Working Paper 91-12
March 1991

Departamento de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Madrid)

A QUANTILE APPROACH TO THE BOX-COX TRANSFORMATION IN RANDOM SAMPLES

Santiago Velilla*

Abstract

This paper presents an alternative approach to the likelihood methods for estimating the parameter λ in the Box-Cox family of transformations when the data arise from a random sample. The method is based on a representation of the quantile function of the variable under consideration. Theoretical properties of the method, its practical applications and comparison with the likelihood approach are studied.

Key Words.
Asymptotic relative efficiency (ARE); Box-Cox transformation; influential observations; Quantile function; Kernel density estimation.

* Departamento de Economía, Universidad Carlos III de Madrid y Departamento de Estadística, Universidad Complutense. 28040-Madrid.

1. Introduction. Let X be a random variable with unknown distribution function F. Let $\{g(.,\lambda)\}$ a family of transformations indexed by the parameter $\lambda\in\Lambda$, where Λ is a non empty set of $\mathbb{R}^m$. A method for modelling F is to suppose that, for some unknown $\lambda\in\Lambda$, $g(X,\lambda)\sim N(\mu,\sigma^2)$. When m=1, a common family of transformations is the family of Box-Cox (1964):

$$X^{(\lambda)}=\begin{cases}\dfrac{X^{\lambda}-1}{\lambda}, & \lambda\neq 0\\ \log X, & \lambda=0.\end{cases} \tag{1.1}$$

In (1.1), X must be positive. If not, X is replaced by X+c (X+c>0). The model is then,

$$X^{(\lambda)}\sim N(\mu,\sigma^2). \tag{1.2}$$

Given $X_1, \ldots, X_n$ a random sample from F, λ is usually estimated by likelihood methods. In this paper I present, following Parzen (1979) suggestions, an alternative approach for estimating λ based on the quantile function of X under (1.2). Section 2 introduces some necessary background and presents some motivation. Section 3 contains the new method and studies its properties. Section 4 compares the new method with the likelihood approach and another quantile method due to Hinkley (1975). Section 5 is devoted to examples. Section 6 contains some final comments and remarks.

2. Background and motivation. The quantile function of **X** is defined to be

$$Q(u)=\inf\{x:\ F(x)\geq u\} \qquad 0<u<1. \tag{2.1}$$

The reader is assumed to have knowledge of the main properties related to (2.1) (see Parzen(1979, sec. 2, 3 and 4); Serfling(1980, sec. 1.1.4.) and Reiss(1989, chap. I) for references). Let $X_1, \ldots, X_n$ a random sample from F. A sample version of the quantile function is obtained by substituting, in (2.1), F by F_n, the empirical distribution function of the sample. This yields

$$Q_n(u)=\inf\{x: F_n(x)\geq u\} \qquad 0<u<1. \qquad (2.2)$$

As defined by (2.2), we have, $Q_n(u)=X_{(j)}$, $(j-1)/n<u\leq j/n$ $(1\leq j\leq n)$.

It is well-known that the transformations in the Box-Cox family are increasing and continuous for all λ. Therefore, under the model (2) (and $X>0$) $\mu+\sigma\phi^{-1}(u) = [Q(u)]^{(\lambda)}$, where $\phi^{-1}(u)$ is the inverse of the distribution function of N(0,1). Given that $([Q(u)]^{(\lambda)})'=q(u)[Q(u)]^{\lambda-1}$ (all λ), where $q(u)=Q'(u)$, we get $f[Q(u)]= (1/\sigma)[Q(u)]^{\lambda-1} \varphi[\phi^{-1}(u)]$, where φ is the density of the standard normal distribution function. Taking logs results in the following relation

$$\log \frac{f[Q(u)]}{\varphi[\phi^{-1}(u)]} = -\log\sigma + (\lambda-1)\log Q(u). \qquad (2.3)$$

Formula (2.3) is a slight modification of a result in Parzen (1979, sec. 12).

(2.3) suggests that if we substitute Q(.) by $Q_n(.)$ and then we put $u=j/n$ for $1\leq j\leq n$ the following approximate relationship among the quantities $U_j=\log\{f[X_{(j)}]/\varphi[\phi^{-1}(j/n+1)]\}$ and $V_j=\log X_{(j)}$, $j=1, \ldots, n$, holds:

$$U_j\approx-\log\sigma+(\lambda-1)V_j, \qquad (2.4)$$

for $j=1, \ldots, n$. Two comments arise from (2.4): (*i*) The adecuacy of the power transformation to attain normality should be indicated by a linear trend in a scatter plot of (V_j, U_j). *(ii)* Estimating λ is equivalent to estimating the slope in a simple linear regression model.

3. The method and its properties. Take $\beta=\lambda-1$ in (2.4). Set, for $j=1, \ldots, n$, $W_j=\log[f(X_j)]$ and $a_j=\log\varphi[\phi^{-1}(j/n+1)]=J(j/n+1)$, where $J(t)=-\frac{1}{2}\log 2\pi - \frac{1}{2}[\phi^{-1}(t)]^2$. Define $D_n=\sum_{j=1}^{n}(V_j-\bar{V})^2$, and $L_n=(1/n)\sum_{j=1}^{n}a_j\log X_{(j)}$. The least-squares "estimate" of β is a random variable of the form $\beta_n=(A_n-1)+B_n$, where $A_n-1=(\sum_{j=1}^{n}V_jW_j-n\overline{VW})/D_n$ and $B_n=(1/D_n)[-nL_n+n\bar{a}\,\bar{V}]$. A natural "estimator" for λ would then be $\beta_n+1=A_n+B_n$. However, for reasons which will appear later, it is convenient to take

$$\lambda_n=A_n$$

as the building block for constructing a new estimator for λ.

Theoretical results. The method is based on the following

THEOREM. Let $X_1, \ldots, X_n$ i.i.d. as a distribution function F with density f. Suppose that $P[X\geq 0]=1$ and that $E[|\log X|^4]$ and $E[|\log f(X)|^4]$ are both finite. The following results hold:

$$\text{(i)}\ L_n \longrightarrow \int_0^1[\log Q(t)]J(t)dt,\ \text{a.e.},$$

$$\text{(ii)}\ D_n/n \longrightarrow \operatorname{var}[\log(X_1)],\ \text{a.e.},$$

and

$$\text{(iii)}\ A_n \sim AN(m, s^2/n),\ \text{for certain m and s.}$$

PROOF. (i) Note first that by Cauchy-Schwartz inequality, if N is $U(0,1)$, $|\int_0^1[\log Q(t)]J(t)dt| \leq E^{1/2}[|\log X|^2]E^{1/2}[|J(N)|^2]<\infty$, since $Q(N)\sim X$ and $\phi^{-1}(N)\sim N(0,1)$. If $h(t)=\log t$, then, $L_n=\frac{1}{n}\sum_{j=1}^{n}J(j/n+1)h[X_{(j)}]$, so

L_n is written in a L-estimate form. If (X_n) are i.i.d. a set of sufficient conditions for (i) to hold is given in Serfling (1980, p. 277 and 279). All the conditions are easy to check in this situation and to see that $|J(t)|\le M[t(1-t]^{-1+(1/r)+\delta}$, given that $J(t)=J(1-t)$, it is enough to proof that $|\phi^{-1}(t)|^2=O([t(1-t)]^{-1+(1/r)+\delta})$, as $t\longrightarrow 1$. In the well known inequality for $x>0, \int_x^\infty e^{-(1/2)y^2}dy \le (1/x)e^{-(x^2/2)}$, put $x=\phi^{-1}(t)\,(t>1/2)$, $r=4$, and $0<\delta<1-(1/r)=3/4$, to obtain, $[t(1-t)]^{1-(1/r)-\delta}|\phi^{-1}(t)|^2\le[\phi^{-1}(t)]^p\exp\{-(q/2)[\phi^{-1}(t)]^2\}$. In the latter expression, $p=1+(1/r)+\delta$ and $q=1-(1/r+\delta)$ are both positive.

(ii) This is a direct consequence of the strong law of the large numbers. (iii) Simple algebra shows that $A_n=g(Z_n)$, where $Z=(\overline{V},\overline{W},(1/n)\sum_{j=1}^{n}V_j^2,(1/n)\sum_{j=1}^{n}V_jW_j)$ and $g(a,b,c,d)=(d-ab+c-a^2)/(c-a^2)$. It's clear that $Z\sim AN(E[Z_1],(1/n)\Sigma)$, where Σ is the 4x4 variance-covariance matrix of the random vector $Z_1=(V_1,W_1,V_1^2,V_1W_1)$ whose existence is guaranteed by the moment conditions given in the statement of the theorem. Therefore, $g(Z)\sim AN(m,(1/n)s^2)$ where

$$m=g[E(Z_1)],\qquad s^2=d'\Sigma d,$$

and d is the gradient vector of g(.) evaluated at $E[Z_1]\in\mathbb{R}^4$.

Remark. The conditions of the theorem on the finiteness of $E[|\log X|^4]$ and $E[|\log f(X)|^4]$ are not empty. By using the formula

$$\frac{d^r}{ds^r}\Gamma(1+s)=\int_0^\infty(\log t)^r t^s e^{-t}dt\ (s\ge 0),\ r=1,2,3,..$$

(see Cramér(1958), p. 125) it is easy to show that the moments above are finite when f belongs to one of each of the family of distributions on $[0,\infty)$: exponential, gamma, Weibull, lognormal and log double-exponential. These families are suitable for transformations in the framework (1.1) and (1.2).

Moment approximations. To assess the usefulness of the theorem above we need to relate the moments of $Z_1=(V_1,W_1,V_1^2,V_1W_1)$ to (μ,λ,σ). Under the model (1.2) , $f(X_1)=(1/2\pi\sigma^2)^{1/2}\exp[-(1/2\sigma^2)(X_1^{(\lambda)}-\mu)^2]X_1^{\lambda-1}$, holds. Taking logs and using (1.1),

$$V_1=\log X_1=(1/\lambda)\log(1+\lambda X_1^{(\lambda)}), \tag{3.1}$$

and

$$W_1=\log f(X_1)=-(1/2)\log 2\pi\sigma^2-(1/2\sigma^2)(X_1^{(\lambda)}-\mu)^2+(\lambda-1)\log X_1. \tag{3.2}$$

(In (3.1), for $\lambda=0$, take the limit when $\lambda\longrightarrow 0$). The main problem in computing expectations of quantities which depend on V_1 and W_1 arises in dealing with $E_\theta[\log X_1]$, with $\theta=(\mu,\lambda,\sigma)$. To overcome this situation, an inmediate approximation is

$$E_\theta[\log X_1]\approx E_\theta[X_1^{(\lambda)}]=\mu \tag{3.3}$$

(3.3) is obviously motivated by $\log(1+t)\approx t\,(t\longrightarrow 0)$ and is sensible for small values of λ. For $\lambda=0$, (3.3) is exact. Furthermore, note that the model (1.2) is itself an approximation and can only be valid, for positive data, when $\lambda=0$. For other techniques related to computing expectations regarding $\log X_1$, see Bickel and Doksum (1981, sec. 6). See also Draper and Cox (1969). As shown below, approximation (3.3) produces very useful results in practice.

In the following, let U denote a standard normal random variable. Using (3.3) sistematically, as well as (3.1) and (3.2), we can write

$$E_\theta[Z_1]\approx E[AS]=AE[S], \text{ and } \quad V_\theta[Z_1]\approx V[AS]=AV[S]A', \tag{3.4}$$

where $S=(U^3,U^2,U,1)'$ and A is the 4×4 matrix given by

$$A=\begin{pmatrix} 0 & 0 & \sigma & \mu \\ 0 & -(1/2) & \sigma(\lambda-1) & -\log\sigma\sqrt{2\pi}+(\lambda-1)\mu \\ 0 & \sigma^2 & 2\mu\sigma & \mu^2 \\ 0 & [(\lambda-1)\sigma^2-(\mu/2)] & [-\sigma\log\sigma\sqrt{2\pi}+2\sigma(\lambda-1)\mu] & [-\mu\log\sigma\sqrt{2\pi}+(\lambda-1)\mu^2] \end{pmatrix}.$$

Observe, finally, that

$$E[S]=\begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \text{ and } V[S]=\begin{pmatrix} 15 & 0 & 3 & 0 \\ 0 & 2 & 0 & 0 \\ 3 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.5}$$

We now relate L_n, D_n and A_n to the triplet (μ,λ,σ).

(1) L_n. Recalling (10), $\log Q(N)= Q_{\log X}(N)\approx Q_{X^{(\lambda)}}(N)= \mu+\sigma\phi^{-1}(N)$ and, using $\phi^{-1}(N)\sim N(0,1)$ and the definition of $J(N)$, we have $\int_0^1 [\log Q(t)]J(t)dt=E[\log Q(N)J(N)]\approx-(\mu/2)(1+\log 2\pi)$.

(2) D_n/n. Exploiting again (10), $\text{var}_\theta[\log X_1]\approx\text{var}_\theta[X_1^{(\lambda)}]=\sigma^2$

(3) A_n. By using (3.4), (3.5) and the definition of g it's not hard to find (after somewhat tedious calculations) that

$$m=g(E_\theta[Z_1])\approx\lambda \quad \text{and} \quad s^2=d'\Sigma d\approx(5/2)/\sigma^2.$$

Practical considerations. The theorem and approximations above provide strongly consistent estimators for μ and σ in an obvious way. For practical purposes about λ, the following points are important: (*i*) We have that $B_n \longrightarrow c\in\mathbb{R}$, a.e.. Since $\int_0^1 J(t)dt=-(1/2)(1+\log 2\pi)$, we get $c\approx 0$. This justifies choosing $\lambda_n=A_n$ as the "estimator" of λ. A_n is an approximate weakly consistent estimator for λ. Notice also that A_n is asymptotically normal with asymptotic mean around λ. (*ii*) A_n depends on the unknown density f. The latter must be replaced by a suitable density estimator $\hat{f}_n(.)$, computed from the sample, to obtain, in obvious notation, the estimator $\hat{A}_n$. When λ is small, a distribution arising from the model (2) is typically long-tailed. Practical experiences performed by the author suggest taking $\hat{f}_n(.)$ as the adaptive kernel estimate defined by

$$\hat{f}_n(x)=n^{-1}\sum_{j=1}^{n}(\lambda_j h_n)^{-1}K\{(\lambda_j h_n)^{-1}(x-X_j)\}, \tag{3.6}$$

where h_n is the bandwidth, $K(.)$ is the gaussian kernel $K(t)=(1/2\pi)^{1/2}\exp[-(t^2/2)]$, and the constants λ_j are the local bandwidth factors defined by $\lambda_j=\{\tilde{f}_n(X_j)/g\}^{-1/2}$, where g is the geometric mean of the $\tilde{f}_n(X_j)$. $\tilde{f}_n(.)$ is itself a pilot estimate of the density which may be taken as the automatic kernel estimate

$$\tilde{f}_n(x)=(nh_n)^{-1}\sum_{j=1}^{n}K(h_n^{-1}(x-X_j)), \tag{3.7}$$

with bandwidth (the same as in (3.6)), $h_n=0.9\ n^{-1/5}\min(R/1.34, St)$, where R and St are, respectively, the interquartile range and the standard deviation of the sample. Proposals (3.6) and (3.7) are taken from Silverman (1986, chapters 3 and 5). *(iii)* Replacing f by $\hat{f}_n$ creates some technical difficulties since asymptotic normality of A_n does not necesarily transmit to $\hat{A}_n$. Keeping this warning in mind an approximate studentization procedure follows. Taking the approximate distribution of $\sigma(2n/5)^{1/2}(\hat{A}_n-\lambda)$ as N(0,1), replace σ by the estimate $(D_n/n)^{1/2}$ to get the approximate $(1-\alpha)\times100\%$ interval for λ

$$\hat{A}_n \pm (5/2D_n)^{1/2}z_{\alpha/2} \tag{3.8}$$

where $z_{\alpha/2}$ is the appropriate quantile of the N(0,1) distribution.

4. Comparisons. Let $(\hat{\mu}_M,\hat{\lambda}_M,\hat{\sigma}_M)$ denote the maximum likelihood estimator of (μ,λ,σ) under the model (2). Following Bickel and Doksum (1981) , I will treat $(\hat{\mu}_M,\hat{\lambda}_M,\hat{\sigma}_M)$ as asymptotically normal with mean (μ,λ,σ) and asymptotic variance-covariance matrix n^{-1} times the information matrix under (1.2). Let $L(\mu,\lambda,\sigma)$ denote the log-likelihood of the sample. The usual $(1-\alpha)$x100% asymptotic interval for λ using likelihood methods is the set of all λ values such that

$$L_{max}(\hat{\lambda}_M)-L_{max}(\lambda)\le(1/2)\chi^2_{1,\alpha} \quad , \qquad (4.1)$$

where $L_{max}(\lambda)=\max_{\mu,\sigma} L(\mu,\lambda,\sigma)$ and $\alpha=P[\chi^2_1\ge\chi^2_{1,\alpha}]$. To compare $\hat{A}_n$ with $\hat{\lambda}_M$ and the interval (3.8) with (4.1), note that: (*i*) Exact computation of $\hat{\lambda}_M$ requires iteration and interval (4.1) is usually handled through a grid of λ-values. In contrast, $\hat{A}_n$ and interval (3.8) are computed directly from the sample. (*ii*) The new method provides the scatter plot of the pairs $(V_j,\hat{U}_j)$ as a useful exploratory tool for assessing the need for transformation of the data. (*iii*) For the case $\lambda=0$, the asymptotic variance of $\hat{\lambda}_M$ is given by $(2/3)/\sigma^2$ (see Hinkley (1975) and Bickel and Doksum (1981)). On the other hand, the asymptotic variance of A_n is $(5/2)/\sigma^2$. Therefore, the approximate ARE (asymptotic relative efficiency) of $\hat{\lambda}_M$ to $\hat{A}_n$ is

$$\hat{ARE}=(5/2)/(2/3)=15/4=3.75 \qquad (4.2)$$

The quantile method, then, has poor efficiency properties. This inconvenient is counterbalanced by the comments in points (*i*) and (*ii*).

Hinkley's (1975) method is complicated to use since it requires solving a trascendental equation first. If $\hat{\lambda}_H$ is Hinkley's estimator, it's asymptotic variance has a very complicated expression. For the exponential case, $f(x)=\exp(-x)$, Hinkley (1975) shows that $\hat{\lambda}_M$ converges to 0.2564 with asymptotic variance 0.314.

For $\hat{A}_n$ it can be shown, using the theorem in section 3 and formula (3.5) that $\hat{A}_n$ converges to 0.392 with asimptotic variance 0.918. Recall that 0.918/0.314=2.92. This is to be compared to (4.2).

5. **Examples.** *a) Simulation results.* Under the model (1.2) with $\mu=0$ and $\sigma=1$, 1000 samples are generated for each combination of $\lambda\in\{0.0,0.25,0.33,0.5\}$ and sample sizes n=10, 25, 49 and 75. The method of simulation consists in generating a sample $(y_1, \ldots, y_n)$ of N(0,1), identifying $y_i=X_i^{(\lambda)}$ and then using (1.1) to get the data $(X_1, \ldots, X_n)$. If $y_i\leq-(1/\lambda)$, y_i is replaced by 0 or $X_i=1$. This yields a sample which follows approximately the model (1.2).

Table 1

The table shows a reasonably satisfactory behaviour of A_n. In practice f, must be replaced by the density estimator $\hat{f}_n(.)$ constructed as in (3.6). I consider in detail, for illustration, the case $\lambda=0$ and n=49.

For a specific generated sample of size 49 under the model (1.2), figures 1 and 2 are, respectively, the scatter plots of (V_j,U_j) and $(V_j,\hat{U}_j)$. Notice the linear trend of both plots. Also A_n=0.027 and $\hat{A}_n$=0.231. The value of $\hat{A}_n$ is clearly inacceptable. Notice that $\hat{A}_n$ is the least-squares estimate of the regression of W_j on V_j. Therefore, the value of $\hat{A}_n$ is badly "influenced" (Cook (1977)) by points not perfectly fitted by the estimate $\hat{f}_n(.)$. A possible remedial action is: (*i*) Since the graph (V_j,W_j) is not linear, use

$(V_j, \hat{U}_j)$ to detect influential points on $\hat{A}_n$. *(ii)* Delete the bad points detected and compute $\hat{A}_n$ with the remaining ones.

In this case, if points 1 and 4 are deleted from the analysis, the new $\hat{A}_n$ equals 0.045.

Figure 1

Figure2

b) A real data example.

Figure 3

Bhattacharyya and Johnson (1977, p. 51) present a data set of size 40 obtained in an epidemiological study. Figure 3 is the plot of pairs $(V_j, \hat{U}_j)$. The linear trend suggests considering an analysis under the model (2). The maximum likelihood estimate is $\hat{\lambda}_M$=-0.215 and the interval (16) (-0.63,0.20). The quantile method yields $\hat{A}_n$=-0.17 with associated interval (-0.95,0.61). The latter interval has length 1.56 while the former has length 0.83. This is not unexpected, in view of the comments regarding the ARE of the two methods in section 4.

6. Final Comments. In this paper, a quantile based approach to the estimation of the one dimensional Box-Cox transformation when the data come in the form of a random sample, is studied. The method is valid only for positive data and small values of λ. This is not a serious limitation since this is the most important case in practice. Small values of λ ($|\lambda| \le 0.5$) are reasonable when it is suspected that $(\mathrm{var}[X])^{1/2} \propto (E[X])^c$, $c \ge 0.5$. ($c=0$ corresponds to taking logs). Extension of the proposed methodology to a general means model of the form $E[X_{ij}]=\mu_i$ ($i=1, \ldots, k$; $j=1, \ldots, n_{ij}$) is fairly straightforward by considering the quantile function in each of the groups. However, the extension to the regression case requires a new definition of the quantile function given by Basset and Koenker (1982). The latter does not allow an easy obtention of an analog of (2.3) and, consequently, extension to the regression case remains as an open problem.

REFERENCES

BASSET, G. & KOENKER, R. (1982). An empirical quantile function for linear models with i.i.d. errors, *J. Am. Stat. Ass.* **77**, 407-15

BHATTACHARYYA, G.K. & JOHNSON, R.A. (1977) *Statistical Concepts and Methods*. New York: J. Wiley.

BICKEL, P.J. & DOKSUM, K.A. (1981). An analysis of transformations revisited, *J. Am. Stat. Ass.*, **76**, 296-311.

BOX, G.E.P. & COX, D.R. (1964). An analysis of transformations, *J.R. Stat. Soc.* B **26**, 211-52.

COOK, R.D. (1977). Detection of influential observations in linear regression, *Technometrics*, **19**, 15-8.

CRAMÉR, H. (1958). *Mathematical Methods of Statistics*. Princeton NJ: Princeton University Press.

DRAPER, N.R. & COX, D.R. (1969). On distributions and its transformations to normality, *J. R. Stat. Soc.* B **31**, 472-76.

HINKLEY, D. (1975). On power transformations to symmetry, *Biometrika*, **62**, 101-12.

PARZEN, E. (1979). Nonparametric Statistical Data Modeling, *J. Am. Stat. Ass.*, **74**, 105-31.

REISS, R.D. (1989). *Approximate Distributions of Order Statistics*. New York: Springer Verlag.

SERFLING, R. (1980). *Approximation Theorems of Mathematical Statistics*. New York: J. Wiley.

SILVERMAN, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Table 1. Means and mean square errors for A_n

n \ λ	0.0	0.25	0.33	0.5
10	0.024 [1.164]	0.251 [1.159]	0.341 [1.125]	0.426 [1.127]
25	0.001 [1.092]	0.323 [1.075]	0.396 [1.057]	0.497 [1.040]
49	0.007 [1.051]	0.342 [1.037]	0.446 [1.028]	0.526 [1.017]
75	0.003 [1.033]	0.368 [1.027]	0.437 [1.020]	0.538 [1.010]

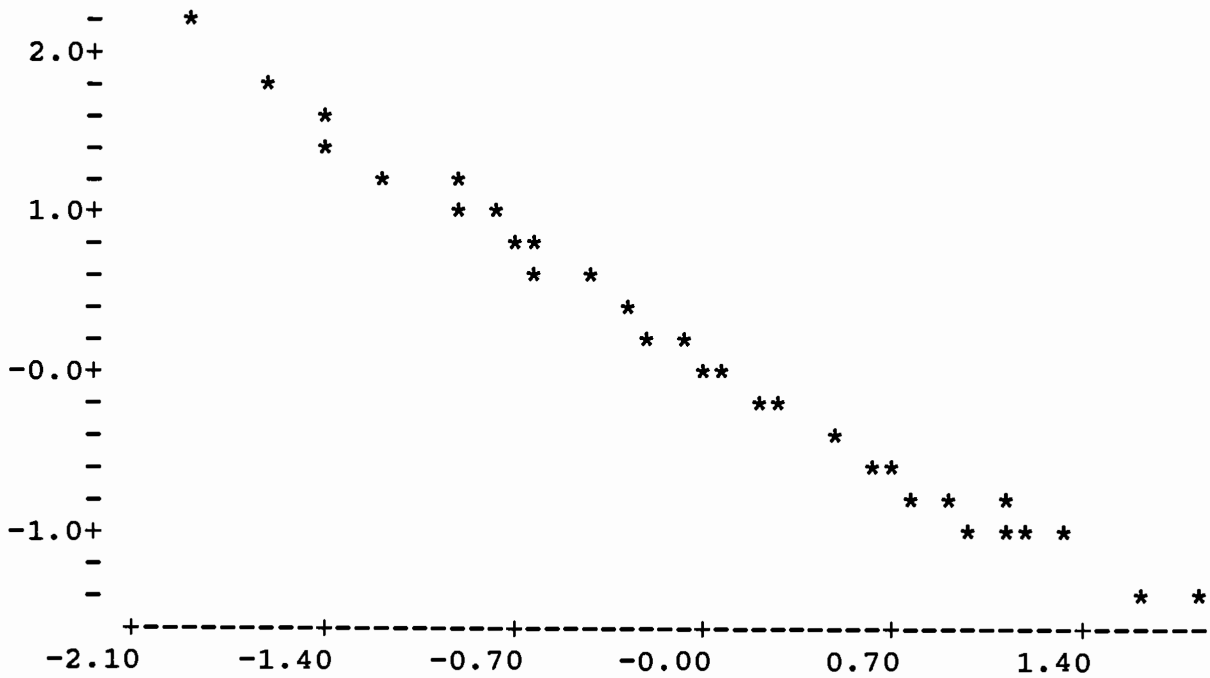

Figure 1

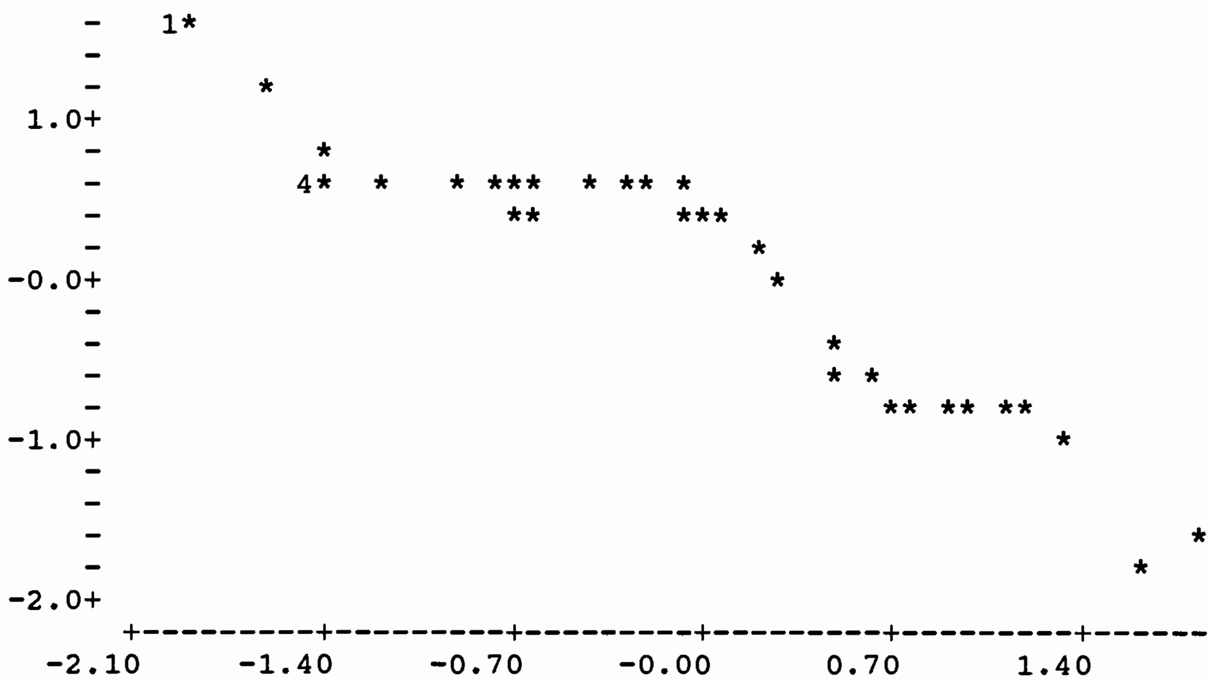

Figure 2

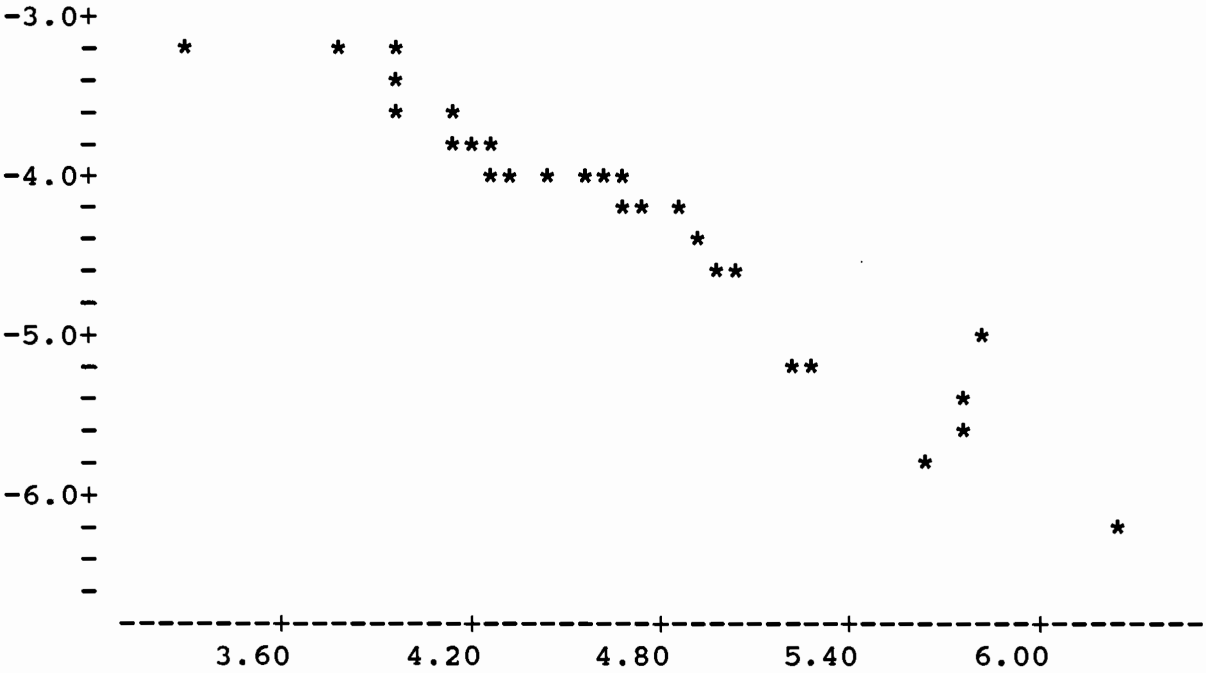

Figure 3

CAPTIONS FOR FIGURES

Fig. 1. Simulation data, $n=49$, $\lambda=0$. Plot of $U_j=\log f(X_j)$ vs. $V_j=\log X_j$.

Fig. 2. Simulation data, $n=49$, $\lambda=0$. Plot of $\hat{U}_j=\log\hat{f}_n(X_j)$ vs. $V_j=\log X_j$

Fig. 3. Epidemiological data. Plot of $\hat{U}_j=\log\hat{f}_n(X_j)$ vs. $V_j=\log X_j$.