The Annals of Probability
1974, Vol. 2, No. 4, 741–744

ON THE ACCURACY OF THE NORMAL APPROXIMATION FOR QUANTILES

BY R.-D. REISS

Mathematisches Institut der Universität Köln

It is well known that the sample p-quantile is asymptotically normally distributed. This paper presents an estimate for the accuracy of the normal approximation. The special intention of this paper is to obtain favorable bounds for moderate samples.

1. Introduction and summary. It is well known that the sample p-quantile is asymptotically normally distributed. This paper presents an estimate for the accuracy of the normal approximation. In view of applications our special intention is to obtain favorable bounds for moderate samples.

The sample p-quantile $x_{p,n}$ based on the ordered sample $x_{(1)} \leqq x_{(2)} \leqq \cdots \leqq x_{(n)}$ is defined by $x_{p,n} \equiv x_{(np)}$ if np is an integer and by $x_{p,n} \equiv x_{([np]+1)}$ otherwise. We shall write $\sigma_p \equiv (p(1-p))^{\frac{1}{2}}$. Φ denotes the standard normal distribution function. For each integer n denote by P^n the independent product of n identical probability measures P.

We shall prove Theorem 1.1. with the help of the Berry–Esséen theorem applied to independent binomial random variables. Note that the best upper bound for the universal constant C occuring in this theorem is given as 0.7975 by van Beeck (1972). Furthermore, Esséen (1956) proved that $C \geqq (3 + 10^{\frac{1}{2}})/6(2\pi)^{\frac{1}{2}}$.

THEOREM 1.1. *Let P be a probability measure with distribution function F such that*

$$F \text{ has a bounded second derivative on } R \text{ (the real line).} \tag{1.2}$$

Let $f \equiv F'$ and $||f'|| \equiv \sup\{|f'(x)| : x \in R\}$. Let $\xi_p \in R$ be such that

$$F(\xi_p) = p \qquad \text{and} \qquad f(\xi_p) > 0 \,. \tag{1.3}$$

Then

$$\sup_{t \in R} \left| P^n \left\{ \frac{n^{\frac{1}{2}} f(\xi_p)}{\sigma_p} (x_{p,n} - \xi_p) < t \right\} - \Phi(t) \right| \leqq n^{-\frac{1}{2}} \left[\frac{6||f'||\sigma_p}{10 f^2(\xi_p)} + \frac{||f'||^2}{f^4(\xi_p) n^{\frac{1}{2}}} + R_{p,n} \right] \tag{1.4}$$

where

$$R_{p,n} \equiv C \frac{1 - 2\sigma_p^{\,2} q_n^{\,2}}{\sigma_p q_n} + \frac{3(|1-2p| + ((\log n)/n)^{\frac{1}{2}})}{10 \sigma_p q_n^{\,2}}$$

(*whenever* $q_n \equiv [1 - \sigma_p^{-2}(|1-2p|((\log n)/n)^{\frac{1}{2}} + (\log n)/n)]^{\frac{1}{2}}$ *is defined*).

Received July 9, 1973; revised November 4, 1973.
AMS 1970 *subject classifications.* Primary 60F05, 62G35.
Key words and phrases. Sample quantiles, normal approximation.

Hence we always assume that $n \geqq 9$. In the case of $p = \frac{1}{2}$ q_n is defined for all $n \geqq 9$. Note that (1.2) and (1.3) imply that $p \in (0, 1)$ and that ξ_p is the unique p-quantile of P. Theorem 1.1 will be derived from

PROPOSITION 1.5. *Denote by Q the uniform distribution on* $[0, 1]$. *Then, for each* $p \in (0, 1)$:

$$\sup_{t \in R} \left| Q^n \left\{ \frac{n^{\frac{1}{2}}}{\sigma_p} (x_{p,n} - p) < t \right\} - \Phi(t) \right| \leqq n^{-\frac{1}{2}} R_{p,n} . \tag{1.6}$$

The methods of proof may be used to obtain an estimate of the accuracy of the normal approximation for the common distribution of several distinct quantiles.

2. Proofs.

(2.1) PROOF OF PROPOSITION 1.5. If $p_{n,t} \equiv p + t\sigma_p/n^{\frac{1}{2}} \notin (0, 1)$ the assertion holds trivially. For $t \geqq n^{\frac{1}{2}}(1 - p)/\sigma_p$ we may prove:

$$Q^n \left\{ \frac{n^{\frac{1}{2}}}{\sigma_p} (x_{p,n} - p) < t \right\} = Q^n\{x_{p,n} < 1\} = 1 .$$

Simple calculations yield

$$\Phi(t) \geqq 1 - 1/t(2\pi)^{\frac{1}{2}} \geqq 1 - \sigma_p/n^{\frac{1}{2}}(1 - p)(2\pi)^{\frac{1}{2}} .$$

As $C \geqq (2\pi)^{-\frac{1}{2}}$ and $(1 - 2\sigma_p{}^2 q_n{}^2)/\sigma_p q_n \geqq (1 - 2\sigma_p{}^2)/\sigma_p \geqq \sigma_p/(1 - p)$ (1.6) follows for $t \geqq n^{\frac{1}{2}}(1 - p)/\sigma_p$. In a similar way we may also prove that (1.6) holds for $t \leqq -n^{\frac{1}{2}} p/\sigma_p$. Hence we may assume that $p_{n,t} \in (0, 1)$.

Let $\sigma_{n,t} \equiv (p_{n,t}(1 - p_{n,t}))^{\frac{1}{2}}$, $q_{n,t} \equiv \sigma_{n,t}/\sigma_p = (1 - t(1 - 2p)/\sigma_p n^{\frac{1}{2}} - t^2/n)^{\frac{1}{2}}$ and $t_n = (\log n)^{\frac{1}{2}}/\sigma_p$.
We have

$$\begin{aligned} Q^n\{(n^{\frac{1}{2}}/\sigma_p)(x_{p,n} - p) < t\} &= Q^n\{x_{p,n} < p_{n,t}\} \\ &= Q^n\{\textstyle\sum_{i=1}^n 1_{(-\infty, p_{n,t})}(x_i) \geqq np\} \\ &= Q^n\{S_{n,t} \leqq t/q_{n,t}\} \end{aligned} \tag{2.2}$$

where $S_{n,t} \equiv -n^{-\frac{1}{2}}\sigma_{n,t}^{-1} \sum_{i=1}^n (1_{(-\infty, p_{n,t})}(x_i) - p_{n,t})$.
By (2.2)

$$\begin{aligned} &|Q^n\{(n^{\frac{1}{2}}/\sigma_p)(x_{p,n} - p) < t\} - \Phi(t)| \\ &\quad \leqq |Q^n\{S_{n,t} \leqq t/q_{n,t}\} - \Phi(t/q_{n,t})| + |\Phi(t/q_{n,t}) - \Phi(t)| . \end{aligned} \tag{2.3}$$

Suppose first that $|t| \leqq t_n$. Applying the Berry–Esséen theorem we obtain an upper bound for the first term on the right side of (2.3).

$$\begin{aligned} &|Q^n\{S_{n,t} \leqq t/q_{n,t}\} - \Phi(t/q_{n,t})| \\ &\quad \leqq C(1 - 2\sigma_{n,t}^2)/n^{\frac{1}{2}}\sigma_{n,t} = C(1 - 2\sigma_p{}^2 q_{n,t}^2)/n^{\frac{1}{2}}\sigma_p q_{n,t} \\ &\quad \leqq C(1 - 2\sigma_p{}^2 q_n{}^2)/n^{\frac{1}{2}}\sigma_p q_n . \end{aligned} \tag{2.4}$$

Next we derive an upper bound for the second term on the right side of (2.3).

Applying Taylor's theorem we obtain

$$q_{n,t} = 1 - (t(1-2p)/\sigma_p n^{\frac{1}{2}} + t^2/n)/2(1 - \vartheta(t(1-2p)/\sigma_p n^{\frac{1}{2}} + t^2/n))^{\frac{1}{2}}$$

for some $\vartheta \in (0, 1)$.
Since $q_{n,t} \geqq q_n$, and for all $|t| \leqq t_n$: $1 - \vartheta(t(1-2p)/\sigma_p n^{\frac{1}{2}} + t^2/n) \geqq q_n^2$ we have

$$\text{(2.5)} \qquad \begin{aligned} |t/q_{n,t} - t| &= |t(1 - q_{n,t})/q_{n,t}| \\ &\leqq t^2(|1-2p|/\sigma_p + t_n/n^{\frac{1}{2}})/2n^{\frac{1}{2}}q_n^2 . \end{aligned}$$

W.l.g. we may assume that $q_n > 0$. Then $1 - t_n|1-2p|/\sigma_p n^{\frac{1}{2}} - t_n^2/n > 0$ and $1 - t(1-2p)/\sigma_p n^{\frac{1}{2}} + t^2/n > 0$ for all $|t| \leqq t_n$. Therefore,

$$\text{(2.6)} \qquad q_{n,t}^2 = 1 - t(1-2p)/\sigma_p n^{\frac{1}{2}} - t^2/n < 2 .$$

(2.5) and (2.6) together imply

$$\text{(2.7)} \qquad \begin{aligned} |\Phi(t/q_{n,t}) - \Phi(t)| &\leqq \Phi'(t/2^{\frac{1}{2}})t^2(|1-2p|/\sigma_p + t_n|n^{\frac{1}{2}})/2n^{\frac{1}{2}}q_n^2 \\ &\leqq 3(|1-2p|/\sigma_p + t_n/n^{\frac{1}{2}})/10n^{\frac{1}{2}}q_n^2 \end{aligned}$$

since $\sup_{t \in R}(\Phi'(t/2^{\frac{1}{2}})t^2/2) \leqq 3/10$.
Hence by (2.3), (2.4) and (2.7) we have proved the inequality in (1.6) for all $|t| \leqq t_n$.
For $|t| \geqq t_n$ we apply a results on "exponential bounds" (see Loève (1963) page 255).
Let $t \leqq -t_n$. By (2.2) for $u = \frac{8}{5}(\log n)^{\frac{1}{2}}\sigma_{n,t}$

$$\text{(2.8)} \qquad \begin{aligned} Q^n\{(n^{\frac{1}{2}}/\sigma_p)(x_{p,n} - p) < t\} &\leqq Q^n\{-S_{n,t} \geqq |t|/q_{n,t}\} \\ &\leqq \exp(-|t|u/q_{n,t} + (u^2/2)(1 + u/2n^{\frac{1}{2}}\sigma_{n,t})) \\ &\leqq \tfrac{3}{2}n^{\frac{32}{25}} . \end{aligned}$$

And

$$\Phi(t) \leqq \exp(-t^2/2)/(2\pi)^{\frac{1}{2}}t \leqq 1/(2\pi)^{\frac{1}{2}}n^{\frac{8}{7}} .$$

Because of $R_{p,n} \geqq C/(1 - (\log n)/n)^{\frac{1}{2}}$ and $n \geqq 9$ this implies

$$|Q^n\{(n^{\frac{1}{2}}/\sigma_p)(x_{p,n} - p) < t\} - \Phi(t)| \leqq R_{p,n}n^{-\frac{1}{2}}$$

for all $t \leqq -t_n$. Since the same inequality also holds for all $t \geqq t_n$, (1.6) is proved.

(2.9) Proof of Theorem 1.1. Let $Z_{p,n} \equiv (n^{\frac{1}{2}}f(\xi_p)/\sigma_p(x_{p,n} - \xi_p)$ and $r_n \equiv n^{\frac{1}{2}}f^2(\xi_p)/\sigma_p\|f'\|$. First we assume that $|t| \leqq r_n$. Because of $P^n\{x_{p,n} < t\} = Q^n\{x_{p,n} < F(t)\}$ we obtain by (1.6) (with $u(n,t) = (n^{\frac{1}{2}}/\sigma_p)(F(\xi_p + (\sigma_p t/f(\xi_p)n^{\frac{1}{2}})) - p)$)

$$\text{(2.10)} \qquad \begin{aligned} &|P^n\{Z_{p,n} < t\} - \Phi(t)| \\ &\quad\leqq |Q^n\{(n^{\frac{1}{2}}/\sigma_p)(x_{p,n} - p) < u(n,t)\} - \Phi(u(n,t))| \\ &\qquad + |\Phi(u(n,t) - \Phi(t)| \\ &\quad\leqq R_{p,n}n^{-\frac{1}{2}} + \max\{|\Phi(t) - \Phi(t + t^2/2r_n)|, |\Phi(t) - \Phi(t - t^2/2r_n)|\} \\ &\quad\leqq R_{p,n}n^{-\frac{1}{2}} + 6\|f'\|\sigma_p/10n^{\frac{1}{2}}f^2(\xi_p) . \end{aligned}$$

Let $\mu \equiv F(\xi_p - f(\xi_p)/||f'||)$. By Chebychev's inequality and because of $p - \mu \geqq f^2(\xi_p)/2||f'||$

$$\begin{aligned} P^n\{Z_{p,n} < -r_n\} &= P^n\{x_{p,n} < \xi_p - f(\xi_p)/||f'||\} \\ &= P^n\{\textstyle\sum_{i=1}^n (1_{(-\infty,\xi_p - f(\xi_p)/||f'||)}(x_i) - \mu) \geqq nf^2(\xi_p)/2||f'||\} \\ &\leqq ||f'||^2/f^4(\xi_p)n\,. \end{aligned} \tag{2.11}$$

Furthermore,

$$\Phi(-r_n) \leqq \Phi'(-r_n)/r_n \leqq ||f'||\sigma_p/(2\pi)^{\frac{1}{2}}f^2(\xi_p)n^{\frac{1}{2}}\,.$$

This together with (2.11) implies for all $t \leqq -r_n$

$$|P^n\{Z_{p,n} < t\} - \Phi(t)| \leqq 6||f'||\sigma_p/10f^2(\xi_p)n^{\frac{1}{2}} + ||f'||^2/f^4(\xi_p)n\,.$$

Since the same inequality also holds for $t \geqq r_n$, (1.4) is proved.

Acknowledgment. The author wishes to thank Professor Dr. J. Pfanzagl for some useful conversations during the course of this work.

REFERENCES

[1] Esséen, C. G. (1956). A moment inequality with an application to the central limit theorem. *Skand. Aktuarietidskr.* **39** 160–170.

[2] Loève, M. (1963). *Probability Theory*, 3rd. ed. Van Nostrand, Princeton.

[3] Van Beeck, P. (1972). An application of Fourier methods to the problem of sharpening the Berry–Esséen inequality. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **23** 187–196.

Mathematisches Institut der Universität Köln
D-5000 Köln 41
Weyertal 86
Federal Republic of Germany