A Last Lecture 1949-2009: Quantiles are "Optimal"

by

Emanuel Parzen

Department of Statistics, Texas A&M University, College Station, Texas, USA

Abstract

This paper discusses (1) our research to provide a framework for almost all statistical methods and philosophies, (2) need to plan the future of the "Science of Statistics" in order to compete for leadership in the practice of the "Statistics of Science", (3) grand unifying ideas of the Science of Statistics, (4) an elegant rigorous proof when quantile function minimizes check loss function which is the basis of quantile regression, (5) exact and approximate confidence quantiles (confidence interval endpoint functions) for parameters p and logodds(p) given a sample of a 0-1 variable.

Keywords statistics of science, science of statistics, mid-probability, mid-distribution, quantile, conditional quantile, confidence quantile, check loss function minimization, quantile regression, proportion p, logit(p)

Mathematics Subject Classification 62F10, 62F03, 62F15

1 United Academic Statistics: Statistics of Science, Science of Statistics, Statistical Education

The concept of "last lecture" or "last journey" implies completion of a quest begun at career's beginning. My quest has been self-education to understand (1) what almost all of statistics is about at the highest level of "analogies between analogies" (Banach, Ulam), (2) how the focus of modern applied statistical research changed over six decades.

One main conclusion (developed in Parzen (2004), (2008), sections 3 and 4 of this paper, and planned book "United Applicable Statistics") is that it is possible for applied statisticians and scientists to (1) integrate the practice of diverse approaches to statistical inference and prediction (parametric, nonparametric, Bayesian, frequentist, algorithmic), and (2) understand the history of how they were inspired (by Bayes, Gauss, Galton, Karl Pearson, Fisher, Neyman, and Tukey).

A second main conclusion is: (1) "quantiles are optimal" in the sense that they are "answer machines" that provide answers to specific practical problems, in contrast with "sub-optimal" which means "right answer to the wrong question"; (2) probability is optimal when calculated by the Science of Statistics using a mathematical (axiomatic or algorithmic) model intended to describe specific features of the real world. Probability is applied to compute P-value of a parameter; quantiles are applied to compute parameter inverse to a P-value.

A model is useful when it is in good correspondence with the real world system that it describes. We support the statisticians' motto "no model is true, but some models are useful" (even optimal in the sense of helping us achieve our goal to provide "approximate answers to the right question"). The concept that probability calculation provides "optimal" evidence for decisions provides an answer to statisticians who question if they really need to study probability theory since they never use it in exploratory data analysis. Our concept of "United Applicable Statistics" is intended to (1) answer concerns of young researchers that knowledge of the history of statistical methods is not helpful or relevant to the methods

required for modern applied statistical analysis of massive data, and (2) concerns of statistical educators about how to provide a path to including Bayesian inference and other alternatives in introductory statistics courses.

Another conclusion is regarding the process (scientific method) of the real world application of statistical thinking, about the empirical correctness of probability conclusions derived from a mathematical model when the repeated sampling interpretation (as relative frequencies in independent repeated experiments) is not applicable. It may not be applicable in exploratory sequential observational investigations; it may be applicable in "contract protocol" statistics such as designed experiments, industrial quality control, and pharmaceutical drug trials. We should explain that the Statistics of Science does not use a rigorous process to apply to reality deductive conclusions derived by the Science of Statistics under the assumption of a probability model of the real world. Conclusions that are statistically significant under a model need to be interpreted to become "rigorously" empirically significant. An important reason for the empirical success of statistics is that it is a meter that measures facts (descriptive statistics, parameters of probability models) that need to be explained (fit) by an empirical science theory for the real phenomena being observed.

This paper reports progress in my ambitious research program, whose goal is to provide a framework for statistical methods for simple data, and integrate

(1) frequentist and Bayesian methods,

(2) nonparametric and parametric methods,

(3) continuous and discrete data analysis,

(4) functional and algorithmic (numerical analysis based) data analysis.

Another integration proposal for academic statisticians is to emulate the "integrate three shields" logo of the Mayo Clinic: (1) practice applied statistics ("Statistics of Science"); (2) research on applicable statistics ("Science of Statistics"); (3) advanced and introductory statistical education.

The "Statistics of Science" (the application of statistical methods and statistical thinking to real problems) is widely practiced by many researchers in many applied fields to seek knowledge by analysis of data.

The "Science of Statistics" (the theory of applicable statistical methods) is practiced by statisticians and has as its goal to provide frameworks that unify (and thus make it easier for applied researchers to know about and apply) the very large number of statistical methods that statisticians have accumulated. I teach that data is expensive (priceless); its analysis deserves to be intensive by varying assumptions to provide several answers to compare. In statistics a question can rarely be answered in 15 minutes! I often warn my students that statisticians run the danger of telling their clients more than they want to know.

Statistical education has missions of reversing: (1) decline in core statistical scholarship, and an increasing gap between statistical methods that are known (by a few) and applied (by many); (2) introductory statistics education emphasis on cookbook recipes by emphasis on statistical thinking and strategies for statistical answers applicable to scientific questions.

Statisticians need to plan the future of the "Science of Statistics" in order to compete for leadership in the practice of the "Statistics of Science". Statisticians should emphasize their vital (and indispensable) role in the success of applied scientific research by providing knowledge of the "Science of Statistics", especially to develop methods extending to complex data the methods that have been successful in the analysis of simple data. Statisticians should get more respect from applied researchers as enabling them to "not reinvent the wheel" by applying methods inspired by problems in other fields.

In Parzen (1977), (1979) I discovered the beauty and utility of quantile functions $Q(P)$, $0 < P < 1$, and "statistic processes" on the interval $0 < P < 1$ whose asymptotic distribution under null hypotheses is a Brownian Bridge. I was inspired by analogies between time series spectral analysis and testing for white noise. Recent reviews are Parzen (2004), (2008) which emphasize confidence quantiles.

My research since 1976 has been about seeking to learn almost all of statistical methods for simple data, especially how to unify and extend statistical methods from traditional

normal models to other standard parametric models, nonparametric models, and algorithmic models. It is unclear how many of the many insights that I have had the pleasure of learning have become widely known and practiced. I believe the gap between what is known and applied is true also of an enormous related literature on model fitting (and goodness of fit) by many statisticians that I believe is overlooked by narrow mainstream academic research.

This paper discusses the following new results.

1. Elegant proof of the check loss function minimizing property of population quantiles which provides the foundation of quantile regression (Koenker (2005)) which are equivalent to conditional quantiles, the quantile function of a conditional distribution;

2. Frequentist statistical inference by probability distributions for parameters (called confidence quantiles which are inverses of mid Pvalues) for the case of one sample proportions p and $logodds(p)$ which deserve to be better taught in introductory statistics courses.

2 Grand Unifying Ideas of Science of Statistics

By grand unifying ideas of the Science of Statistics we mean concepts which enable us to better understand and apply the details of statistical methods. Important in my framework are the following ideas for parametric inference and nonparametric inference.

I. For parametric inference (on parameters of parametric models), the goal of a statistical analysis is "evidence" obtained from data as a basis for "decisions" (including confidence intervals and hypothesis tests). Evidence are best expressed as probability distributions (quantiles) for our knowledge of parameters given the data: frequentist confidence quantiles and Bayesian posterior quantiles, denoted $Q(P;\text{parameter}|\text{parameter estimator})$, $0 < P < 1$; they are endpoint function of frequentist confidence intervals and Bayesian credible intervals.

Estimating equations for confidence quantiles are obtained from quantile of sampling

distribution for parameter estimator given parameter:

$$Q(P; \text{parameter estimator}|\text{parameter})$$

under a "stochastic order" condition that it is an increasing function of the parameter for P fixed.

II. For nonparametric inference (modeling and comparing distributions without specifying finite parametric models), traditional methods can be unified by formulating them as comparison of two distributions, accomplished by a sample comparison distribution function $\tilde{D}(u)$, $0 < u < 1$, which is an estimator of a population comparison distribution $D(u)$ with comparison density $d(u)$.

The education of Ph.D. statisticians should include answers to questions of the form: "What is (basic topic name) and what are its statistical applications?" Basic frequentist topics include: Brownian Bridge and its orthogonal expansions; mid-distribution and mid-quantile; confidence intervals and confidence quantiles; Wilson Hilferty formula for Γ quantiles; comparison density, conditional comparison density; Renyi information; model identification and input modeling for simulation; changepoint analysis; extensions to censored data.

I believe that the integration of statistical practice, research, and teaching would benefit from a map that codes all statistical methods (that provides a place for every method, and every method in its place). A framework that I propose is a three dimensional description (a,b,c):

(a) data type: $Y, (X, Y)$; X and Y binary, discrete, continuous, multivariate, or functional; includes one sample, two sample, multiple sample, regression (bivariate) data;

(b) inference type: parametric, nonparametric, frequentist confidence, Bayesian posterior, censored, dependent, information divergence criterion (Karl Pearson chi squared, Fisher likelihood); inference conclusions by probability distribution of parameters (Bayes); model selection; prediction.

(c) probability model type: likelihood model $f(Y|X,$ parameters); stochastic functional model $Y = h(X,$ parameters, noise); formulas for f, h, noise distribution.

Quantiles can help develop United Statistics (unity of statistical methods) which includes following methods and concepts that deserve to be more widely practiced: quantiles as inverse distribution function, mid-quantiles, sample mid-quantiles, identification quantile for explanatory data analysis; classification of distributions by tails (ends) exponents, the ends justify the means, quantile based parameter estimators; convergence in quantile, quantile and extreme value limit theorems; comparison distribution, comparison density as density (likelihood) ratio, components for goodness of fit and tests of homogeneity; conditional quantile as quantile regression, copula, change comparison distribution; confidence quantiles as inverse P-values, midP-values, quantiles of sampling distributions as functional models, Bayesian posterior quantiles. A comprehensive book on quantile based all statistical methods could be called "Parametric and Nonparametric Statistics: Quantile Mechanics", a name recently introduced to describe deriving quantiles as solutions to differential equations (see Wilkedia article "Quantile Functions" by W. T. Shaw).

3 Quantile Minimizes Check Loss Function

The population mean $E[Y]$ of a random variable Y can be justified as the constant c minimizing the mean square error $E[|Y-c|^2]$. One proves this using the identity

$$E[|Y-c|^2] = E[|Y-E[Y]|^2] + (E[Y]-c)^2. \tag{3.1}$$

Define $\text{Pain}(c) = (E[Y]-c)^2$; the mean $E[Y]$ equals value of c minimizing mean square error, or equivalently minimizes the pain $\text{Pain}(c)$.

Given a random sample $Y_1, \cdots, Y_n$ denote the sample mean by $M(Y)$ or $\tilde{E}[Y]$; it is equal to $\text{SUM}(Y_j)/n$, and by the above reasoning equals the constant c minimizing the sample mean square error $\tilde{E}[|Y-c|^2]$.

Conditional expectation $E[Y|X]$ can be shown to equal the function $c(X)$ minimizing

mean square prediction error $E[|Y - c(X)|^2]$. One has identity for a fixed value of X

$$E[|Y - c(X)|^2|X] = E[|Y - E[Y|X]|^2|X] + (E[Y|X] - c(X))^2. \tag{3.2}$$

Using the Mean Fundamental Formula $E[E[Y|X]] = E[Y]$, obtain (by taking expectation with respect to X) that for any function $c(X)$

$$E[|Y - c(X)|^2] = E[VAR[Y|X]] + E[(E[Y|X] - c(X))^2]. \tag{3.3}$$

Define pain $\text{Pain}(c(X)) = E[(E[Y|X] - c(X))^2]$. The function $c(X)$ minimizing pain is the conditional mean $E[Y|X]$. Choosing $c(X) = E[Y]$ yields Variance Fundamental Formula

$$VAR[Y] = E[VAR[Y|X]] + VAR[E[Y|X]]. \tag{3.4}$$

Conditional means are computed by estimating equations derived from the property that $Y - E[Y|X]$ and $c(X)$ are uncorrelated for any function $c(X)$.

Our aim is to establish analogous properties for the important methods of conditional quantile function $Q(P; Y|X)$ and unconditional quantile function $Q(P), 0 < P < 1$, of a random variable Y with distribution function $F(y) = Pr[Y \leq y]$. Our proof will guide us to the general definition of the inverse distribution function or quantile function

$$Q(P) = \inf\{y : F(y) \geq P\}. \tag{3.5}$$

We call P an exact value if there is a y such that $F(y) = P$; then $F(Q(P)) = P$ and $Q(P)$ is the smallest value of y such that $F(y) = P$. If F is continuous then all P in $0 < P < 1$ are exact. If F is continuous and strictly increasing (probability density $f(y) = F'(y) > 0$) then $Q(P)$ is the unique value y satisfying $F(y) = P$.

Some important general properties (see Parzen (2004)): with probability 1 $Q(F(Y)) = Y$; $Q(U) = Y$ in distribution where U is Uniform(0,1); $Q(P; g(Y)) = g(Q(P; Y))$ if g is non-decreasing and continuous from the left. When F is continuous $F(Y) = U$ is Uniform(0,1).

We define for F discrete the discrete probability integral transform $F\text{mid}(Y)$, defining mid-distribution function $F\text{mid}(y) = F(y) - .5p(y)$, probability mass function $p(y) =$

$Pr[Y = y]$. The sample probability integral transform is defined to be $U = \tilde{F}\text{mid}(Y)$. We call U a pseudo-observation. Sample quantile function (five quantile summary, sample median, quartiles, mid-quartile MQ, twice interquartile range DQ, boxplot) are computed from sample mid-quantile function $\tilde{Q}\text{mid}(P)$, $0 < P < 1$; for a discrete random variable it has a beautiful asymptotic theory discussed by Ma, Genton, Parzen (2009). Diagnostics of symmetry, outliers, tails are provided by sample identification quantile

$$QIQ(P) = (\tilde{Q}\text{mid}(P) - MQ)/DQ.$$

A minimization criterion definition of quantile $Q(P)$ is obtained from check loss function for random variable Y defined by Koenker (2005) who proves property only for discrete sample distributions.

DEFINITION: For $0 < P < 1, L(x; P) = (1 - P)(-x)I(x < 0) + PxI(x > 0)$. Note $L(x; .5) = .5|x|$, $L(x; P) = x(P - I(x < 0))$. We propose "pain" as a name of a penalty function whose minimization is equivalent to the minimization of an objective function.

THEOREM 3.1. Assume Y has finite mean; then tails of random variable Y obey $yPr[|Y| > y]$ tends to 0 as y tends to ∞. Quantile $Q(P)$ equals constant c minimizing $E[L(Y - c; P)]$ because for any constants c and m

$$E[L(Y - c; P)] = E[L(Y - m; P)] + \text{Pain } (c, m) \tag{3.6}$$

where formula for pain is

$$\text{Pain}(c, m) = \int_c^m (P - F(y))dy, \text{for } m > c; \tag{3.7}$$
$$\text{Pain}(c, m) = \int_m^c (F(y) - P)dy, \text{for } m < c.$$

Without assumption about tails of Y choose $M1$ and $M2$ satisfying $M1 < c, m < M2$. One can verify

$$E[L(Y - c; P)I(M1 < Y < M2)] = E[L(Y - m; P)I(M1 < Y < M2)] + \text{ Pain}(c, m)$$
$$+(m - c)(F(M1)(1 - P) - (1 - F(M2))P).$$

For $m = Q(P)$ and fixed c, choose $M1$ and $M2$ so that $F(M1)$ and $1 - F(M2)$ are small enough so that following optimization property holds:

$$E[L(Y - c; P)I(M1 < Y < M2)] \geq E[L(Y - m; P)I(M1 < Y < M2)].$$

Proof: We first prove that formula (3.7) for pain implies $Q(P)$ is the value of c minimizing $E[L(Y - c; P)]$. Fix m, a candidate for the minimizing c. The condition that pain is non-negative for all c implies

$$F(y) - P \geq 0 \text{ for all } y \geq m; P - F(y) \geq 0 \text{ for all } y \leq m. \tag{3.8}$$

The value of m satisfying this condition is $m = Q(P) = \inf\{y : F(y) \geq P\}$.

To prove formula for pain verify that by integration by parts

$$\int_{M1}^{c} (c - y)dF(y) = \int_{M1}^{c} F(y)dy - (c - M1)F(M1) \tag{3.9}$$
$$\int_{c}^{M2} (y - c)dF(y) = \int_{c}^{M2} (1 - F(y))dy - (M2 - c)(1 - F(M2)).$$

Verify

$$E[L(Y - c; P)] = (1 - P) \int_{-\infty}^{c} F(y)dy + P \int_{c}^{\infty} (1 - F(y))dy \tag{3.10}$$

and apply this formula to calculate $E[L(Y - c; P)] = E[L(Y - m; P)] + \text{Pain}(c, m)$.

SAMPLE UNIVARIATE QUANTILE: Analogous conclusions holds for a sample $Y_1, \ldots, Y_n$ of Y with sample expectation $\tilde{E}[L(Y - c; P)]$, sample distribution $\tilde{F}(y)$, sample quantile $\tilde{Q}(P)$.

QUANTILE REGRESSION: For a bivariate sample (X, Y) with finite mean $E[Y]$ and conditional mean $E[Y|X]$ above reasoning proves that conditional quantile $Q(P; Y|X)$ is function $c(X)$ minimizing $E[L(Y - c(X); P)]$. This fact motivates methods of quantile regression, pioneered by Roger Koenker and exposited in Koenker (2005), to numerically compute conditional quantiles.

CONDITIONAL QUANTILE ESTIMATION: We outline improvements to our approach (Parzen (2004)) to estimating conditional quantiles.

THEOREM 3.2. Use the elegant formula: with probability 1 $Y = Q_Y(F_Y(Y))$ to infer immediately by the formula for the quantile function of a monotone transformation for Y continuous or discrete

$$Q(P; Y|X) = Q_Y(Q(P; F_Y(Y)|X)). \tag{3.11}$$

Our improved approach to estimation from a sample of the conditional quantile given X of $F_Y(Y)$ is novel because it starts with transform observations (X_j, Y_j) to pseudo-observations (U_j, V_j) where

$$U_j = \tilde{F}\text{mid}_X(X_j), V_j = \tilde{F}\text{mid}_Y(Y_j) = (\text{rank}(Y_j) - .5)/n \tag{3.12}$$

applying rank command (of R statistical computing platform) which assigns rank (Y_j) to be average of ranks of all Y values equal to Y_j.

COPULA DENSITY: Estimation of conditional density of V given U has applications in many fields and has many names: copula density or dependence density, or conditional comparison density

$$d(u, v) = f_{X,Y}(Q_X(u), Q_Y(v))/f_X(Q_X(u))f_Y(Q_Y(v)). \tag{3.13}$$

Smoothing methods available to complete the estimation of conditional quantile of V given U include bivariate kernel density estimation, wavelets, exponential family.

4 Confidence Quantiles, Parameters p, logodds(p)

Introductory statistics textbooks present frequentist inference for 0-1 random variable Y with parameter $p = Pr[Y = 1]$ that does not follow the advice to teach the most accurate methods recommended for practical use. This section applies confidence quantile methods to provide new accurate formulas for the endpoint function of (mid-probability based) confidence intervals for parameters p and logodds(p).

DATA: Denote by K the number of values 1 observed in n trials, sample probability $\tilde{p} = K/n = \tilde{Pr}[Y = 1]$ =Proportion$[Y_1, \ldots, Y_n = 1]$. Observed value of random variable K is

often denoted Kobs.

SAMPLING DISTRIBUTION OF PARAMETER ESTIMATOR GIVEN PARAMETER: Exact sampling distribution of random variable K given value p of parameter is Binomial(n, p); therefore

$$E[K] = np, VAR[K] = np(1-p), E[\tilde{p}] = p, VAR[\tilde{p}] = p(1-p)/n. \tag{4.1}$$

We call p a moment parameter since it is the mean of the sufficient statistic $\tilde{p}$. Because of its role in exponential family representation of probability law of Y we define "inverse parameter" (or natural parameter) logodds$(p) = \log(p/(1-p))$. To avoid unnecessary jargon we use logodds(p) rather than usual notation logit(p).

Frequentist inference of p and logodds(p) starts with exact or approximate sampling distributions for $\tilde{p}$ and logodds$(\tilde{p})$ given the value p. An intuitive statement of asymptotic Normal approximations is, using Z as a generic symbol for a Normal(0,1) random variable, random variable representation of distributions

$$\tilde{p} = p + \sqrt{(p(1-p)/n)}Z \tag{4.2}$$

$$\log \text{odds}(\tilde{p}) = \log \text{odds}(p) + \sqrt{(1/np(1-p))}Z. \tag{4.3}$$

When one is approximating the distribution of a discrete random variable by a continuous Z Normal(0,1) one usually states the approximation with a continuity correction. We prefer to state it as an approximation of mid-probabilities and the mid-distribution function:

$$F\text{mid}(x; \tilde{p}) = \text{Mid}Pr[\tilde{p} \leq x] = Pr[p + \sqrt{(p(1-p)/n)}Z \leq x] \text{ approximately.} \tag{4.4}$$

PROBABILITY DISTRIBUTION OF PARAMETER p GIVEN OBSERVED VALUE OF PARAMETER ESTIMATOR $\tilde{p}$: Traditional frequentist statistical inference about p given $\tilde{p}$ provides methods for confidence intervals and hypothesis tests which I regard as decisions to be made by the scientist who must make a "subjective" choice of significance level. The job of the statistician is to provide to the decision maker "objective"

evidence stated as a probability distribution for p given $\tilde{p}$obs, the observed value of the random variable $\tilde{p}$.

The Bayesian approach, which regards p as a random variable with an assumed prior distribution (often described by hyperparameters of a conjugate prior distribution), computes the posterior distribution of p, best described by the posterior quantile

$$Q(P; p|\tilde{p}\text{obs, prior hyperparameters}) = Q(P; p|\text{posterior hyperparameters.}) \tag{4.5}$$

The frequentist approach regards p as an unknown constant for which we have uncertain knowledge given $\tilde{p}$obs which we represent by the probability distribution of a (pseudo) random variable, denoted $p|\tilde{p}$, whose ("confidence") distribution is best described by its quantile function, called the confidence quantile,

$$Q(P; p|\tilde{p}\text{obs})$$

The confidence quanntile is the endpoint function of traditional confidence intervals; a 95% confidence interval for p can be represented

$$Q(.025; p|\tilde{p}\text{obs}) < p < Q(.975; p|\tilde{p}\text{obs}). \tag{4.6}$$

Our justification (and interpretation) of this interval is that it is equivalent to an interval of values of the parameter p whose MID-PVALUE satisfy

$$.025 < \text{MID-PVALUE}(p; \tilde{p}\text{obs}) < .975. \tag{4.7}$$

PVALUE AND MID-PVALUE: We define the concept of PVALUE as a function of the parameter p given $\tilde{p}$obs:

$$\text{PVALUE}(p; \tilde{p}\text{obs}) = Pr[K \geq K\text{obs } |p]. \tag{4.8}$$

Agresti and Gottard (2006) in their paper "reducing conservatism of exact small sample methods of inference for discrete data" demonstrate the benefits in practice of replacing

PVALUE by

$$\text{MID-PVALUE}(p; \tilde{p}\text{obs}) = \text{Mid}Pr[K \geq K\text{obs}|p] = \tag{4.9}$$
$$= Pr[K \geq K\text{obs}|p] - .5Pr[K = K\text{obs}|p].$$

DEFINITION: CONFIDENCE QUANTILE UNDER STOCHASTIC ORDER CONDITION ON SAMPLING DISTRIBUTION: Assume stochastic order condition

$$Q(P; \tilde{p}|p) \text{ is increasing function of } p, \text{ for } P \text{ fixed;}$$
$$1 - F(y; \tilde{p}|p) \text{ is increasing function of } p, \text{ for } y \text{ fixed;}$$
$$P = \text{MID-PVALUE}(p; \tilde{p}\text{obs}) \text{ is increasing function of } p, \text{ for } \tilde{p} \text{ obs fixed.}$$

Confidence quantile $Q(P; p|\tilde{p}\text{obs})$ is defined as value of p at which MID-PVALUE function equals P. From its definition we obtain a general estimating equation

$$Q(1 - P; \tilde{p}|Q(P; p|\tilde{p}\text{obs})) = \tilde{p}\text{obs}. \tag{4.10}$$

Confidence quantile is frequentist summary of our knowledge about p. The Bayesian posterior quantile assuming a flat uninformative conjugate prior usually coincides with confidence quantile. *This motivates us to algorithmically manipulate confidence quantiles as if they were usual quantiles of axiomatic probability.*

In practice we recommend computing a quantile $Q(P; p|\tilde{p}\text{obs}))$ at P values in Pvec

$$\text{Pvec} = c(.005, .01, .025, .05, .1, .25, .5, .75, .9, .95, .975, .99, .995).$$

This table provides evidence for traditional confidence interval and hypothesis testing decisions.

MID-PVALUE OF BINOMIAL EXPRESSED IN TERMS OF RANDOM VARIABLE WITH BETA DISTRIBTION: In terms of the order statistics of a random sample of Uniform(0,1) random variable one can give elegant proofs of following probability

theory facts:

$$Pr[\text{Binomial}(n,p) \geq k] = Pr[\text{Beta}(k, n-k+1) \leq p] \tag{4.11}$$

$$P = \text{Mid}Pr[\text{Binomial}(n,p) \geq k] = Pr[\text{Beta}(k+.5, n-k+.5) \leq p] \text{ approximately.} \tag{4.12}$$

THEOREM 4.1. Endpoints of mid-probability confidence intervals for p, equivalently exact confidence quantile of parameter p obeys

$$Q(P; p|\text{data, exact approx}) = Q(P; \text{Beta}(a^*, b^*)) \tag{4.13}$$
$$a^* = K + .5, b^* = n - K + .5, n^* = a^* + b^*, p^* = a^*/n^*.$$

Confidence quantile, which is inverse of P =MID-PVALUE$(p|\tilde{p})$, is identical with posterior quantile for conjugate prior Beta(.5,.5) Jeffreys flat prior.

We justify the very important equation (4.12) by inequality (4.14) proved in his book on Bayesian methods by Leonard (1999, p. 136) to give a frequentist interpretation of the Bayesian posterior quantile with Jeffrey's prior:

$$Pr[\text{Binomial}(n,p) \geq k] \leq Pr[\text{Beta}(k+.5, n-k+.5) \leq p] \tag{4.14}$$
$$\leq Pr[\text{Binomial}(n,p) \geq (k+1)].$$

THEOREM 4.2. Novel formula for endpoints of exact confidence intervals for logodds(p), equivalently confidence quantile of logodds(p):

$$Q(P; \text{logodds}(p)|\tilde{p}\text{obs, exact}) = \text{logodds}(p^*) + Q(P; \log F(2a^*, 2b^*)).$$

Proof: Apply the algorithm that confidence quantiles enjoy same calculus as quantiles, and known relations between Beta and F distributions to deduce the confidence quantile of the parameter logodds(p).

Before discussing normal approximations to confidence quantiles of p and logodds(p) we note the more accurate normal approximation of natural parameter logodds(p) may be used to compute approximate confidence quantile for mean parameter p by following theorem.

THEOREM 4.3. Improved approximate confidence quantile for p will be obtainable from improved approximate quantile for logodds(p), since $p = \text{inv}(\text{logodds}(p))$, defining $\text{inv}(t) = \exp t/(1+\exp t)$.

$$Q(P; p|\tilde{p}\text{obs}) = \text{inv}(Q(P;\text{logodds }(p)|\tilde{p}\text{obs})). \tag{4.15}$$

THEOREM 4.4. Normal Approximation for Confidence Quantile of p. Assume $p|\tilde{p}$obs=Beta (a^*, b^*). Define $n^{**} = n^*p^*(1-p^*)$. Note $1/n^{**} = (1/a^*) + (1/b^*)$.

$$Q(P; p|\tilde{p}\text{obs,normal}) = p^* + \sqrt{(p^*(1-p^*)/(n^*+1))}Q(P; Z). \tag{4.16}$$

THEOREM 4.5. Novel Normal Approximation for Confidence Quantile of logodds(p), applying improved normal approximation for distribution of log F.

$$Q(P; \text{logodds}(p)|\tilde{p}\text{obs, normal})=\text{logodds}(p^*) + (-1/3a^*) + (1/3b^*) + \sqrt{(1/n^{**})}Q(P; Z). \tag{4.17}$$

We omit details of the proof of the normal approximation theorems. The normal approximation to Beta(a^*, b^*) is well known (see Leonard (1999), p. 119) and is usually recommended for $a^* > 5, b^* > 5$. Our new "bias corrected" approximaton to log $F(2a^*, 2b^*)$ may provide increased accuracy for $a^* < 5$ or $b^* < 5$. There is an extensive literature on approximations to the distribution of log $F(2a^*, 2b^*)$ which is Fisher's original version of the F distribution.

The important point is that our improved bias corrected normal approximation to log F provides more accurate for small samples approximate confidence quantile for logodds(p) and therefore for p =inv(logodds(p)). They also apply to important problems of confidence intervals for log odds ratio to measure association and test independence in two samples p_1, p_2 or 2 by 2 table.

INCREASING PIVOT METHOD OF COMPUTING CONFIDENCE QUANTILES: In practice confidence quantiles such as $Q(P; p|\tilde{p}\text{obs})$ are most easily found exactly or approximately from a pivot $T\text{in}(p; \tilde{p}\text{obs})$ obeying the conditions

(1) Tin is an increasing function of p, we write Tin rather than T to emphasize that it is chosen to be increasing function;

(2) as a random variable (function of $\tilde{p}$) Tin has for all values of p distribution of random variable T.

THEOREM 4.6. Estimating equation for confidence quantile given $\tilde{p}$obs

$$T\text{in}(Q(P;p|\tilde{p}\text{obs});\tilde{p}\text{obs}) = Q(P;T), 0 < P < 1 \tag{4.18}$$

One solves the estimating equation for the confidence quantile numerically or explicitly.

For Bernoulli parameter p one may verify by computing derivative $T\text{in}\ \prime(p;\tilde{p}\text{obs})$ that increasing pivot is

$$T\text{in}(p|\tilde{p}) = (p-\tilde{p})/\sqrt{(p(1-p)/n)} = Z \tag{4.19}$$

where Z is Normal(0,1). To TEST HYPOTHESIS $p = p_0$ one compares $T\text{in}\ (p_0|\tilde{p}\text{obs})$ to quantiles of Z.

THEOREM 4.7. WILSON (1927) explicit formula for confidence quantile of parameter p given observed value $\tilde{p}$: Define $c(P) = Q(P;Z)/\sqrt{(n)}$, $c2(P) = |c(P)|^2 = |Q(P;Z)|^2/n$. By solving quadratic equation one obtains

$$Q(P;p|n,\tilde{p}) = (\tilde{p} + c2(P)/2) + c(P)\sqrt{(\tilde{p}(1-\tilde{p}) + (c2(P)/4))}/(1+c2(P)) \tag{4.20}$$

When $\tilde{p} = 0$

$$\begin{aligned} Q(P) &= c2(P)/2 + c(P)\sqrt{(c2(P)/4)}/(1+c2(P)) \\ &= 0, P < .5; = |Q(P;Z)|^2/(n+|Q(P;Z)|^2) \end{aligned} \tag{4.21}$$

Numerical comparisons show that Wilsons formula for approximate pivot confidence quantile is a good approximation to the exact inverse MID-PVALUE confidence quantile

$$Q(P;p|\tilde{p}) = Q(P;\text{Beta}\ (K+.5, n-K+.5)) \tag{4.22}$$

TWO SAMPLE INFERENCE: Applying the algorithm that mechanics of confidence quantiles are identical with mechanics of Bayesian posterior quantiles we can apply one sample confidence quantiles to derive exact and approximate two independent samples confidence quantiles.

Acknowledgments

The author is grateful to *Communications in Statistics: Theory and Methods* Associate Editor, Nitis Mukhopadhyay, and three anonymous reviewers for their comments and useful suggestions.

References

Agresti, Alan and Gottard, Anna. (2006). *Reducing conservatism of exact small sample methods of inference for discrete data*. COMSTAT 2006: Proceedings in Computational Statistics. Springer.

Leonard, T. and Hsu, John S. J. (1999). *Bayesian Methods*. Cambridge University Press.

Koenker, Roger. (2005). *Quantile Regression*. Cambridge University Press.

Ma, Y., Genton, M. G., Parzen, E. (2009). Asymptotic properties of sample quantiles of discrete distributions. *Annals of the Institute of Statistical Mathematics*.

Parzen, E. (1977). *Nonparametric statistical data science: A unified approach based on density estimation and testing for white noise*. Technical report. Statistical Science Division. SUNY at Buffalo.

Parzen, E. (1979). Nonparametric statistical data modeling. *Journal American Statistical Association*, 74:105–131.

Parzen, E. (2004). Quantile probability and statistical data modeling. *Statistical Science*, 19:652–662.

Parzen, E. (2008). United statistics, confidence quantiles, Bayesian statistics. *Journal Statistical Planning and Inference*, 138:2777–2785.

Wilson, E. B. (1927). Probable inference, the Law of Succession, and statistical inference. *Journal of the American Statistical Association*, 22:209–212.