lorenz curve $L_n(p)$ represents the proportion of the total lifetime contributed by the least fortunate $p \cdot 100$ percent of the patients; for example, 50% of the patients contribute only 20% of the total lifetime. The sample Lorenz curve can also be used to compare the heterogeneity of the survival patterns of two groups of patients. To illustrate this, we give in Figure 7.3 the Lorenz curves for the data on the survival times of guinea pigs considered by Doksum (1974). The Lorenz curve for the "control

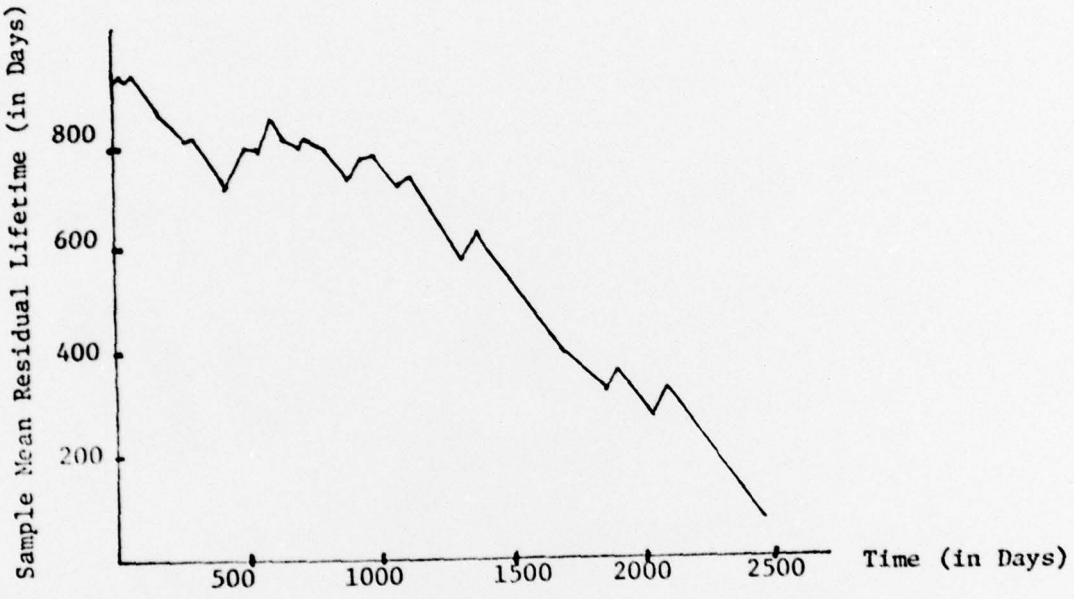


Figure 7.1--Sample mean residual lifetime versus time of leukemia patients

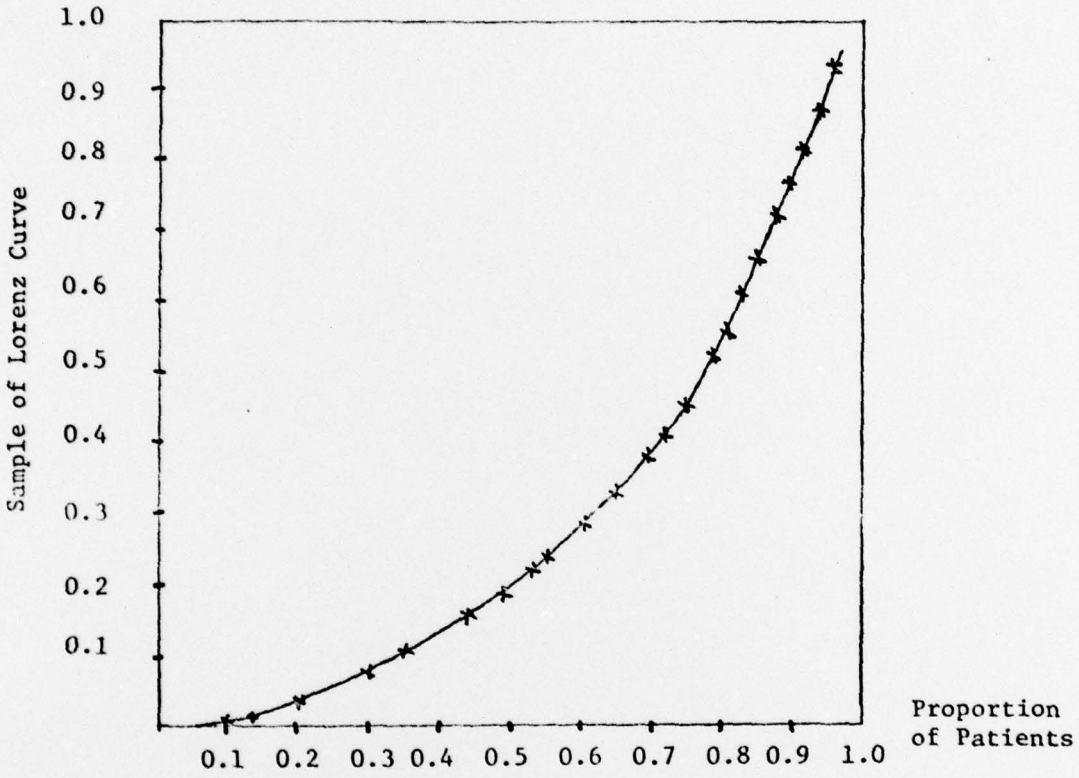


Figure 7.2--Sample Lorenz curve versus proportion of leukemia patients

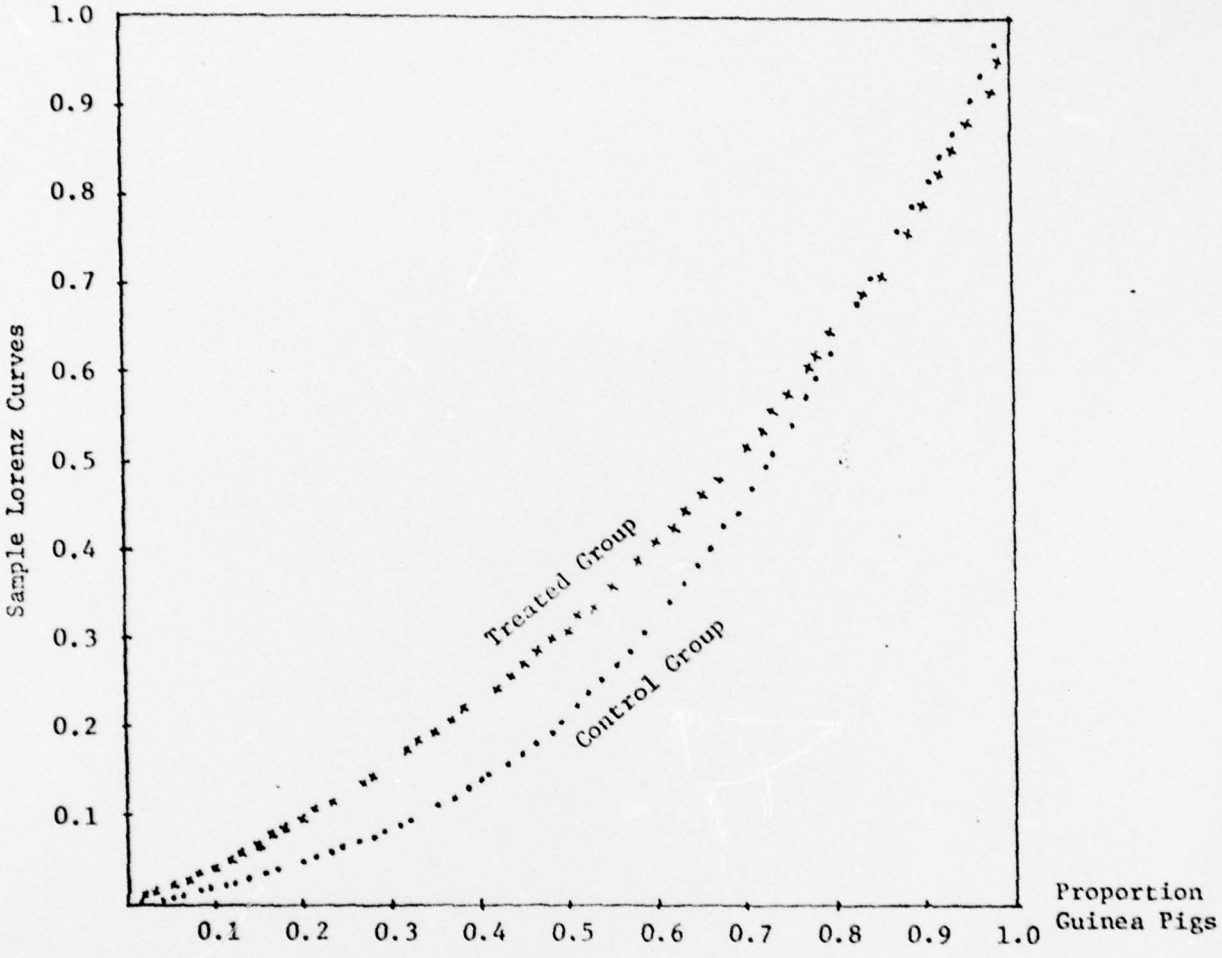


Figure 7.3--Sample Lorenz curves versus proportion of guinea pigs

group" lies below the Lorenz curve for the "treatment group" for p less than about .8 ; the curves cross near $p \approx .8$. Thus, initially the treatment group is less heterogeneous than the control group and the reverse is true later on. This can also be verified by an inspection of the actual data.

ACKNOWLEDGMENTS

We are grateful to Joseph L. Gastwirth and Mitchell Gail, who were kind enough to provide us with a preprint of Gail and Gastwirth (1977b). Several comments by Frank Proschan are also gratefully acknowledged.

REFERENCES

BARLOW, R. E. (1968). Likelihood ratio tests for restricted families of probability distributions. *Ann. Math. Statist.* 39 547-560.

BARLOW, R. E. (1977). Geometry of the total time on test transform. Operations Research Center Report ORC 77-11, University of California, Berkeley.

BARLOW, R. E., D. J. BARTHOLOMEW, J. M. BREMNER and H. D. BRUNK (1972). *Statistical Inference under Order Restrictions*. John Wiley and Sons, Inc., New York.

BARLOW, R. E. and R. CAMPO (1975). Total time on test process and applications to failure data analysis. *Reliability and Fault Tree Analysis* (R. E. Barlow, J. Fussell, and N. D. Singpurwalla, eds.). Conference volume published by SIAM, Philadelphia.

BARLOW, R. E. and K. DOKSUM (1972). Isotonic tests for convex ordering. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* 1 293-323.

BARLOW, R. E. and F. PROSCHAN (1969). A note on a test for monotone failure rate based on incomplete data. *Ann. Math. Statist.* 40 595-600.

BARLOW, R. E. and F. PROSCHAN (1975). *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart, and Winston, New York.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. John Wiley and Sons, Inc., New York.

BRYSON, M. C. and M. M. SIDDIQUI (1969). Some criteria for ageing. *J. Amer. Statist. Assoc.* 64 1472-1485.

DOKSUM, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Ann. Statist.* 2 267-277.

GAIL, M. H. (1977). A scale-free goodness-of-fit test for the exponential distribution based on the Lorenz curve. Ph.D. Dissertation, The George Washington University.

GAIL, M. H. and J. L. GASTWIRTH (1977a). A scale-free goodness-of-fit test for the exponential distribution based on the Lorenz curve. (Submitted for publication.)

GAIL, M. H. and J. L. GASTWIRTH (1977b). A scale-free goodness-of-fit test for the exponential distribution based on the Gini statistic. (Submitted for publication.)

GASTWIRTH, J. L. (1971). A general definition of the Lorenz curve. *Econometrica* 39 1037-1039.

GASTWIRTH, J. L. (1972). The estimation of the Lorenz curve and Gini index. *Rev. Economics and Statist.* 54 306-316.

HOEFFDING, W. (1949). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* 19 293-325.

HOLLANDER, M. and F. PROSCHAN (1975). Tests for the mean residual life. *Biometrika* 62 585-593.

LANGBERG, N., R. LEON, and F. PROSCHAN (1978). To appear.

MARSHALL, A. W. and F. PROSCHAN (1965). Maximum likelihood estimation for distributions with monotone failure rate. *Ann. Math. Statist.* 36 69-77.

SHORACK, G. (1972). Convergence of quantile and spacing process with applications. *Ann. Math. Statist.* 43 1400-1411.

THOMPSON, W. A., Jr. (1976). Fisherman's luck. *Biometrics* 32 265-271.