A NOTE ON QUANTILES IN LARGE SAMPLES[1]

By R. R. Bahadur

University of Chicago

1. Introduction. Let $F(x)$ be a probability distribution function on the real line. Let ξ be a fixed point and let

$$F(\xi) = p. \tag{1}$$

It is assumed that F has at least two derivatives in some neighborhood of ξ, that $F''(x)$ is bounded in the neighborhood, and that $F'(\xi) = f(\xi) > 0$. These assumptions imply, in particular, that $0 < p < 1$ and that ξ is the unique p-quantile of F.

Let $\omega = (X_1, X_2, \cdots \text{ ad inf})$ be a sequence of independent random variables X_i with each X_i distributed according to F. For each $n = 1, 2, \cdots$, let $Y_n = Y_n(\omega)$ be the sample p-quantile when the sample is $(X_1, \cdots, X_n)$. Let $Z_n = Z_n(\omega)$ be the number of observations X_i in the sample $(X_1, \cdots, X_n)$ such that $X_i > \xi$. This note points out that, with $q = 1 - p$,

$$Y_n(\omega) = \xi + [(Z_n(\omega) - nq)/n \cdot f(\xi)] + R_n(\omega) \tag{2}$$

where R_n becomes negligible as $n \to \infty$. It is shown here that

$$R_n(\omega) = O(n^{-3/4} \log n) \quad \text{as} \quad n \to \infty \tag{3}$$

with probability one, but the exact order of R_n is not known at present.

The above representation of Y_n gives new insight into the well known result that $n^{\frac{1}{2}}(Y_n - \xi)$ is asymptotically normally distributed with mean 0 and variance $v = pq/f^2(\xi)$. It gives an easy access, via the multivariate central limit theorem for zero-one variables, to the asymptotic joint distribution of several quantiles in samples from a multivariate distribution [2]. The representation also shows that the law of the iterated logarithm holds for quantiles, i.e.,

$$\limsup_{n\to\infty} [n^{\frac{1}{2}}(Y_n - \xi)/(2 \log \log n)^{\frac{1}{2}}] = v^{\frac{1}{2}}, \tag{4}$$
$$\liminf_{n\to\infty} [n^{\frac{1}{2}}(Y_n - \xi)/(2 \log \log n)^{\frac{1}{2}}] = -v^{\frac{1}{2}}$$

with probability one.

The proof in the following section may be outlined as follows. Let $F_n(x, \omega)$ be the sample distribution function when the sample is $(X_1, \cdots, X_n)$, i.e., $F_n(x, \omega) = (\text{The number of } X_i \leqq x \text{ in the sample})/n$. It is shown that, with I_n a suitable neighborhood of ξ, $F_n(x, \omega) \doteq F_n(\xi, \omega) + F(x) - F(\xi)$ uniformly

Received 3 January 1966.

[1] This research was supported in part by Research Grant No. NSF GP 3707 from the Division of Mathematical Physical and Engineering Sciences of the National Science Foundation, and in part by the Statistics Branch, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.