Journal of Statistics: Advances in Theory and Applications
Volume 9, Number 2, 2013, Pages 99-103

DISTRIBUTION OF SAMPLE MEDIAN

MOON JUNG CHO and EUNGCHUN CHO

U. S. Bureau of Labor Statistics
Office of Survey Methods Research
2 Massachusetts Avenue NE
Washington, DC 20212
USA
e-mail: Cho.Moon@bls.gov

Seoul National University
Seoul, Korea

Abstract

Probability density of the median of samples of even size is given. Examples from populations with uniform distribution and exponential distribution are presented.

1. Introduction

The probability density function (pdf) of the median of the samples of size $(2k+1)$ from a population with pdf $f(x)$ is given by

$$g(x) = (2k+1)\binom{2k}{k} f(x)F(x)^k(1-F(x))^k, \tag{1}$$

2010 Mathematics Subject Classification: 60G17, 62D05.

Keywords and phrases: sampling, sample median, probability density, samples of even size.

Received June 19, 2013

where $f(x)dx = dF(x)$. See [1]. It can be rewritten as

$$g(x) = \frac{f(x)F(x)^k(1 - F(x))^k}{B(k+1,\ k+1)}, \tag{2}$$

where $B(k+1,\ k+1) = \Gamma(k+1)\Gamma(k+1)/\Gamma(2k+2)$. An asymptotic distribution $N(m, 1/4f(m)^2 n)$, the normal distribution with parameters m (the median of the population) and $1/4f(m)^2 n$, is given for large sample size n. See [1]. But it is derived from $g(x)$ for samples of odd size. The corresponding $g(x)$ for samples of even size is not available in the literature.

2. Median of Samples of Size $2k$

We give the exact pdf of the sample median for $n = 2k$.

Theorem 1. *Let P be a population with pdf $f(x)$. The pdf of the median of samples of size $2k$ from P is given by*

$$g(x) = \frac{4k}{B(k,\ k)} \int_0^\infty f(x-h)f(x+h)F(x-h)^{k-1}(1 - F(x+h))^{k-1} dh. \tag{3}$$

Proof. Let m be the median of a sample $\{x_1, x_2, \ldots, x_{2k}\}$. Then there exist, say x_i and x_j, in the sample such that $x_i = m - h$ and $x_j = m + h$ for some $h \geq 0$ and $(k-1)$ elements less than or equal to x_i and the rest greater than or equal to x_j. The probability for x_i to be in the interval $[m-h, m-h+dx]$ and for $(k-1)$ elements to be less than or equal to x_i is $f(m-h)F(m-h)^{k-1}dx$. Similar argument for x_j shows the probability for m to be in the interval $[x, x+dx]$ is proportional to the integral

$$I = \int_0^\infty f(x-h)f(x+h)F(x-h)^{k-1}(1 - F(x+h))^{k-1} dh.$$

Counting the number of all corresponding arrangements of the elements in the sample, we see the probability

$$\Pr(m \in [x,\ x + dx]) = 2k^2 \binom{2k}{k} I\, dx.$$

Substituting the factor $2k^2\binom{2k}{k}$ by $4k/B(k,\ k)$, we have

$$g(x) = \frac{4k}{B(k,\ k)} \int_0^{\infty} f(x - h) f(x + h) F(x - h)^{k-1} (1 - F(x + h))^{k-1}\, dh.$$

Remark. The integral in the formula does not have a closed form in general and the subsequent estimation of the expected value, variance, and higher moments requires quadrature method. However, the following examples give relatively simple forms of $g(x)$, and direct evaluation of the integral is possible.

Example 1. Uniform distribution on $[0, 1]$, sample size $2k$. If the population pdf $f(x) = I_{[0,1]}$, then

$$g(x) = \begin{cases} \dfrac{4k}{B(k, k)} \displaystyle\int_0^{\infty} I_{[0,1]}(x - h) I_{[0,1]}(x + h)(x - h)^{k-1}(1 - x - h)^{k-1}\, dh, & \text{for } 0 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

The expected value of sample median $\int_0^1 x g(x) dx = \frac{1}{2}$ for all $k \ge 1$. For $k = 1$, we have

$$g(x) = \begin{cases} 4 \min(x,\ 1 - x), & \text{for } 0 \le x \le 1, \\ 0, & \text{otherwise,} \end{cases}$$

and the variance

$$\int_0^1 (x - \frac{1}{2})^2 g(x) dx = \int_0^{1/2} (x - \frac{1}{2})^2 4x\, dx + \int_{1/2}^1 (x - \frac{1}{2})^2 4(1 - x) dx$$

$$= \frac{1}{24},$$

as expected. For $k = 2$, we have

$$g(x) = \begin{cases} 8\alpha_x(2\alpha_x^2 - 3\alpha_x + 6x - 6x^2), & \text{for } 0 \le x \le 1, \\ 0, & \text{otherwise,} \end{cases}$$

where $\alpha_x = \min(x, 1-x)$. The variance is $1/30$. The asymptotic distribution $N(1/2, 1/8k)$ overestimates the variance by three times for $k = 1$ and by two times for $k = 2$.

Example 2. Exponential distribution, sample size $2k$. Let the population pdf

$$f(x) = \begin{cases} e^{-x}, & \text{for } 0 \le x, \\ 0, & \text{otherwise.} \end{cases}$$

The population median is log 2. It follows from the theorem that

$$g(x) = \begin{cases} \dfrac{4k}{B(k,k)} \displaystyle\int_0^x e^{-x+h}(1 - e^{-x+h})^{k-1}(e^{-x-h})^k dh, & \text{for } 0 \le x, \\ 0, & \text{otherwise.} \end{cases}$$

For $k = 1$, we have

$$g(x) = \begin{cases} 4xe^{-2x}, & \text{for } 0 \le x, \\ 0, & \text{otherwise,} \end{cases}$$

the expected value

$$EV = \int_0^\infty 4x^2 e^{-2x} dx = 1,$$

and the variance

$$\text{Var} = \int_0^\infty (x-1)^2 4xe^{-2x} dx = \frac{1}{2}.$$

For $k = 2$, we have

$$g(x) = \begin{cases} 48\,(e^{x} - x - 1)e^{-4x}, & \text{for } 0 \leq x, \\ 0, & \text{otherwise,} \end{cases}$$

the expected value is $\frac{5}{6}$ and the variance $\frac{17}{72}$. The asymptotic distribution $N(\log 2,\, 1/2k)$ has variance $1/4$ for $k = 2$.

Acknowledgement

The authors thank John Eltinge for his many helpful comments on sample median and the referee for the thoughtful review. The views expressed in this paper are those of the authors and do not necessarily reflect the policies of the U.S. Bureau of Labor Statistics. Eungchun Cho's work at Seoul National University was supported by the Korea Research Foundation and the Korean Federation of Science and Technology Societies Grant funded by the South Korean Government (MOEHRD, Basic Research Promotion Fund).

Reference

[1] M. G. Kendall and A. Stuart, Advanced Theory of Statistics, Volume 1, Distribution Theory, 3rd Edition, pp 325-326, Hafner Publishing, New York, 1969.

■