Statistical Science
2004, Vol. 19, No. 4, 652–662
DOI 10.1214/088342304000000387
© Institute of Mathematical Statistics, 2004

Quantile Probability and Statistical Data Modeling

Emanuel Parzen

Abstract. Quantile and conditional quantile statistical thinking, as I have innovated it in my research since 1976, is outlined in this comprehensive survey and introductory course in quantile data analysis. We propose that a unification of the theory and practice of statistical methods of data modeling may be possible by a quantile perspective. Our broad range of topics of univariate and bivariate probability and statistics are best summarized by the key words. Two fascinating practical examples are given that involve positive mean and negative median investment returns, and the relationship between radon concentration and cancer.

Key words and phrases: Mid-distribution transform, percent function, percentile function, quantile function, monotone transform, parameter inverse pivot quantile function, confidence Q–Q curve, quantile–quartile function $QIQ(u)$, density quantile, quantile density, conditional quantile, comparison distribution, comparison density, Bayesian inference using quantile simulation, bivariate dependence, component correlations.

0. PHILOSOPHY

Quantile and conditional quantile statistical methods are not widely practiced in introductory statistics courses. They were pioneered by Galton (1889), who computed medians and quartiles of conditional distributions of heights of sons given heights of parents, and discovered that they had constant scale and linear location. Galton thus pioneered regression, correlation, bivariate normal distributions and conditional normal distributions. Many facts about quantiles have a long history and were known before 1900 (see Hald, 1998).

Quantile statistical thinking, as I have innovated it in my research since Parzen (1979), is outlined in this paper. My teaching philosophy has as its maxim: to earn more, learn more, and believe that learning a lot (answering all related questions) is easier than learning little (answering only the questions asked).

I teach that statistics (done the quantile way) can be simultaneously frequentist and Bayesian, confidence intervals and credible intervals, parametric and nonparametric, continuous and discrete data. My first step in data modeling is identification of parametric models; if they do not fit, we provide nonparametric models for fitting and simulating the data. The practice of statistics, and the modeling (mining) of data, can be elegant and provide intellectual and sensual pleasure. Fitting distributions to data is an important industry in which statisticians are not yet vendors. We believe that unifications of statistical methods can enable us to advertise, "What is your question? Statisticians have answers!"

1. PROBABILITY LAW OF RANDOM VARIABLE Y

Concepts to describe the probability distribution of a random variable Y include distribution function $F(y) = P[Y \le y]$, quantile function $Q(u) = F^{-1}(u)$, probability mass function $p(y) = P[Y = y]$, probability density function $f(y) = F'(y)$ and mid-distribution function $F^{\mathrm{mid}}(y) = F(y) - 0.5p(y)$.

To denote the distinct concepts of $p(y)$ and $f(y)$, the same letter should not be used; using the same letter is detrimental to quantile domain and Bayesian reasoning. A discrete random variable can be described by $p(y)$ and a continuous variable can be described by $f(y)$.

Emanuel Parzen is Distinguished Professor, Department of Statistics, Texas A&M University, College Station, Texas 77843-3143, USA (e-mail: eparzen@stat.tamu.edu).

Important examples of continuous distributions are standard exponential $f(y) = e^{-y}$ and $F(y) = 1 - e^{-y}$, and standard normal $\phi(y)$, $\Phi(y)$. Location–scale models for continuous random variables Y represent $Y = \mu + \sigma Y_0$, where Y_0 has standard distribution $F_0(y)$. Then $F(y) = F_0((y-\mu)/\sigma)$. The Normal$(\mu, \sigma)$ distribution has $F(y) = \Phi((y-u)/\sigma)$.

2. MID-DISTRIBUTION TRANSFORM

The mid-distribution function concept $F^{\text{mid}}(y)$ is important for discrete distributions, especially sample distribution functions. When F is continuous, $U = F(Y)$ is Uniform$(0, 1)$. When F is discrete we use the mid-distribution transform $W = F^{\text{mid}}(Y)$; it has mean $E(W) = 0.5$ and variance

$$\text{Var}(W) = \tfrac{1}{12}\big(1 - E[p^2(Y)]\big).$$

This elegant formula for Var(W) is important for applications to data with ties (cf. Heckman and Zamar, 2000). Proof of the mean and variance of W can be obtained by defining Y to have values y_j with probability p_j, $u_j = p_1 + \cdots + p_j$ and $u_j^{\text{mid}} = u_j - 0.5p_j$. Verify

$$\begin{aligned}(u_j^2 - u_{j-1}^2)/2 &= p_j u_j^{\text{mid}},\\ (u_j^3 - u_{j-1}^3)/3 &= p_j(u_j^2 + u_j u_{j-1} + u_{j-1}^2)/3\\ &= p_j|u_j^{\text{mid}}|^2 + p_j^3/12.\end{aligned}$$

3. SAMPLE DISTRIBUTION FUNCTION

A sample $Y_1, \ldots, Y_n$ has: (1) a sample distribution function

$$\widetilde{F}(y) = \widetilde{P}[Y \le y] = \frac{1}{n}\sum_{t=1}^{n} I(Y_t \le y),$$

where $I(Y \le y) = 1$ or 0 as $Y \le y$ or $Y > y$; (2) a sample probability mass function $\tilde{p}(y) = \widetilde{P}[Y = y]$; and (3) a sample mid-distribution function $\widetilde{F}^{\text{mid}}(y) = \widetilde{F}(y) - 0.5\tilde{p}(y)$. A continuous version $\widetilde{F}^{\text{c}}(y)$ of the discrete sample distribution $\widetilde{F}(y)$ is defined below.

4. PERCENT FUNCTION

The distribution function can be denoted $u = F(y) = u(y)$ and called the percent function since $u(y)$ is the percent of the population whose values are less than or equal to y. Percent is similar to the p value of the statistic T under a null hypothesis H_0 about the distribution of T.

5. PERCENTILE FUNCTION

The percentile or quantile function is the inverse $y = Q(u) = F^{-1}(u) = y(u)$ of $u = F(y) = u(y)$. We call u the percent of y and call y the percentile of u.

To rigorously define $y = Q(u)$ suppose first that u is in the range of F; there exists a value y such that $u = F(y)$. Define $y = Q(u)$ to be the smallest y such that $u = F(y)$ and $F(Q(u)) = u$. The general definition of quantile function, for $0 \le u \le 1$, is

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

The graph of $y = Q(u)$ is a rotation of the mirror image of the graph of $u = F(y)$. Experts on perception report that rotating a picture often helps us see patterns. Verify geometrically that

$$\int_{-\infty}^{\infty} |F_1(y) - F_2(y)|\, dy = \int_0^1 |Q_1(u) - Q_2(u)|\, du.$$

The quantile $y = Q(u)$ of the standard exponential is

$$u = F(y) = 1 - e^{-y}, \quad y = Q(u) = -\log(1-u).$$

The quantile of standard Normal$(0, 1)$ is $\Phi^{-1}(u)$. An excellent approximation for v large of quantile $Q_v(u)$ of Gamma$(v)/v$ is given by the Wilson–Hilferty transformation

$$Q_v(u) \approx \left(\left(1 - \frac{1}{9v}\right) + \frac{1}{3}\frac{1}{v^{0.5}}\Phi^{-1}(u)\right)^3.$$

6. QUANTILE FORMULA FOR MEAN AND VARIANCE

$$E(Y) = \int_{-\infty}^{\infty} y\, dF(y) = \int_0^1 Q(u)\, du,$$

$$\text{Var}(Y) = \int_0^1 \big(Q(u) - E(Y)\big)^2\, du.$$

For a sample $Y_1, \ldots, Y_n$, the sample mean $\bar{Y}$ should be computed NOT by

$$\bar{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t,$$

but by

$$\bar{Y} = \frac{1}{n}\sum_{j=1}^{n} Y(j; n) = \int_0^1 \widetilde{Q}(u)\, du,$$

where $Y(1; n) \le \cdots \le Y(n; n)$ are the order statistics of the sample and $\widetilde{Q}(u)$ is the sample quantile function.

Quantile thinking defines statistics as summation done by sorting (ranking) data before adding.

A mean can be a misleading summary of a distribution; one should always plot the quantile function to learn skewness and tails, and outliers (see Appendix 1 for a very practical example).

The sample mean $\bar{Y} = \tilde{\mu} = \tilde{E}[y]$ is the mean of the sample distribution. The sample variance should be defined as the variance of the sample distribution, that is,

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}(Y_t - \bar{Y})^2.$$

We believe teaching statistics is made difficult by the popular definition of sample variance as

$$S^2 = \sum_{t=1}^{n}\frac{(Y_t - \bar{Y})^2}{n-1};$$

S^2 should be called the adjusted sample variance and accompanied by our general definition of sample variance.

7. PERCENTILE METHOD OF SIMULATION

Quantile function $Q(u)$ can be used to simulate Y from U, which is Uniform$(0,1)$ by $Y =_d Q(U)$. One can show

$$P[Q(U) \le y] = P[U \le F(y)] = F(y) = P[Y \le y].$$

8. CREDIBLE INTERVALS

A $1-\alpha$ credible interval for Y can be obtained from

$$\begin{aligned} P\big[&y(\alpha/2) = Q(\alpha/2) \\ &\le Y \le Q\big(1-(\alpha/2)\big) = y\big(1-(\alpha/2)\big)\big] \\ &= 1-\alpha. \end{aligned}$$

Let θ be a parameter of a probability model for Y. Given a prior distribution, one can compute the quantile function $Q(u)$ of the posterior distribution of θ given data. One can express the Bayesian credible interval for θ, with credibility $1-\alpha$, as

$$\begin{aligned} P\big[&\theta(\alpha/2) = Q(\alpha/2) \\ &\le \theta \le Q\big(1-(\alpha/2)\big) = \theta\big(1-(\alpha/2)\big)|\text{data}\big] \\ &= 1-\alpha. \end{aligned}$$

9. CONFIDENCE INTERVAL AND PARAMETER INVERSE PIVOT QUANTILE FUNCTION

Let θ be a parameter of a probability model $f(y|\theta)$. Regard θ as a constant to be estimated. Assume we can form a pivot $\widetilde{T}(\theta)$ that (1) is a function of θ and the data, which is increasing in θ, and (2) has a distribution, when θ is the true parameter value, that is identical with the distribution of random variable T with quantile function $Q_T(u)$. Define $\theta(u)$, $0 < u < 1$, by

$$\widetilde{T}(\theta(u)) = Q_T(u), \quad \theta(u) = \widetilde{T}^{-1}(Q_T(u)).$$

We call $\theta(u)$ the parameter inverse pivot quantile function. It satisfies $F_T[\widetilde{T}(\theta(u))] = u$.

Conventional confidence intervals and hypothesis tests can be expressed in terms of $\theta(u)$. A $1-\alpha$ confidence interval for θ is

$$\theta(\alpha/2) \le \theta \le \theta\big(1-(\alpha/2)\big),$$

because when θ is the true parameter value, the set of samples for which

$$\begin{aligned} Q_T(\alpha/2) &= \widetilde{T}(\theta(\alpha/2)) \\ &\le \widetilde{T}(\theta) \le \widetilde{T}\big(\theta\big(1-(\alpha/2)\big)\big) \\ &= Q_T\big(1-(\alpha/2)\big) \end{aligned}$$

has probability $1-\alpha$. The rejection region $\theta_0 \le \theta(\alpha)$ has probability $P[\widetilde{T}(\theta_0)) \le Q_T(\alpha)] = \alpha$ under the hypothesis $H_0 : \theta = \theta_0$.

Our concept $\theta(u)$ should be compared with the concept $\hat{\theta}_\alpha$, which is defined in the bootstrap percentile method of confidence intervals (see Davison and Hinkley, 1997, page 193) as a random variable that is an endpoint of a confidence interval. Davison and Hinkley defined $P[\theta < \hat{\theta}_\alpha] = \alpha$; the probability function P should be denoted P_θ to emphasize that it is calculated under the assumption that θ is the true parameter value. Our more rigorous definition of $\theta(u)$ writes the probability statement

$$P_\theta\big[\widetilde{T}(\theta) \le \widetilde{T}(\theta(u))\big] = P[T \le Q_T(u)] = u.$$

The concept of parameter inverse pivot quantile $\theta(u)$ facilitates computation of confidence intervals for several confidence levels between 0.5 and 0.99 in order to discover any asymmetry in the confidence interval about the point estimator of θ.

10. QUANTILE FUNCTION OF MONOTONE TRANSFORMATIONS

A distribution function $F(y)$ is nondecreasing and continuous from the right. A quantile function is nondecreasing and continuous from the left. For $g(y)$ nondecreasing and continuous from the left we define $g^{-1}(z) = \sup\{y : g(y) \le z\}$. A beautiful and powerful property of quantile functions is the formula for the quantile function of $g(Y)$:

$$Q_{g(Y)}(u) = g(Q_Y(u)).$$

11. INVERSE PROPERTIES OF QUANTILES UNDER INEQUALITIES

To prove the monotone transform theorem we use the fact that, in general, the inverse properties of quantile functions hold under inequalities: $F(Q(u)) \ge u$,

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

Similarly, for $g(y)$ nondecreasing and continuous from the left,

$$g(y) \le t \quad \text{if and only if} \quad y \le g^{-1}(t).$$

The formula for $Q_{g(Y)}(u)$ follows from

$$\begin{aligned} F_{g(Y)}(t) &= P[g(Y) \le t] \\ &= P[Y \le g^{-1}(t)] \\ &= F_Y(g^{-1}(t)) \end{aligned}$$

and equivalently the inequalities

$$\begin{aligned} &F_{g(Y)}(t) \ge u, \quad F_Y(g^{-1}(t)) \ge u, \\ &g^{-1}(t) \ge Q_Y(u), \qquad t \ge g(Q_Y(u)). \end{aligned}$$

12. SAMPLE QUANTILE FUNCTION

For theory we use the sample quantile function defined by $\tilde{y}(u) = \widetilde{Q}(u) = \widetilde{F}^{-1}(u)$; it is piecewise constant and can be expressed in terms of order statistics $Y(1;n) \le \cdots \le Y(n;n)$ of the sample:

$$\widetilde{Q}(u) = Y(j;n), \quad (j-1)/n < u \le j/n.$$

We can think of sample percentile as a fractional order statistic

$$\tilde{y}(u) = Y([un];n),$$

where $[un] = j$ if $j - 1 < un \le j$.

For practice we would like a definition of sample quantile whose sample median $\tilde{y}(0.5)$ agrees with the usual definition: if $n = 2m+1$, $\tilde{y}(0.5) = Y(m+1;n)$; if $n = 2m$, $\tilde{y}(0.5) = 0.5(Y(m;n) + Y(m+1;n))$. A definition of sample quantile function which yields these formulas is the continuous version sample quantile $\widetilde{Q}^{\mathrm{c}}(u)$. If the sample consists of distinct values, define $\widetilde{Q}^{\mathrm{c}}(u)$ as piecewise linear connecting values

$$\widetilde{Q}^{\mathrm{c}}((j-0.5)/n) = Y(j;n).$$

Many computer programs (such as SPlus and Excel) use ad hoc definitions,

$$\begin{aligned} \widetilde{Q}^{\mathrm{c}}\big((j-1)/(n-1)\big) &= Y(j;n), \\ \widetilde{Q}^{\mathrm{c}}\big(j/(n+1)\big) &= Y(j;n), \\ \widetilde{Q}^{\mathrm{c}}\big((j-a)/(n+1-2a)\big) &= Y(j;n) \end{aligned}$$

for some constant a. For example, 9, 10, 11, 21, 26, 48, 56, 60, 60, 99 has a sample lower quartile 11 by definition $a = 0.5$ and 13.5 by definition $a = 1$.

Our definition extends to the case of ties in the sample. Denoting distinct values in the sample by $y_1, \ldots, y_r$, define $\widetilde{Q}^{\mathrm{c}}$ as piecewise linear connecting

$$\widetilde{Q}^{\mathrm{c}}(\widetilde{F}^{\mathrm{mid}}(y_j)) = y_j.$$

We consider $\tilde{y}(u) = \widetilde{Q}^{\mathrm{c}}(u)$ to be a definition of fractional order statistic. A continuous version sample distribution $\widetilde{F}^{\mathrm{c}}(y)$ is defined as piecewise linear connecting $\widetilde{F}^{\mathrm{c}}(y_j) = \widetilde{F}^{\mathrm{mid}}(y_j)$.

13. CONFIDENCE INTERVAL FOR QUANTILE FUNCTION

Let Y be continuous. The parameter $\theta = Q(p)$ can be defined from $F(y)$ by $F(\theta) - p = 0$. An estimator $\hat{\theta}$ of θ is defined to satisfy

$$\widetilde{F}^{\mathrm{c}}(\theta) - p = 0;$$

therefore, $\hat{\theta} = \widetilde{Q}^{\mathrm{c}}(p)$. A confidence interval for θ can be obtained by defining a pivot $\widetilde{T}(\theta)$, a function of θ and data, by

$$\widetilde{T}(\theta) = \frac{\widetilde{F}^{\mathrm{c}}(\theta) - p}{(p(1-p)/n)^{0.5}} \stackrel{d}{=} Z,$$

whose distribution independent of θ is approximately Z Normal$(0,1)$. We are using the asymptotic distribution of $\widetilde{T}(\theta)$ when θ is the true parameter value. The parameter inverse pivot quantile function $\theta(u)$, $0 \le u \le 1$, is defined to satisfy

$$\widetilde{T}(\theta(u)) = Q_Z(u);$$

explicitly,

$$\widehat{Q}(p;u) = \theta(u) = \widetilde{Q}^{\mathrm{c}}\big(p + \big(p(1-p)/n\big)^{0.5}\big) Q_Z(u).$$

We claim that (1) a conventional large sample $1-\alpha$ confidence interval for $\theta = Q(p)$ can be expressed as

$$\theta(\alpha/2) \le \theta \le \theta(1-(\alpha/2));$$

(2) a $1-\alpha$ significance test of hypothesis $\theta = \theta_0$ is rejected if $\theta_0 \le \theta(\alpha)$ or $\theta_0 \ge \theta(1-\alpha)$, depending on whether the alternative hypothesis is $\theta_0 \le \theta$ or $\theta \le \theta_0$; (3) point estimation of θ is $\theta(0.5) = \widetilde{Q}^{\mathrm{c}}(p)$. For extensions, see Rosenkrantz (2000).

14. QUARTILES, MEDIAN, QUARTILE, LOCATION AND SCALE

Important summary of a quantile function $Q(u)$ are quartiles $Q1 = Q(0.25)$, $Q3 = Q(0.75)$, and median $Q2 = Q(0.5)$. Nonparametric measures of location are $Q2$ and midquartile

$$MQ = 0.5(Q1 + Q3).$$

A measure of scale is interquartile range $IQR = Q3 - Q1$. As a measure of scale we prefer twice the interquartile range:

$$IQR2 = 2(Q3 - Q1).$$

A measure of skewness is $(Q2 - MQ)/IQR2$; its absolute value is bounded by 0.25.

General measures of scale have the form $\int_0^1 J_0(u) \cdot Q(u)\,du$ for suitable score functions $J_0(u)$ such as $J_0(u) = \Phi^{-1}(u)$ or $J_0(u) = u - 0.5$.

The Shapiro–Wilks statistic to test normality of a random variable Y is a sample version of the squared correlation

$$\rho^2\big(Q(u), \Phi^{-1}(u)\big) = \frac{[\int_0^1 \Phi^{-1}(u)Q(u)\,du]^2}{\int_0^1 [Q(u) - \int_0^1 Q(s)\,ds]^2\,du}.$$

As a test statistic, we recommend $\log \rho^2$ because it is compared with zero and is an entropy difference statistic since it is the difference of two estimators of $\log \sigma^2$.

15. SAMPLE Q–Q PLOT

A sample Q–Q plot compares a sample with a continuous quantile $Q_0(u)$, representing a model by plotting quantile functions

$$\big(Q_0(\widetilde{F}^{\mathrm{mid}}(y_j), y_j)\big) = \big(Q_0(u_j^{\mathrm{mid}}), \widetilde{Q}^{\mathrm{c}}(u_j^{\mathrm{mid}})\big),$$

where $y_1 < \cdots < y_r$ are distinct values in the sample and $u_j^{\mathrm{mid}} = \widetilde{F}^{\mathrm{mid}}(y_j)$. We believe these widely used plots are difficult to interpret. It helps to align the plots by making the functions equal at $u = 0.25$ and $u = 0.75$. This is accomplished by plotting quantile–quartile functions (defined in the next section)

$$\big(Q_0 IQ_0(u_j^{\mathrm{mid}}), \widetilde{Q}^{\mathrm{c}} I\widetilde{Q}^{\mathrm{c}}(u_j^{\mathrm{mid}})\big).$$

We outline ideas for research on the concept of "confidence Q–Q curves" to compare a model Q_0 with sample quantile $\widetilde{Q}$ of data Y. The lower confidence Q–Q curve joins linearly (using notation from Section 13),

$$\big(Q_0(\widetilde{F}^{\mathrm{mid}}(y_j)), \widehat{Q}(\widetilde{F}_0^{\mathrm{mid}}(y_j); \alpha/2)\big);$$

the upper confidence Q–Q curve joins linearly,

$$\big(Q_0(\widetilde{F}^{\mathrm{mid}}(y_j)), \widehat{Q}(\widetilde{F}_0^{\mathrm{mid}}(y_j); 1-(\alpha/2))\big),$$

Model Q_0 fits data $\widetilde{Q}^{\mathrm{c}}$ if a line exists between lower and upper confidence Q–Q curves. If the graph of $y = g(x)$ fits between the confidence curves, we conclude $Y =_d g(X)$ since $Q_Y(u) = g(Q_0(u))$, where X has quantile $Q_0(u)$. An important goal is to identify transformations of the data to normality or exponential. For a positive random variable Y, the hazard function $H(Y)$ has the property that it is exponential (see Parzen, 1979). Of related interest is the shift function $\Delta(x) = G^{-1}(F(x)) - x$ studied by Doksum (1974).

16. QUANTILE–QUARTILE FUNCTION $QIQ(u)$

We define quantile–quartile function $QIQ(u)$ of quantile $Q(u)$ as

$$QIQ(u) = \frac{Q(u) - 0.5(Q(0.25) + Q(0.75))}{2(Q(0.75) - Q(0.25))}.$$

Verify that $QIQ(0.25) = -0.25$ and $QIQ(0.75) = 0.25$.

If $QIQ(u) > 1$ or $QIQ(u) < -1$, we call u a Tukey outlier, since the value $y = Q(u)$ lies outside the fences as defined by John Tukey in his pioneering work on exploratory data analysis. The measure of skewness is $QIQ(0.5)$. The measures of tail behavior are $QIQ(0.05)$ and $QIQ(0.95)$ (see Table 1).

TABLE 1
Quantile–quartile diagnostics of tail

	***QIQ* diagnostic**
Left tail	
Short	$-0.5 < QIQ(0.05) < -0.25$
Medium	$-1 < QIQ(0.05) < -0.5$
Long	$QIQ(0.05) < -1$
Right tail	
Short	$0.25 < QIQ(0.95) < 0.5$
Medium	$0.5 < QIQ(0.95) < 1$
Long	$1 < QIQ(0.95)$

17. FOLIO OF *QIQ* PLOTS AND DATA MODELING

For data analysis, we plot the sample quantile–quartile function $\widetilde{Q}^{\mathrm{c}} I \widetilde{Q}^{\mathrm{c}}(u)$. From this normalized graph, we can identify the shape of probability models to fit to the data. To compare the fit of a location–scale model $Q(u) = u + \sigma Q_0(u)$, we plot the sample quantile–quartile function and $Q_0 I Q_0(u)$ on the same graph.

From the sample quantile–quartile function, we can diagnose symmetry and the tail behavior of the data, identify a standard distribution that might fit the data, and diagnose goodness of fit of models to the data. The study of a folio of *QIQ* plots would enable a statistician to identify distributions that fit the data, and identify distributions (especially Normal) that do NOT fit the data. An example is studied in Appendix 2.

18. DENSITY QUANTILE AND QUANTILE DENSITY FUNCTIONS

If F is continuous, $F(Q(u)) = u$ for all u. Taking derivatives,

$$f(Q(u))Q'(u) = 1.$$

Define the density quantile function $fQ(u) = f(Q(u))$, the quantile density function $q(u) = Q'(u)$ and the score function

$$J(u) = -(fQ(u))' = \frac{-f'(Q(u))}{f(Q(u))}.$$

In practice we assume representation near 0 and 1 as regularly varying functions:

$$fQ(u) = u^{\alpha_0} L(u),$$
$$fQ(1-u) = u^{\alpha_1} L(u),$$

where $L(u)$ is a slowly varying or loglike function that satisfies, for fixed $y > 0$,

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

An example of a slowly varying function is $L(u) = (-\log u)^{\beta}$.

We call α_0 and α_1 tail exponents; they are used to classify tail behavior as short ($\alpha < 1$), medium ($\alpha = 1$) or long ($\alpha > 1$). The concept of tail behavior is widely used by statisticians to describe nonnormal distributions; tail exponents provide rigorous concepts of tail behavior.

19. ASYMPTOTIC DISTRIBUTION OF SAMPLE QUANTILES

When Y is continuous, $U = F_Y(Y)$ is Uniform$(0, 1)$ and $Y = Q_Y(U)$. The sample quantile of Y can be represented as

$$\widetilde{Q}_Y(u) = Q_Y(\widetilde{Q}_U(u)).$$

By the delta method of large sample theory,

$$n^{0.5}\big(\widetilde{Q}_Y(u) - Q(u)\big) - q_Y(u) n^{0.5}\big(\widetilde{Q}_U(u) - u\big) \xrightarrow{P} 0.$$

We can show

$$n^{0.5}\big(\widetilde{Q}_U(u) - u\big) \xrightarrow{d} B(u),$$

where $B(u), 0 \le u \le 1$, is a Brownian bridge, a zero mean Gaussian process with covariance kernel $E[B(u_1)B(u_2)] = \min(u_1, u_2) - u_1 u_2$. We can conclude that

$$n^{0.5} f_Y Q_Y(u)\big(\widetilde{Q}_Y(u) - Q_Y(u)\big) \xrightarrow{d} B(u).$$

The parameters μ and σ in a location scale model, $Q_Y(u) = \mu + \sigma Q_0(u)$, $f_Y Q_Y(u) = \frac{1}{\sigma} f_0 Q_0(u)$, then satisfy approximately a regression model

$$f_0 Q_0(u) \widetilde{Q}_Y(u) = \mu f_0 Q_0(u) + \sigma f_0 Q_0(u) Q_0(u) + \frac{\sigma}{\sqrt{n}} B(u).$$

Using the reproducing kernel Hilbert space theory of continuous parameter regression we can derive asymptotically efficient estimators $\hat{\mu}$ and $\hat{\sigma}$, which are linear combinations of order statistics. We can also solve data compression problems of selecting a small number of values $u_1, \ldots, u_k$ such that $\widetilde{Q}(u_1), \ldots, \widetilde{Q}(u_k)$ have as much information for estimation and modeling as the whole quantile function.

20. CONDITIONAL QUANTILE FUNCTION

When observing (X, Y), the mean and variance approach to statistical reasoning emphasizes conditional mean $E[Y|X = x]$ and conditional variance, which are mean and variance of the conditional distribution

$$F_{Y|X=x}(y) = P[Y \le y | X = x].$$

The conditional quantile is defined as

$$Q_{Y|X=x}(u) = F^{-1}_{Y|X=x}(u).$$

We call this formula a brute force approach to calculating conditional quantile. An alternative can be developed using the fact that conditional probability

has properties analogous to the properties of probability. Therefore, for $g(y)$ nondecreasing and continuous from the left,

$$Q_{g(Y)|X=x}(u) = g(Q_{Y|X=x}(u)).$$

We can show that $F(Q(u)) = u$ if u is in the range of F, and $Q(F(y)) = y$ if y is in the range of Q. A random variable Y is in the range of Q with probability 1. Therefore we have:

THEOREM 1. *The powerful representation is* $Y = Q_Y(F_Y(Y))$ *with probability* 1.

Note that Y is equal in distribution to $Q(U)$, where U is Uniform$(0, 1)$. When Y is discrete, $F(Y)$ is not uniform; still $Y = Q(F(Y))$. The representation of Y as a transform of $F(Y)$ yields:

THEOREM 2. *The conditional quantile representation is*

$$Q_{Y|X=x}(u) = Q_Y(s),$$

where $s = Q_{F(Y)|X=x}(u)$.

To compute s we write

$$\begin{aligned} u &= F_{F(Y)|X=x}(s) = P[F(Y) \le s | X = x] \\ &= P[Y \le Q_Y(s) | X = x] = F_{Y|X=x}(Q_Y(s)). \end{aligned}$$

The relationship between u and s is a special case of the concept of comparison distribution.

21. COMPARISON DISTRIBUTION P–P PLOTS

A fundamental problem of statistics is comparison of two distributions F and G, and testing hypothesis $H_0 : F(y) = G(y)$.

If we let $u = G(y)$ and $y = G^{-1}(u)$, we can express the hypothesis as

$$H_0 : F(G^{-1}(u)) = u.$$

We can write $H_0 : D(u; G, F) = u$, where $D(u; G, F)$ is the comparison distribution function whose definition is given for (1) F and G both continuous, (2) F and G both discrete, and (3) F discrete (data) and G continuous (model). A comparison distribution was called a relative distribution by Handcock and Morris (1999).

When F and G are both continuous with probability densities $f(y)$ and $g(y)$, we assume also $F \ll G$, defined $g(y) = 0$ implies $f(y) = 0$. Then

$$D(u) = D(u; G, F) = F(G^{-1}(u))$$

satisfies $D(0) = 0$ and $D(1) = 1$. The comparison density is defined as

$$d(u; G, F) = f(G^{-1}(u))/g(G^{-1}(u)).$$

When F and G are discrete with probability mass functions $p_F(y)$ and $p_G(y)$, we assume $p_G(y) = 0$ implies $p_F(y) = 0$, and define first the comparison density function

$$d(u; G, F) = p_F(G^{-1}(u))/p_G(G^{-1}(u)).$$

The comparison distribution is defined as

$$D(u) = D(u; G, F) = \int_0^u d(s; G, F)\, ds.$$

We verify that $D(u)$ is piecewise linear between its values at $u_j = G(y_j)$, where $y_1 < \cdots < y_r$ are probability mass points of G and

$$D(u_j) = F(G^{-1}(u_j)) = F(y_j).$$

The graph of $D(u)$ joins $(G(y_j), F(y_j))$ and is called a P–P plot.

22. COMPARISON DENSITY REJECTION SIMULATION

The graph of $d(u)$ provides a rejection method of simulation, which generates a sample $Y_1, \ldots, Y_n$ from F as an acceptable subset of a sample $X_1, \ldots, X_m$ from G when there exists a bound c, $d(u) \le c$ for all u. Generate independent Uniform$(0, 1)$ U_1 and U_2. If $U_2 \le d(U_1)/c$, accept $X = G^{-1}(U_1)$ as an observed value of Y; otherwise reject X. The probability of acceptance is $1/c$. To prove the acceptance–rejection rule, verify that the area under $d(u)$ from 0 to $G(y)$ equals $D(G(y)) = F(y)$. The probability that $U_1 \le G(y)$ and $U_2 \le d(U_1)/c$ has probability $F(y)/c$. The event $Y \le y$ can be shown to have probability $F(y)$.

23. BAYESIAN THEOREM FOR POSTERIOR DISTRIBUTIONS

Parametric statistical inference assumes a probability model that depends on a parameter θ to be estimated. Bayesian inference assumes a prior distribution for the parameter θ, which is a probability mass function $p(\theta)$ if θ is discrete and is a probability density $f(\theta)$ if θ is continuous. The model for Y given θ is a probability mass function $p(Y|\theta)$ if Y is discrete and is a probability density function $f(Y|\theta)$ if Y is continuous.

TABLE 2
Bayes' formula

	Y discrete	Y continuous
θ discrete	$p(\theta\|Y)/p(\theta) = p(Y\|\theta)/p(Y)$	$p(\theta\|Y)/p(\theta) = f(Y\|\theta)/f(Y)$
θ continuous	$f(\theta\|Y)/f(\theta) = p(Y\|\theta)/p(Y)$	$f(\theta\|Y)/f(\theta) = f(Y\|\theta)/f(Y)$

The posterior distribution of θ given data Y is described by $p(\theta|Y)$ or $f(\theta|Y)$. To compute it we apply Bayes' theorem, which we state as a generalization of the basic statement of Bayes' theorem for events A and B: $P[A|B]/P(A) = P[B|A]/P[B]$. See Table 2.

24. BAYESIAN INFERENCE USING QUANTILE SIMULATION

The most informative way to compute the posterior distribution is by the posterior quantile function $Q_{\theta|Y}(u)$ using

$$Q_{\theta|Y}(u) = Q_\theta(s),$$
$$s = D^{-1}(u; F_\theta, F_{\theta|Y}),$$
$$u = D(s; F_\theta, F_{\theta|Y}).$$

We can simulate a sample from the posterior distribution using a sample from the prior distribution via rejection simulation and a formula for the comparison density $d(s; F_\theta, F_{\theta|Y})$.

When θ and Y are both continuous,

$$d(s) = d(s; F_\theta, F_{\theta|Y}) = f_{\theta|Y}(Q_\theta(s))/f_\theta(Q_\theta(s))$$
$$= f_{Y|\theta=Q_\theta(s)}(Y)/f_Y(Y).$$

Monte Carlo simulation chooses independent Uniform$(0, 1)$ S and U; accept $\theta = Q_\theta(S)$ if

$$\frac{d(S)}{\max_s d(s)} = \frac{f_{Y|\theta=Q_\theta(S)}(Y)}{\max_\theta f_{Y|\theta}(Y)} \geq U.$$

We compare the likelihood of Y under $\theta = Q_\theta(S)$ with the maximum likelihood of Y.

25. BIVARIATE DEPENDENCE DENSITY AND COMPONENT CORRELATIONS

To model and measure the dependence of bivariate data (Y, X), general tools are dependence density (or copula density)

$$d_{Y,X}(s,t) = d\big(s; F_Y, F_{Y|X=Q_X(t)}\big)$$

and component correlations

$$C_{Y,X}(j,k) = \int_0^1 \int_0^1 ds\, dt\, d_{Y,X}(s,t)\phi_{Y,j}(s)\phi_{X,k}(t)$$

for suitable orthonormal score functions. Note

$$\int_0^1 ds\, d_{Y,X}(s,t)\phi_{Y,j}(s)$$
$$= E[\phi_{Y,j}(F_Y(Y))|X = Q_X(t)],$$
$$C_{Y,X}(j,k)$$
$$= E[\phi_{Y,j}(F_Y(Y))\phi_{X,k}(F_X(X))].$$

One way to construct orthonormal score functions is

$$\phi_{Y,j}(s) = g_j(F_Y^{-1}(s)),$$

where $g_j(u)$ are orthonormal functions of y.

Empirical component correlations, estimated from data, are

$$C_{Y,X}(j,k) = \widetilde{E}\big[\phi_{Y,j}(\widetilde{F}_Y^{\text{mid}}(Y))\phi_{X,k}(\widetilde{F}_X^{\text{mid}}(X))\big].$$

To estimate $d_{Y,X}(s,t)$ we recommend logistic regression to estimate it as a function of t for s fixed. Apply it as a function of s for fixed t to compute the conditional quantile $Q_{Y|X=Q_X(t)}(u), 0 < u < 1$, by rejection simulation from the unconditional quantile $Q_Y(s), 0 < s < 1$.

The asymptotic joint distribution of

$$n^{0.5} f_Y Q_Y(s)\big(\widetilde{Q}_Y(s) - Q_Y(s)\big)$$

and

$$n^{0.5} f_X Q_X(t)\big(\widetilde{Q}_X(t) - Q_X(t)\big)$$

is bivariate normal with zero means and a covariance matrix of indicator random variables $I(F_Y(Y) \leq s)$ and $I(F_X(X) \leq t)$.

APPENDIX 1: INVESTMENT STRATEGY WITH POSITIVE MEAN GAIN AND NEGATIVE MEDIAN GAIN

Investors should be aware that a stock market trading strategy can result in a positive mean gain, but negative gains for most investors. Each week an investor invests in an IPO (initial public offering) and sells after a week with gain 80% with probability 0.5 and loss

60% with probability 0.5. Let Y denote profit after two trades (2 weeks) with an initial investment of \$10,000:

$$Y = \begin{cases} 22{,}400 & \text{if both trades gain,} \\ -2800 & \text{if one trade gains, one trade loses,} \\ -8400 & \text{if both trades lose.} \end{cases}$$

The probability mass function and mid-distribution of Y are

y	$p(y)$	$F^{\text{mid}}(y)$
−8,400	1/4	1/8
−2,800	1/2	1/2
22,400	1/4	7/8

The average gain $E(Y) = 2100$; the median $Q2 = -2800$. In other words, the strategy is "winning," since the mean is positive, but actually losing since the median is negative.

Quartiles are found by interpolation. $Q1 = 6533$, $Q3 = 21{,}000$; quantile–quartile analysis $MQ = 7233.5$, $IQR2 = 55{,}066$, $(MIN - MQ)/IQR2 = -0.284$ and $(MAX - MQ)/IQR2 = 0.275$. These diagnostics indicate very short tails, which occur when we have bimodality (two groups of small observed values and large observed values).

APPENDIX 2: EXPLORATORY DATA ANALYSIS COMPARISON OF TWO SAMPLES

Is high indoor radon concentration related to cancer of children in the home? To study this question radon concentration was measured in two types of houses: houses in which a child diagnosed with cancer had been residing and houses with no recorded cases of childhood cancer. From Devore (2004, page 43, Example 1.20) we obtained the following data on radon concentration in cancer and noncancer houses (numbers in parentheses indicate the number of repetitions of the value):

Cancer houses: 3 5 6 7 8 9(2) 10(3) 11(4) 12 13(2) 15(3) 16(3) 17 18(3) 20 21(2) 22(2) 23(2) 27 33 34 38 39 45 57 210

Noncancer houses: 3(2) 5 6(2) 7(3) 8(2) 9(4) 11(5) 12(2) 13 14 17(2) 21(2) 24(2) 29(4) 33 38 39 55(2) 85

Table 3 lists summary quantiles of the two samples. The conclusions of our data analysis follow:

Compare the location (means, medians) *of the two samples*: Cancer houses radon has a greater location parameter than do noncancer houses radon. What can be done about an extreme observation of 210 in cancer houses that inflates the mean?

TABLE 3
Numerical summary and diagnostics for radon concentration in cancer and noncancer homes

	Cancer houses	**Noncancer houses**
Sample size n	42	39
Number of distinct values	26	19
Sample mean $\bar{Y}$	22.8	19.2
Sample SD	31.7	17.0
$S/\sqrt{n}$	4.8	2.7
Sample MIN	3	3
Sample MAX	210	85
Next to MIN	5	5
Next to MAX	57	55
$Q1$	11	8
$Q2$	16	12
$Q3$	22	26.5
MQ	16.25	17.25
$IQR2$	22	36
$\frac{Q2-MQ}{IQR2}$	−0.02	−0.15
Conclusion	Symmetric	Skew
Upper fence $= MQ + IQR2$	38.5	53.5
Upper outliers	39, 45, 57, 210	55, 55, 85
$\frac{MIN-MQ}{IQR2}$	−0.61	−0.40
Conclusion	Normal with outliers	Exponential

Compare scale: The interquartile range (preferred to standard deviation) indicates that the variability of radon in noncancer homes is greater than the variability of radon in cancer homes.

Side by side box plots (which the reader can draw using Table 3): Radon in noncancer homes has a skew distribution, whereas radon in cancer homes is symmetric. Noncancer radon variability is greater than cancer radon variability.

Identification of probability laws: Noncancer homes diagnostics indicate fit by exponential distribution; cancer homes diagnostics indicate fit by normal distribution with outliers.

Comparison of two samples: The most general way to compare distributions of radon in cancer homes and noncancer homes is to plot a comparison distribution or P–P plot of

$$\left(F_{\text{radon|cancer}}(y_j),\, F_{\text{radon|nocancer}}(y_j)\right)$$

evaluated at values y_j obtained by pooling the values in each sample. Intuitively we consider $F_{\text{radon|cancer}}(y)$

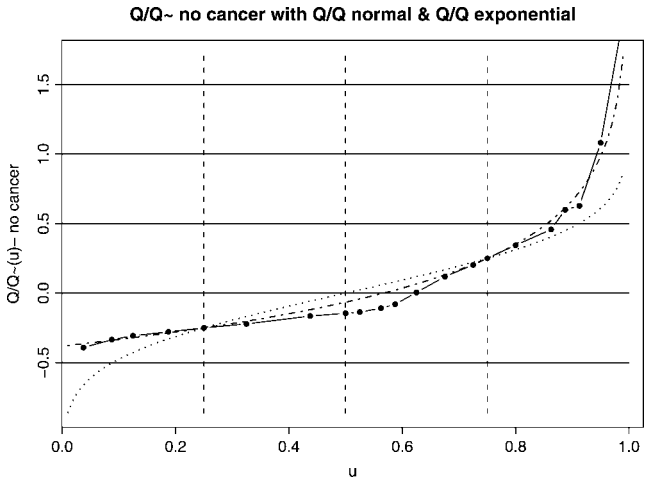

FIG. 1. *Plots of QIQ(u) for exponential and normal distributions, and sample distribution of radon in noncancer homes. Our speculation that the exponential fits the data is strengthened by this plot of the QIQ curves.*

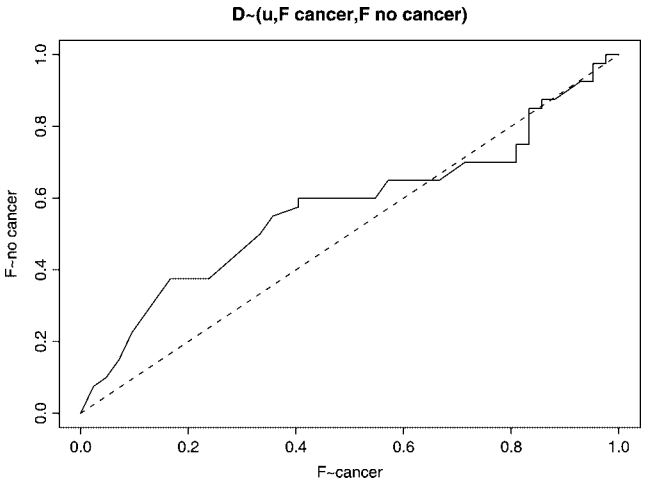

FIG. 3. *Tests of the hypothesis that the two samples (radon in noncancer and cancer homes) have the same distribution by P−P plot of the two sample distribution functions which estimate the comparison distribution* $D(u; F_{\text{cancer}}, F_{\text{nocancer}})$.

to be the conditional distribution $F_{Y|X=x}(y)$ of $Y =$ radon concentration given $X =$ type of home, cancer or noncancer. Theory discussed in Parzen (1999) leads us to recommend as the most general method that one plot

$$\left(F_{\text{radon}}(y_j),\, F_{\text{radon|nocancer}}(y_j)\right),$$

where $F_{\text{radon}}(y)$ is the distribution of radon in the pooled sample.

Quantile–quartile *QIQ* plots and comparison distribution plots are given in Figures 1–4.

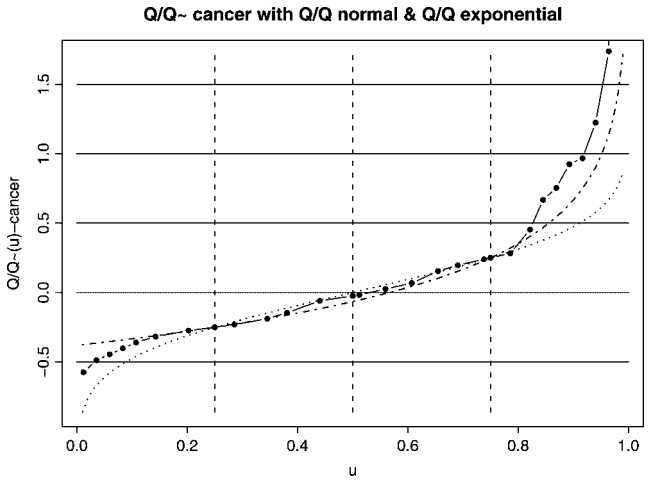

FIG. 2. *Plots of QIQ(u) for exponential and normal distributions, and sample distribution of radon in cancer homes. This plot of the QIQ curves supports our speculation that normal with outliers fits the data, but also suggests that for a better fit we should consider as a model a Weibull distribution. The dots on the sample QIQ curve represent the distinct values* y_j^Q *in the sample plotted at* $u_j = \widetilde{F}^{\text{mid}}(y_j)$; *we define* $y_j^Q = (y_j - MQ)/IQR2$. *These values are connected linearly to form sample QIQ(u). Note that sample QIQ plots always have dots at* $(0.25, -0.25)$, $(0.75, 0.25)$ *and* $(0.5, QIQ(0.5))$ *which diagnose skewness. We do not usually plot the quantile function* $Q(u)$ *because information about shape comes from* $QIQ(u)$.

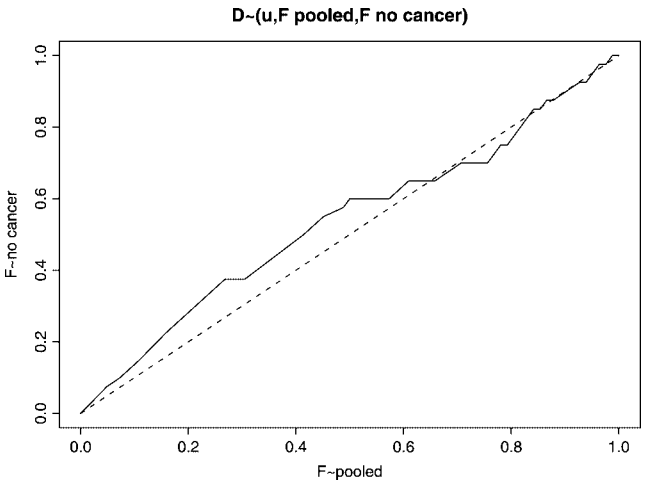

FIG. 4. *Plots of an estimate of* $D(u; F_{\text{pooledsample}}, F_{\text{nocancer}})$. *Studying the plots in Figures* 3 *and* 4 *shows why we believe the second graph may be more useful as well as able to be plotted in general. Both graphs are plotted at the distinct values in the pooled sample.*

REFERENCES

DAVISON, A. C. and HINKLEY, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge Univ. Press.

DEVORE, J. (2004). *Probability and Statistics for Engineering and the Sciences*, 6th ed. Brooks/Cole, Belmont, CA.

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

GALTON, F. (1889). *Natural Inheritance*. Macmillan, London.

HALD, A. (1998). *A History of Mathematical Statistics from 1750 to 1930*. Wiley, New York.

HANDCOCK, M. and MORRIS, M. (1999). *Relative Distribution Methods in the Social Sciences*. Springer, New York.

HECKMAN, N. and ZAMAR, R. H. (2000). Comparing the shapes of regression functions. *Biometrika* **87** 135–144.

PARZEN, E. (1979). Nonparametric statistical data modeling (with discussion). *J. Amer. Statist. Assoc.* **74** 105–131.

PARZEN, E. (1989). Multi-sample functional statistical data analysis. In *Statistical Data Analysis and Inference* (Y. Dodge, ed.) 71–84. North-Holland, Amsterdam.

PARZEN, E. (1990). Unification of statistical methods for continuous and discrete data. In *Computing Science and Statistics*: *Proc. Symposium on the Interface* (C. Page and R. LePage, eds.) 235–242. Springer, New York.

PARZEN, E. (1992). Comparison change analysis. In *Nonparametric Statistics and Related Topics* (A. K. Md. E. Saleh, ed.) 3–15. North-Holland, Amsterdam.

PARZEN, E. (1993). Change PP plot and continuous sample quantile function. *Comm. Statist. Theory Methods* **22** 3287–3304.

PARZEN, E. (1994). From comparison density to two sample analysis. In *Proc. First US/Japan Conference on the Frontiers of Statistical Modeling*: *An Informational Approach* (H. Bozdogan et al., eds.) **3** 39–56. Kluwer, Dordrecht.

PARZEN, E. (1996). Concrete statistics. In *Statistics of Quality* (S. Ghosh, W. Schucany and W. Smith, eds.) 309–332. Dekker, New York.

PARZEN, E. (1999). Statistical methods mining, two sample data analysis, comparison distributions, and quantile limit theorems. In *Asymptotic Methods in Probability and Statistics* (B. Szyszkowicz, ed.). North-Holland, Amsterdam.

ROSENKRANTZ, W. (2000). Confidence bands for quantile functions: A parametric and graphic alternative for testing goodness of fit. *Amer. Statist.* **54** 185–190.