

2004

## Approximation for The Expectation of A Function of The Sample Mean

Rasul A. Khan  
Cleveland State University, R.KHAN@csuohio.edu

Follow this and additional works at: [https://engagedscholarship.csuohio.edu/scimath\\_facpub](https://engagedscholarship.csuohio.edu/scimath_facpub)

 Part of the [Mathematics Commons](#)

[How does access to this work benefit you? Let us know!](#)

---

### Repository Citation

Khan, Rasul A., "Approximation for The Expectation of A Function of The Sample Mean" (2004).  
*Mathematics Faculty Publications*. 163.  
[https://engagedscholarship.csuohio.edu/scimath\\_facpub/163](https://engagedscholarship.csuohio.edu/scimath_facpub/163)

This Article is brought to you for free and open access by the Mathematics and Statistics Department at EngagedScholarship@CSU. It has been accepted for inclusion in Mathematics Faculty Publications by an authorized administrator of EngagedScholarship@CSU. For more information, please contact [library.es@csuohio.edu](mailto:library.es@csuohio.edu).

# APPROXIMATION FOR THE EXPECTATION OF A FUNCTION OF THE SAMPLE MEAN

RASUL A. KHAN

Let  $\bar{X}_n$  be the mean of a random sample of size  $n$  from a distribution with mean  $\mu$  and variance  $\sigma^2$ . Under some conditions it is shown that  $Ef(\bar{X}_n) = f(\mu) + (\sigma^2/2n)f''(\mu) + O(n^{-2})$ , and  $\text{var}(f(\bar{X}_n)) = (\sigma^2/n)(f'(\mu))^2 + O(n^{-2})$ , where  $f$  is a continuous function with a suitable growth condition. This complements a result of Lehmann [(1991). *Theory of Point Estimation*. Wadsworth, California] and Cramér [(1946). *Mathematical Methods of Statistics*. Princeton University Press, Princeton, N.J.] for wider application. An illustrative example is given to show an application where the usual approximations do not apply.

## INTRODUCTION

Let  $X_1, X_2, \dots, X_n$  be iid random variables with mean  $\mu$  and variance  $\sigma^2$ , and set  $\bar{X}_n = (X_1 + \dots + X_n)/n$ . Let  $\mathcal{A} \subset \mathcal{R}$  be an interval such that  $P(X_1 \in \mathcal{A}) = 1$ , and let  $f$  be a real function on  $\mathcal{A}$ . Quite often one needs to find an approximation for  $Ef(\bar{X}_n)$  and  $\text{var}(f(\bar{X}_n))$ . Assuming the existence of some higher moments of  $X_1$ , Cramér (1946) and Lehmann (1991) (see also Bickel *et al.*, 2001, pp. 306–308) do provide such approximations under suitable conditions on  $f$ . However, as we point out, these approximations do not apply in certain cases. The usual approximation is given by Lehmann's Theorem 5.1 (cf. Lehmann, 1991, p. 106; Bickel *et al.*, 2001, p. 308). It requires that the derivatives  $f^{(i)}(x)$ ,  $1 \leq i \leq 4$ , exist for all  $x \in \mathcal{A}$  and  $|f^{(4)}(x)| \leq M$  for all  $x \in \mathcal{A}$  where  $M$  is a constant. Then under the assumption of fourth moment, the result is:

$$Ef(\bar{X}_n) = f(\mu) + \frac{\sigma^2}{2n}f''(\mu) + O(n^{-2}). \quad (1)$$

A similar result is given for  $\text{var}(f(\bar{X}_n))$ . A stronger approximation under a weaker condition on  $f$  is given by Cramér (1946). The approximations are the best possible under the given

conditions. However, the boundedness of the fourth derivative is the main reason why the above approximation may not be applicable in certain cases of interest. As an example, suppose that  $X_i$  is  $N(\mu, 1)$ , and let  $f(x) = e^x$ . Then  $Ee^{\bar{X}_n} = \exp(\mu + 1/2n)$ , and Eq. (1) obviously holds although Theorem 5.1 (cf. Lehmann, 1991, p. 106) cannot be applied. The result due to Cramér (1946, pp. 353–355) is much stronger but is not applicable to our example due to his second condition. In fact, Cramér's second condition (Cramér, 1946, p. 354) applied to  $f(\bar{X}_n)$  becomes  $|f(\bar{X}_n)| \leq Cn^p$  for some constant  $C$  and  $p > 0$  for all the values of  $X_i$ s. This condition is somewhat restrictive although it works for many examples. Consider again the preceding normal example with the same function  $f(x) = e^x$ . Let the observed values of  $X_i$  be  $x_i = 2^i$ ,  $1 \leq i \leq n$ . Clearly,  $\bar{X}_n = 2(2^n - 1)/n$ , and  $e^{\bar{X}_n} \geq 2^{n+1}/n$  if  $n \geq 3$ . Thus  $|e^{\bar{X}_n}| \not\leq Cn^p$  for all the values of  $X_i$ s even though Eq. (1) holds. So even Cramér's result may not apply in certain cases. The object of this note is to obtain the existing results for a suitable class of functions under the stronger assumption of moment generating function (m.g.f.). The assumption of m.g.f. allows a larger class of functions with weaker conditions on  $f$  for which Eq. (1) holds. The given result merely complements the usual approximations such as Eq. (1). The required conditions do not improve Lehmann's Theorem 5.1 or Cramér's result when applied to  $f(\bar{X}_n)$ . Similar results are possible for some order statistics.

## MAIN RESULT

In what follows, let  $\mathcal{F}$  denote a class of continuous functions on  $\mathcal{A}$  such that

$$|f(x)| = O(e^{\alpha|x|}) \quad \text{as } |x| \rightarrow \infty, \quad \text{for some } \alpha > 0. \quad (2)$$

Obviously, bounded functions belong to  $\mathcal{F}$  with  $\alpha = 0$ . Moreover, if  $\mathcal{A}$  is a closed finite interval, then  $f$  is bounded and such functions belong to  $\mathcal{F}$  with  $\alpha = 0$ , and 'as  $|x| \rightarrow \infty$ ' becomes unnecessary. In fact, bounded functions are in  $\mathcal{F}$  for  $\alpha \geq 0$ . The following two lemmas are needed in the sequel. The first lemma is due to Chernoff (1952, p. 495). Throughout the article we use  $I$  to denote the usual indicator function.

**LEMMA 1 (Chernoff)** *Let  $X_1, X_2, \dots, X_n$  be iid random variables with mean  $\mu$ , and assume that  $X_1$  has a finite m.g.f.  $\phi(\theta)$  for  $\theta \in J$  containing zero. Then for any  $\delta > 0$  there exist numbers  $\rho$  and  $\rho_1$  ( $0 < \rho, \rho_1 < 1$ ) such that*

$$P(\bar{X}_n - \mu \geq \delta) \leq \rho_1^n \quad \text{and} \quad P(|\bar{X}_n - \mu| \geq \delta) \leq 2\rho^n.$$

**LEMMA 2** *Let  $f \in \mathcal{F}$ , and let  $E|f(\bar{X}_n)| < \infty$ . Then under the conditions of Lemma 1:*

$$Ef(\bar{X}_n)I\{|\bar{X}_n - \mu| \geq \delta\} = O(1)(\rho^n + \rho_1^n) = O(n^{-2}).$$

*Proof* Since  $f \in \mathcal{F}$  and Eq. (2) holds, there exists a large finite number  $N$  and a constant  $C$  such that  $|f(x)| \leq C \exp(\alpha|x|) \forall |x| \geq N$ . However, this implies that  $|f(x)| \leq C \exp(\alpha|x|) \forall |x - \mu| \geq N_1$  where  $N_1 = N + |\mu|$ . Let  $\omega(y) = \{y: \delta \leq |y - \mu| \leq N_1\}$  and  $\bar{\omega}(y) = \{y: |y - \mu| > N_1\}$ . Clearly,

$$Ef(\bar{X}_n)I\{|\bar{X}_n - \mu| \geq \delta\} = Ef(\bar{X}_n)I(\omega(\bar{X}_n)) + Ef(\bar{X}_n)I(\bar{\omega}(\bar{X}_n)). \quad (3)$$

Since  $|f(\bar{X}_n)| \leq M$  on  $\omega(\bar{X}_n)$  for some constant  $M$  free of  $n$ , Lemma 1 implies

$$E|f(\bar{X}_n)|I(\omega(\bar{X}_n)) \leq MP(|\bar{X}_n - \mu| \geq \delta) \leq 2M\rho^n. \quad (4)$$

Also,  $|f(\bar{X}_n)| \leq C \exp(\alpha|\bar{X}_n|)$  on  $\bar{\omega}(\bar{X}_n)$ , and we have

$$E|f(\bar{X}_n)|I(\bar{\omega}(\bar{X}_n)) \leq CE \exp(\alpha|\bar{X}_n|)I(\bar{\omega}(\bar{X}_n)).$$

Moreover, Schwarz inequality gives

$$E \exp(\alpha|\bar{X}_n|)I(\bar{\omega}(\bar{X}_n)) \leq (E \exp(2\alpha|\bar{X}_n|)P(|\bar{X}_n - \mu| \geq N_1))^{1/2}. \quad (5)$$

Since  $\exp(2\alpha|\bar{X}_n|) \leq \exp(2\alpha\bar{X}_n) + \exp(-2\alpha\bar{X}_n)$ , and  $X_1$  has the m.g.f.  $\phi(\theta)$ , we obtain

$$E \exp(2\alpha|\bar{X}_n|) \leq \phi^n\left(\frac{2\alpha}{n}\right) + \phi^n\left(\frac{-2\alpha}{n}\right).$$

It is easy to see that

$$\phi^n\left(\frac{2\alpha}{n}\right) + \phi^n\left(\frac{-2\alpha}{n}\right) \rightarrow \exp(2\alpha\mu) + \exp(-2\alpha\mu) \text{ as } n \rightarrow \infty.$$

Hence  $E \exp(2\alpha|\bar{X}_n|) = O(1)$ , and Eq. (5) gives

$$E \exp(\alpha|\bar{X}_n|)I(\bar{\omega}(\bar{X}_n)) \leq \sqrt{O(1)\rho^n} = O(1)\rho^{n/2},$$

which, combined with Eqs. (3) and (4), proves the lemma. ■

The main result is given by the following theorem.

**THEOREM 1** *Let  $X_1, X_2, \dots, X_n$  be iid random variables with mean  $\mu$  and variance  $\sigma^2$ , and assume that  $X_1$  has a finite m.g.f. Let  $f$  be a continuous function on  $\mathcal{A}$  satisfying Eq. (2) (i.e.  $f \in \mathcal{F}$ ) where  $\mathcal{A}$  is an interval such that  $P(X_1 \in \mathcal{A}) = 1$ . Suppose that the first four derivatives of  $f$  are continuous in  $(\mu - \delta, \mu + \delta)$  for some  $\delta > 0$ . Then*

$$Ef(\bar{X}_n) = f(\mu) + \frac{\sigma^2}{2n} f''(\mu) + O(n^{-2}), \quad (6)$$

and

$$\text{var}(f(\bar{X}_n)) = \frac{\sigma^2}{n} (f'(\mu))^2 + O(n^{-2}). \quad (7)$$

*Proof* We use primes for the first two derivatives, and use  $f^{(k)}(\mu)$  to denote the  $k$ th derivative of  $f$  at  $\mu$  with the convention that  $f^{(0)}(\mu) = f(\mu)$ . Let  $P(x) = \sum_{k=0}^4 ((x - \mu)^k / k!) f^{(k)}(\mu)$  be the Taylor polynomial, and consider the Taylor expansion of  $f$  in  $(\mu + \delta, \mu - \delta)$  as

$$\begin{aligned} f(x) &= P(x) + \frac{(x - \mu)^4}{4!} (f^{(4)}(\mu + \eta(x - \mu)) - f^{(4)}(\mu)), \quad 0 \leq \eta \leq 1, \\ &= P(x) + R(x). \end{aligned}$$

It is well known that

$$E(\bar{X}_n - \mu)^3 = O(n^{-2}) \quad \text{and} \quad E(\bar{X}_n - \mu)^4 = O(n^{-2}),$$

and hence

$$EP(\bar{X}_n) = f(\mu) + \frac{\sigma^2}{2n} f''(\mu) + O(n^{-2}). \quad (8)$$

Let  $0 < \delta_1 < \delta$  and set  $T_n = Ef(\bar{X}_n)I\{|\bar{X}_n - \mu| < \delta_1\}$ . By Lemma 2 we have

$$Ef(\bar{X}_n) = T_n + Ef(\bar{X}_n)I\{|\bar{X}_n - \mu| \geq \delta_1\} = T_n + O(n^{-2}). \quad (9)$$

Clearly,

$$\begin{aligned} T_n &= EP(\bar{X}_n)I\{|\bar{X}_n - \mu| < \delta_1\} + ER(\bar{X}_n)I\{|\bar{X}_n - \mu| < \delta_1\} \\ &= EP(\bar{X}_n) - EP(\bar{X}_n)I\{|\bar{X}_n - \mu| \geq \delta_1\} + ER(\bar{X}_n)I\{|\bar{X}_n - \mu| < \delta_1\}. \end{aligned}$$

Since  $P(x) \in \mathcal{F}$ , by Lemma 2 we have

$$T_n = EP(\bar{X}_n) + O(n^{-2}) + ER(\bar{X}_n)I\{|\bar{X}_n - \mu| < \delta_1\}. \quad (10)$$

We now consider the remainder term. Let  $Z_n = \mu + \eta(\bar{X}_n - \mu)$ ,  $0 \leq \eta = \eta_n \leq 1$ . Obviously,  $|\bar{X}_n - \mu| < \delta_1$  implies  $|\bar{X}_n - \mu| \leq \delta_1$ , and  $Z_n$  is in the closed interval  $[\mu - \delta_1, \mu + \delta_1]$ . Since  $f^{(4)}(x)$  is continuous in  $[\mu - \delta_1, \mu + \delta_1]$ , hence  $\phi(\bar{X}_n) = (1/4!)|f^{(4)}(Z_n) - f^{(4)}(\mu)|$  remains bounded by some constant  $K$ . Thus we obtain

$$E|R(\bar{X}_n)|I\{|\bar{X}_n - \mu| < \delta_1\} \leq KE(\bar{X}_n - \mu)^4 = O(n^{-2}).$$

This, combined with Eqs. (8)–(10), proves Eq. (6). Moreover, since  $\text{var}(f(\bar{X}_n)) = Ef^2(\bar{X}_n) - (Ef(\bar{X}_n))^2$ , Eq. (6) applied to  $\psi(x) = f^2(x)$  leads to Eq. (7).  $\blacksquare$

*Remark 1* In Theorem 1 it is not necessary to assume the existence and continuity of the first four derivatives if one is content with a slightly weaker result. Suppose we assume that the first two derivatives of  $f$  are continuous in  $(\mu - \delta, \mu + \delta)$  for some  $\delta > 0$ . Then repeating the same proof with a second-order polynomial one can show that Eqs. (6) and (7) in Theorem 1 hold except that  $O(n^{-2})$  is to be replaced by  $o(n^{-1})$  which is implied by a refinement of the preceding argument.

*Remark 2* In general, the growth condition in the class  $\mathcal{F}$  cannot be improved. To see this, let  $X_1, \dots, X_n$  be iid  $\chi^2$  random variables with two degrees of freedom. Clearly,  $S_n = X_1 + \dots + X_n$  has an  $\chi^2$ -distribution with  $2n$  degrees of freedom having density  $g_n(x) = K x^{n-1} e^{-x/2}$ ,  $x \geq 0$ , where  $K = 1/2^n \Gamma(n)$ . If  $f(x) = e^{x^\alpha}$  ( $\alpha > 1$ ), then

$$Ef(\bar{X}_n) = K \int_0^\infty x^{n-1} \exp\left(\frac{x(x^{\alpha-1} - (1/2)n^\alpha)}{n^\alpha}\right) dx = +\infty \quad \forall n \geq 1.$$

*Remark 3* Theorem 1 is complementary to the existing result of the type (1) under local conditions but stronger assumption of the m.g.f. to gain a larger class of functions  $\mathcal{F}$ . However, such an advantage would be lacking without it. For example, let  $X$  be a continuous random variable with density  $f(x) = (8/(3\pi(1 + (x - \mu)^2)^3))$ ,  $-\infty < x < \infty$ . It can be verified that  $EX = \mu$ ,  $\sigma^2 = 1/3$ . In this case Theorem 1 does not apply due to lack of m.g.f. while the approximation (1) still remains valid.

The following is an illustrative example where neither Lehmann's nor Cramér's theorem applies.

*Example 1* Let  $X_1, X_2, \dots, X_n$  be a random sample from a Poisson distribution with mean  $\mu > 0$ . Let  $f(x) = \sqrt{x}$ ,  $x > 0$ , and  $\bar{X}_n = (X_1 + \dots + X_n)/n$ . Since Lemmas 1 and 2 are applicable to the Poisson distribution, and  $f = \sqrt{x} \in \mathcal{F}$ , Theorem 1 applies. It is easily verified that

$$E\sqrt{\bar{X}_n} = \sqrt{\mu} - \frac{1}{8n\sqrt{\mu}} + O(n^{-2}),$$

and

$$\text{var}(\sqrt{\bar{X}_n}) = \frac{1}{4n} + O(n^{-2}).$$

Cramér (1946, p. 358) considers  $s/\bar{x}$  as an estimator of  $\sigma/\mu$ , the coefficient of variation (CV) and he notes that his theorem does not apply for the normal distribution. Thus his approximations for the mean and the variance of the estimator of the CV are based on the positively truncated normal distribution with the added condition that  $\sigma/\mu$  be fairly small. The next example discusses this problem in the normal case.

*Example 2* Let  $X_1, X_2, \dots, X_n$  be iid  $N(\mu, \sigma^2)$  random variables, and let  $\bar{X}_n = (\sum_1^n X_i)/n$  and  $s_n^2 = \sum_1^n (X_i - \bar{X}_n)^2/(n-1)$  be the respective unbiased estimates of  $\mu$  and  $\sigma^2$ . Consider estimation of the CV  $\lambda = \sigma/\mu$ . It is natural to use  $s_n/\bar{X}_n$  as an estimator of  $\lambda$ . Unfortunately,  $E(s_n/\bar{X}_n)$  does not exist even when  $\mu \neq 0$ . However, it is possible to modify the estimator to overcome this deficiency. Suppose that the sign of  $\mu$  is known, and therefore there is no loss of generality in assuming that  $\mu > 0$ . To modify the estimator, let  $\alpha_n = 1/n^2$  and define  $\hat{\lambda} = s_n/\sqrt{(\alpha_n + \bar{X}_n^2)}$ . Let  $f_n(x) = 1/\sqrt{(\alpha_n + x^2)}$ . Clearly,  $E\hat{\lambda} = Es_n E f_n(\bar{X}_n)$ . It is well known that

$$Es_n = \frac{\Gamma(n/2)}{\Gamma((n-1)/2)} \sqrt{\frac{2}{n-1}} \sigma.$$

Using the asymptotic expansion of gamma function it is easy to see that

$$Es_n = \left(1 - \frac{1}{4n} + O(n^{-2})\right) \sigma. \quad (11)$$

Even though  $f$  depends on  $n$  through  $\alpha_n = 1/n^2$ , it is easily seen that Theorem 1 is still applicable, and

$$Ef_n(\bar{X}_n) = f_n(\mu) + \frac{\sigma^2}{2n} f_n''(\mu) + O(n^{-2}).$$

Moreover, it can be verified that

$$Ef_n(\bar{X}_n) = \frac{1}{\mu} + \frac{\sigma^2}{n\mu^3} + O(n^{-2}), \quad (12)$$

and hence it follows from Eqs. (11) and (12) that

$$E\hat{\lambda} = \frac{\sigma}{\mu} + n^{-1} \left( \frac{\sigma^3}{\mu^3} - \frac{\sigma}{4\mu} \right) + O(n^{-2}).$$

Similarly, we have

$$Ef_n^2(\bar{X}_n) = \frac{1}{\mu^2} + \frac{3\sigma^2}{n\mu^4} + O(n^{-2}).$$

Since  $\text{var}(\hat{\lambda}) = E s_n^2 E f_n^2(\bar{X}_n) - (E \hat{\lambda})^2 = \sigma^2 E f_n^2(\bar{X}_n) - (E \hat{\lambda})^2$ , we obtain

$$\text{var}(\hat{\lambda}) = \frac{\sigma^4}{n\mu^4} + \frac{\sigma^2}{2n\mu^2} + O(n^{-2}) = \frac{\sigma^2}{2n\mu^2} \left( 1 + \frac{2\sigma^2}{\mu^2} \right) + O(n^{-2}).$$

Thus

$$E \hat{\lambda} \doteq \frac{\sigma}{\mu} \quad \text{and} \quad \text{var}(\hat{\lambda}) \doteq \frac{\sigma^2}{2n\mu^2} \left( 1 + \frac{2\sigma^2}{\mu^2} \right),$$

as noted by Cramér (1946, p. 358) for positively truncated normal sample under the condition that  $\sigma/\mu$  be fairly small. Of course, we have shown it for a normal distribution without the restriction that  $\sigma/\mu$  be fairly small.

Similar techniques can be used to obtain improved asymptotic approximations for functions of order statistics as long as the exponential rates for tail probabilities are valid. This is illustrated by the following example.

*Example 3* Let  $X_1, X_2, \dots, X_n$  be iid random variables with uniform distribution on  $(0, \theta)$  ( $\theta > 0$ ), and let  $X_{(n)} = \max(X_1, \dots, X_n)$ . Then the probability density function of  $X_{(n)}$  is  $g_n(x) = (n/\theta)(x/\theta)^{n-1}$ ,  $0 < x < \theta$ , and  $= 0$ , otherwise. If  $0 < \delta < \theta$ , it is easy to see that  $P(|X_{(n)} - \theta| \geq \delta) = (1 - (\delta/\theta))^n = \gamma^n$ . One can easily verify that

$$E(X_{(n)} - \theta) = -\frac{\theta}{n+1} = -\frac{\theta}{n} + \frac{\theta}{n^2} + O(n^{-3}),$$

and

$$E(X_{(n)} - \theta)^2 = \frac{2\theta^2}{(n+1)(n+2)} = \frac{2\theta^2}{n^2} + O(n^{-3}).$$

Then using the above tail probability and under the same conditions on  $f$  as in Theorem 1, it can be shown that

$$E f(X_{(n)}) = f(\theta) - \frac{\theta}{n} f'(\theta) + \frac{1}{n^2} (\theta f'(\theta) + \theta^2 f''(\theta)) + O(n^{-3}).$$

Similarly,  $\text{var}(f(X_{(n)})) = (\theta^2/n^2)(f'(\theta))^2 + O(n^{-3})$ .

## References

- Bickel, P. J. and Doksum, K. A. (2001). *Mathematical Statistics*, Vol. I. Prentice Hall, New Jersey.
- Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.*, **23**, 493–507.
- Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press, Princeton, N.J.
- Lehmann, E. L. (1991). *Theory of Point Estimation*. Wadsworth, California.