The Korean Communications
in Statistics Vol. 3, No. 3, 1996
pp. 179-186

# A Nonparametric Procedure for Bioassay by using Conditional Quantile Processes

**Ho Kim**[1)]

## Abstract

Bioequivanence models arise typically in bioassays when new preparations are compared against standard ones by means of responses on some biological organisms. Relative potency measures provide nice interpretations for such bioequivalence and their estimation constitutes the prime interest of such studies. A conditional quantile process based on the k-nearest neighbor method is proposed for this purpose. An alternative procedure based on Kolmogrov-Smirnov type estimator has also been considered along with. ARIC ultrasound data are analyzed as examples.

## 1. Introduction

In a biological assay (bioassay), a new (test) preparation (T) and an old (standard) one (S) are compared by means of the reactions that follow their applications to biological organisms. The main interest under investigation is to examine how biological organisms react to the two preparations and to evaluate the relative potency of the test preparation with respect to the standard. To achieve this goal, one typically consider the following model

$$E(Y_T) = h_T(x) = h_S(\rho x) = E(Y_S) \tag{1}$$

for all $x$ and for some $\rho > 0$. In many cases, to get more symmetry for the tolerance distributions, one typically take log transformation. Then the mode becomes

$$E(Y_T) = h_T(x) = h_S(\mu + x) = E(Y_S), \text{ for all } x \text{ for some } \mu \in R. \tag{2}$$

The model (2) is termed the fundamental assumption of an assay. The main interest is to estimate the relative potency $\rho$ and to verify the fundamental assumption.

Parametric procedures for these problems are usually based on the assumptions of the distributions of $Y$ (normal, lognormal, logistic or loglogistic), these procedures are discussed in detail in Finney (1964). Using (generalized) linear models, the usual maximum likelihood

1) Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599-7400, USA.

estimator has many optimal properties, such as consistency and asymptotic normality under the assumed distribution. However, these parametric estimates generally not very robust against departures from the assumed form of the distribution function. Rank based estimates of relative potency have been considered by Sen(1963), Shorack(1966) and Rao and Little(1976), among others. These estimates are invariant under the choice of any monotone transformation on the dose, and, besides being robust, are generally quite efficient for normal, log-normal, logistic or other common forms of the tolerance distributions. Bhattacharya and Gangopadhyay (1990) proposed k-nearerst neighbor (k-NN) estimator of a conditional quantile. We can use the conditional quantile of the response instead of the expectation in (1), to test the equivalence of a test preparation with respect to a standard one.

Let $\{(X_i, Y_i), i\geq 1\}$ be a sequence of i.i.d. random vectors with a distribution function $\pi(x,y), (x,y)\in R^2$. Let $F(x)=\pi(x,\infty), x\in R$ and let $G(y|x)$ be the conditional d.f. of $Y$ given $X=x$, for $y\in R, x\in R$. A conditional quantile function (of $Y$ given $X=x$) is defined by

$$\xi_p(x)=\inf\{y: G(y|x)\geq p\}, x\in R, (0<p<1).$$

Consider the transformation $(X_i, Y_i)\rightarrow(Z_i, Y_i)$ where $Z_i=|X_i-x|$. For the collection $\{(Z_1, Y_1),\cdots,(Z_n, Y_n)\}$ of r.v.'s, let $Z_{n1}<\cdots<Z_{nn}$ be the order statistics corresponding to $Z_1,\cdots,Z_n$, and let $Y_{n1},\cdots,Y_{nn}$ be the induced order statistics. (i.e., $Y_{ni}=Y_j$ if $Z_{ni}=Z_j$, for $i,j=1,\cdots,n$). For every positive integer $k(\leq n)$, the k-NN (nearest neighbor) empirical d.f. of Y (with respect to $x$) is given by

$$\hat{G}_{nk}(y)=k^{-1}\sum_{i=1}^{k}1(Y_{ni}\leq y),\quad y\in R,$$

where 1(A) stands for the indicator function of the set A. The k-NN estimator of $\xi$ due to Bhattacharya and Gangopadhyay (1990) is defined by

$$\begin{aligned}\hat{\xi}_{nk} &= \text{the } [kp]-\text{th order statistic of } Y_{n1},\cdots,Y_{nk}\\ &= \inf\{y: \hat{G}_{nk}(y)\geq k^{-1}[kp]\}.\end{aligned}$$

For the asymptotic normality results, usually, one needs that $k_n=O(n^{\lambda})$, for some $\lambda\in(0,1)$. Bhattacharya and Gangopadhyay (1990) proved the asymptotic normality of the k-NN estimator for $\lambda=4/5$, .i.e., for every $t\in[a,b]$ as n$\rightarrow\infty$,

$$n^{2/5}[\hat{\xi}_{n,[n^{4/5}t]}-\xi]\rightarrow N(\beta t^2, p(1-p)t^{-1}/g(\xi)^2),$$

where

$$\beta(\xi)=-[f(x)G_{xx}(\xi|x)+2f(x)G_x(\xi|x)]/[23f^3(x)g(\xi)]$$

and

$$G_x(y|x) = (\partial/\partial x)G(y|x),$$
$$G_{xx}(y|x) = (\partial^2/\partial x^2)G(y|x).$$

Sen and Gangopadhyay (1990) showed the bias term $(\beta t^2)$ can be eliminated when $k_n = O(n^{4/5-2\delta})$. They showed that for every $k : k = [tn^{4/5-2\delta}]$, $t \in [a, b]$, as $n \to \infty$

$$n^{2/5-\delta}[\hat{\xi}_{nk} - \xi] \to N(0, t^{-1}p(1-p)/g^2(\xi)).$$

They also considered the construction of (local) bootstrap confidence intervals for $\xi$ based on the k-NN methods, and its asymptotic normality.

## 2. Conditional Quantile Process

Let $\xi(x)$ be the p-quantile of the d.f. $G(\cdot|x)$, $x \in [x_l, x_u] = X \subset R$, and let $\hat{\xi}_{n,k_n}(x)$ be the k-NN estimator for $k_n = O(n^{4/5-2\delta})$ where $Z_{ni}$ are defined with respect to the pivot x. Note that these $Z_{ni}$ are the induced order statistics corresponding to the nearest neighbors of x (i.e. those $X_i$'s which are closest to x). As such, if we consider two distinct points, say $x_1$ and $x_2$, then the two sets of $X_i$ forming the corresponding nearest neighbors would have at most $k_n$ members in common. The proportion of the two subsets' common elements will be asymptotically $o(n^{-1}k_n)$, and we may immediately say that $\hat{\xi}_{n,k_n}(x_1)$ and $\hat{\xi}_{n,k_n}(x_2)$ are asymptotically independent. We set $W_n(x) = \{W_n(x), x \in X\}$ by letting

$$W_n(x) = k_n^{1/2}[\hat{\xi}_{n,k_n}(x) - \xi(x)], \quad x \in X.$$

Then the asymptotic behavior of $W_n$ is depicted entirely by the pointwise asymptotic normality and the stochastic dependence pattern only in a shrinking neighborhood of fixed points.

We discussed the pointwise asymptotic normality in the introduction for $k_n = tn^{4/5-2\delta}$, $(0 < a < t < b < \infty)$. We have, as $n \to \infty$,

$$W_n(x) \to W(x), \quad \text{for every } x \in X,$$

where $W(x)$ is normal with mean 0 and variance $pq/(g^2(\xi(x)|x))$. About the stochastic dependency, we refer to Gangopadhyay and Sen (1993). Shrinking neighborhoods are defined by the sets of $x$ relative to a given $x_0$ as

$$I_n(x_0, K) = \{x : |x - x_0| \le Kn^{-1}k_n\},$$

where K is a finite positive number. For any two points, say, $x_n$ and $x_n'$, belonging to $I_n(x_0, K)$, let $k(x_n, x_n')$ be the number of elements in common between the two neighbor sets. Consider the intersection set $I_n(x_n, x_n')$

$$[x_n-(2nf(x_0))^{-1}k_n, x_n+(2nf(x_0))^{-1}k_n]\cap$$
$$[x_n'-(2nf(x_0))^{-1}k_n, x_n'+(2nf(x_0))^{-1}k_n].$$

Let $m_n(x_n, x_n')$ be the length of this intersection set $I_n(x_n, x_n')$. Then by Bahadur representation, we may again claim that

$$n^{-1}k(x_n, x_n')-f(x_0)m_n(x_n, x_n')=o(n^{-1}k_n) \ a.s. \text{ as } n\to\infty.$$

And we may conclude that $W_n(x_n)$ and $W_n(x_n')$ have asymptotic correlation

$$nf(x_0)m_n(x_n, x_n')/k_n, \tag{3}$$

and it is easy to show that for both $x_n$ and $x_n'$ belonging to $I_n(x_0, K)$,, (3) has a limit (which vanishes when $m_n(x_n, x_n')$=0). In order to justify this asymptotic analysis, we need to extend the Bahadur representation for a conditional quantile for a suitable convergence sequence $\{x_n\}$(to $x_0$), where $x_n\in I_n(x_0, K)$. To facilitate this, we consider a stochastic process

$$W_n^0(t)=W_n(x_0+n^{-1}k_nt), \ t\in[t_{0,t_1}].$$

Consider a compact interval $C=[-K, K]$ where $K(<\infty)$ is a positive number so chosen that $Kf(x_0)\le 1$. In passing, we may remark that (3) possesses a limit, when we set $x_n=x_0+n^{-1}k_nt_1$ and $x_n'=x_0+n^{-1}k_nt_2$, with $t_1, t_2\in C$. This is given by

$$[t_1f(x_0)-1/2, t_1f(x_0)+1/2]\cap[t_2f(x_0)-1/2, t_2f(x_0)+1/2]$$

which, for $t_2>t_1$ is non-null when $(t_2-t_1)f(x_0)\le 1$, and that is the reason we set $Kf(x_0)\le 1$. And under the assumption of the finite fourth moment of $Z^*_{ni}=G^{-1}\circ G(Z_{ni}|Y_{ni})$ for $i=1,\cdots, n$,

$$W_n^0(C)=\{W_n^0(t), \ t\in C\}$$

is asymptotically tight, where $W_n^0(t)=W_n(x_0+n^{-1}k_nt)$ and $C=[-K, K]$ such that $Kf(x_0)\le 1$, for $k_n=O(n^{4/5-2\delta})$. This, in turn, implies the weak convergence of $W_n(x)$ to $W(x)$ for $x\in X$.

To estimate the relative potency, we consider Kolmogorov-Smirnov type estimator $\mu$ which

minimize

$$\sup_{x\in X} | \hat{\xi}^T_{n,k_n}(x) - \hat{\xi}^S_{n,k_n}(x+\mu)|.$$

Let $\hat{\xi}^T_{n,k_n}$ and $\hat{\xi}^S_{n,k_n}$ be the k-NN estimators of the conditional medians of the test and the standard treatments. Then we have two independent stochastic processes

$$W^T_n(x) = k_n^{1/2}[\ \hat{\xi}_{n,k_n}{}^T(x) - \xi_T(x)] \to W_T(x),\ x \in X \text{as } n\to\infty$$
$$W^S_n(x) = k_n^{1/2}[\ \hat{\xi}_{n,k_n}{}^S(x) - \xi_S(x)] \to W_S(x),\ x \in X \text{as } n\to\infty$$

$W_T(x)$ and $W_S(x)$ are the Gaussian processes defined before. Under Ho : $\xi_T(x) = \xi_S(x+\mu)$,

$$W^T_n(x) - W^{S(x+\mu)}_n = k_n^{1/2}\{\ \hat{\xi}^T_{n,k_n}(x) - \hat{\xi}^S_{n,k_n}(x+\mu)\}$$
$$\to W_T(x) - W_S(x+\mu)\ \ x \in X,$$

where

$$W_T(x) - W_S(x+\mu) \sim N(0, pq(1/(g(\xi(x)|x)) + 1/(g(\xi(x+\mu)|x+\mu))))$$

for all $x \in X$ We may do the test by bootstrap sampling method; for m-dimensional $x' = (x_1, \cdots, x_m)$, we have m-dimensional multivariate normality after putting $\mu = \hat{\mu}$. We can estimate the variance matrix by bootstrapping method.

## 3. Numerical Examples

In this section, we demonstrate numerical examples. We use $p = 1/2$, the median of the response. Atherosclerosis Risk in Communities (ARIC) is a prospective investigation of the etiology and natural history of atherosclerosis and the etiology of clinical atherosclerotic disease in four US communities (Forsyth County, North Carolina; Jackson, Mississippi; suburbs of Minneapolis Minnesota; and Washington County, Maryland) (The ARIC investigators (1989)). Examinations included ultrasound scanning of carotid artery. The data for the example included black female participants with second visit from all four communities. The response variable is the average carotid artery far wall thickness(WT) scanned by Beta-mode ultrasound equipment. The explanatory variable is Body Mass Index (BMI) which is defined by weight/height2 (Kg/m2). The group variable is the hypertension medicine history which is one if the participant has taken hypertension lowering medication in past two weeks, zero otherwise. The main interest is to test the bioequivalence of the carotid artery far wall thickness between the two groups and to estimate the relative potency. The data has sample size 623 and 735 for the hypertension and the normotension groups, respectively.

To test the bioequivalence between the two groups and to estimate the relative potency, we

assume the bioequivalence model (parallelism of the two response curves) which is:

$$\text{Ho}: \xi^T(x)-\xi^S(x+\mu)=0 \text{ for all } x\in X. \tag{4}$$

We need to check the parallelism before doing any statistical inferences based on the assumption (4). Figure 1 shows the estimated conditional median curves over BMI, and we decided that there is no violation of the parallelism. WT curve of the hypertension group is greater than that of the normotension group for all BMI's. To estimate $\mu$, we discussed the mini-max rule in section 2. Figure 2 shows

$$\sup_{x\in X} | \hat{\xi}_n^T(x)- \hat{\xi}_n^S(x+\mu)| \tag{5}$$

as a function of $\mu$. From the figure, we estimate $\mu$ is 10 which minimizes (5), i.e., the WT curve of hypertension patients is shifted to the left by 10 units of BMI compared to the WT curve of the normotension group.

To test (4) when $\mu=10$, as we discussed in section 2, we use

$$k_n^{1/2}[\ \hat{\xi}_n^T(x)- \hat{\xi}_n^S(x+\mu)]\rightarrow N_m(0,\Sigma),$$

for $x=(x_1,\cdots,x_m)'$ as $n\rightarrow\infty$. We reject Ho if

$$Q=k_n[\ \hat{\xi}^T{}_n(x)- \hat{\xi}^S{}_n(x+\mu)]'\ \hat{\Sigma}_n^{-1}[\ \hat{\xi}^T{}_n(x)- \hat{\xi}^S{}_n(x+\mu)]>\chi^2.$$

where $\hat{\Sigma}_n$ is the variance matrix of $[\ \hat{\xi}^T{}_n(x)- \hat{\xi}^S{}_n(x+\mu)]$ which is asymptotically a diagonal matrix, where $\chi^2$ is the upper $\alpha$ percentile from $\chi^2$ distribution with m d.f. Also, we employ the following

$$Q_d=k_n\sum_{i=1}^{m}(\ \hat{\xi}^T{}_n(x_i)- \hat{\xi}^S{}_n(x+\mu))^2\ \tilde{\sigma}_{ii}^{-1},$$

where $\tilde{\sigma}_{ii}$ is a bootstrap estimator of $\text{Var}(\ \hat{\xi}^T{}_n(x_i)- \hat{\xi}^S{}_n(x_i+\mu))$. We use $x'=(15,20,25,30,35,40,45)$ with 200 bootstrap iterations to estimate $\hat{\Sigma}_n$ and $\tilde{\sigma}_{ii}$. We have Q=20.36 with p-value 0.005 and $Q_d$=18.87 with p=0.009. Therefore, we reject the hypothesis (5) which means the difference between the two WT curves are statistically significant. The difference $Q-Q_d$=1.30 can be interpreted as the increase of the $\chi^2$ due to the non-zero off-diagonal elements in $\hat{\Sigma}_n$. Since $Q-Q_d$ is small, we can say that the independence assumption of the conditional quantile over $x\in X$ is not violated.

The difference of the mean WT's between the two groups is 0.04974. We subtract the difference from the hypertension group and apply the test again. The values Q=5.88 with p=0.55 and Qd=5.56 with p=0.59 were obtained. Our conclusions are as follows: 1) the WT curve of the hypertension group is significantly higher than that of the normotension group for the entire range of BMI, 2) but hypertension effect-adjusted WT curves are equivalent for the two groups; WT curves have the same shape for the two groups.

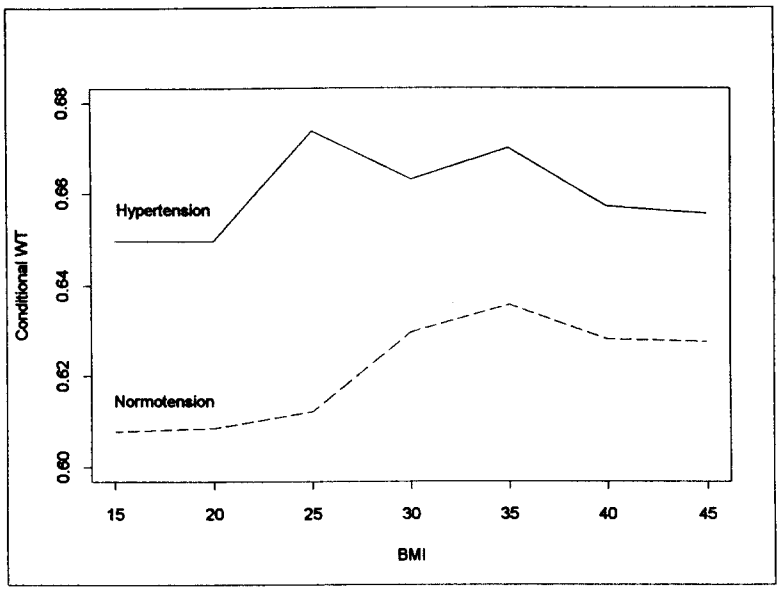


Figure 1. Estimated conditional median carotid artery far wall thickness conditioned on BMI for two groups.

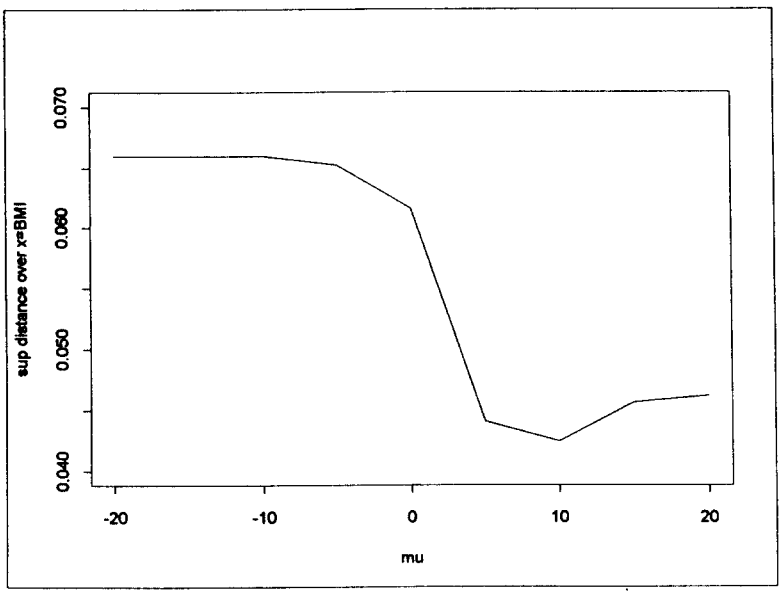


Figure 2. $\sup_{x \in X} | \hat{\xi}^T{}_n(x) - \hat{\xi}^S{}_n(x+\mu)|$ vs. $\mu$

We have several questions which may become future research topics. There may be more than two responses in which we need to consider the correlation structure among the responses. To test the fundamental assumption, we can try to obtain the distribution of $\sup_{x \in X} | W^T(x) - W^S(x+\mu) |$ which would make our test of the fundamental assumption of the bioequivalence models simpler. The Kolmogorov type estimator of the relative potency may possess some appealing properties such as asymptotic normality or consistency. One interesting question is how to select k-NN samples for the processes. In this paper, we select x's by the same increment over the range of interest. We may select exclusive samples to guarantee independence among them, or we may allow some dependency by selecting not totally exclusive samples.

## References

[1] The Aric Investigators (1989). The Atherosclerosis Risk in Communities Study : Design and Objectives, *American Journal of Epidemiology*, Vol. 129, 687-702.

[2] Bhattacharya, P.K and Gangopadhyay, A.K. (1990). Kernel and nearest neighbor estimation of a conditional quantile, *The Annals of Statistics*, Vol. 18(3), 1400-1415.

[3] Finney, D.J. (1964). *Statistical Methods in Biological Assay*, 2nd edn, Charles Griffin, London.

[4] Gangopadhyay, A.K. and Sen, P.K. (1993). Contiguity in Bahadur-type representation of a conditional quantile and application in conditional quantile process, in J.K.G. et al.(ed.), *Statistics and Probability, a Raghu Rai Bahadur Festschrift*, Wiley Eastern, 219-232, New Delhi.

[5] Rao, P.V. and Little, R.C. (1976). An estimator of relative potency, *Communications in Statistics A*, Vol. 5, 183-189.

[6] Sen, P.K. (1963). On the estimation of relative potency in dilute (-direct) assays by distribution-free methods, *Biometrics*, Vol. 19, 532-552.

[7] Sen, P.K. and Gangopadhyay, A.K.(1990). Bootstrap confidence intervals for conditional quantile functions, *Sankya : The Indian Journal of Statistics A*, Vol. 52, 346-363.

[8] Shorack, G.R. (1966). Graphical procedures for using distribution-free methods in the estimation of relative potency in dilution (-direct) assays, *Biometrics*, Vol. 22, 610-619.