UNIVERSITY OF LONDON

Goldsmiths
UNIVERSITY OF LONDON

The Sample Median Theorem [1]

Math Notes | Larry Cui

April 2, 2022

The Central Limit Theorem is one of the gems of probability. Another important theorem is about the sample median. It says that the density of the median of a sample, if the sample size goes large enough, will also follow a normal distribution. This note will explore the mean and variance of the sample median.

1 Background: Order Statistics

Suppose the random variables $X_1, X_2, \ldots, X_n$ constitute a sample of size n from an infinite population with continuous density. We re-order the variables from smallest to largest, and rename them so that Y_1 is the smallest, Y_n is the largest. Y_r is call the r^{th} **order statistic** of the sample. We know the pdf of the order statistic as follows.

Order Statistic pdf For a random sample of size n from an population having continuous pdf $f(y)$ and cdf $F(y)$, the probability density of $Y_{\min}$ (the minimum), $Y_{\max}$ (the maximum) and Y_r (any r^{th} variable), the pdf is:

$$
\begin{aligned}
f_{Y_{\min}} &= nf(y)F(y)^{n-1} \\
f_{Y_{\max}} &= nf(y)(1-F(y))^{n-1} \\
f_{Y_r} &= \frac{n!}{(r-1)!1!(n-r)!}F(y)^{r-1}(1-F(y))^{n-r}f(y)
\end{aligned}
$$

Proof: We omit the proofs for $Y_{\min}$, $Y_{\max}$ and Y_r. But with following illustration, readers are easy to understand why pdf takes the form like that for Y_r:

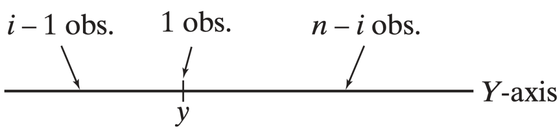

Figure 1: r^{th} order pdf

[1] This note is based on Steven J. Miller's paper: *The Probability Lifesaver: Order Statistics and the Median Theorem.*

2 Description of Sample Median Theorem

For a sample of odd size ($n = 2m+1$), the sample median is defined as Y_{m+1}. For a sample of even size ($n = 2m$), the sample median is defined as $(Y_m + Y_{m+1})/2$. Without losing generosity, we use Y_{m+1} as median for further discussion.

Compared to referring to μ as the mean of a population, we use $\tilde{\mu}$ as reference to the population median, and $\tilde{y}$ to the sample median value.

Let's assume the population has a continuous density. By definition of the median, we know that:

$$\int_{-\infty}^{\tilde{\mu}} f(y)\,dy = \frac{1}{2} \qquad \text{and} \qquad F(\tilde{\mu}) = \frac{1}{2}$$

Sample Median Theorem

Let a sample of size n, $n = 2m+1$, be taken from an infinite population with a density function $f(y)$ that is non-zero at the population median $\tilde{\mu}$ and continuously differentiable in a neighborhood of $\tilde{\mu}$. The sample median $\tilde{y}$ distribution is approximately normal with:

(1) Mean: $\tilde{\mu}$

(2) Variance: $1/8f(\tilde{\mu})^2m$

3 Intermediate Step to the Proof of Median Theorem

We use $g(\tilde{y})$ to denote the density function for sample median, though we don't know if it's a normal distribution or not yet. As mentioned above, the median density is simply the $(m+1)^{\text{th}}$ order statistic:

$$g(\tilde{y}) = \frac{(2m+1)!}{m!m!}[F(\tilde{y})]^m f(\tilde{y})[1-F(\tilde{y})]^m \tag{1}$$

We have Stirling's formula ($n! = n^n e^{-n}\sqrt{2\pi n}$) to substitute the factorial part of equation (1):

$$\begin{aligned} \frac{(2m+1)!}{m!m!} &= \frac{(2m+1)(2m)!}{(m!)^2} \\ &= \frac{(2m+1)(2m)^{2m}e^{-2m}\sqrt{2\pi 2m}}{(m^m e^{-m}\sqrt{2\pi m})^2} \\ &= \frac{(2m+1)4^m}{\sqrt{\pi m}} \end{aligned}$$

As a result, we have a simplified function below,

$$g(\tilde{y}) = \frac{(2m+1)4^m f(\tilde{y})}{\sqrt{\pi m}}[F(\tilde{y})]^m[1-F(\tilde{y})]^m \tag{2}$$

In this intermediate step, we need to prove that the sample median $\tilde{y}$ is asymptomatically close to the population median $\tilde{\mu}$. A formal way to say this:

Lemma Population density $f(y)$ is continuously differentiable including in some neighborhood of $\tilde{\mu}$. Then for any $c > 0$, we have

$$\lim_{m\to\infty} \text{Prob}(|\tilde{y} - \tilde{\mu}| \geqslant c) = 0$$

This Lemma is equivalent to say:

$$\lim_{m\to\infty} \text{Prob}(\tilde{y} \leqslant \tilde{\mu} - c) = 0 \qquad \text{and} \qquad \lim_{m\to\infty} \text{Prob}(\tilde{y} \geqslant \tilde{\mu} + c) = 0$$

We will work on the first statement, the second is the same as long as we prove the first one.

Take a second look at the factor in equation (2), $([F(\tilde{y})][1-F(\tilde{y})])^m$, we can see two features of the base:

(1) it has a maximum of $1/4$ when $F(\tilde{y}) = 1/2$. But since $F(\tilde{\mu}) = 1/2$, it means $F(\tilde{\mu})$ sees its maximum.

(2) when $F(\tilde{y})$ gets closer to $F(\tilde{\mu})$ from left, the base is increasing, so are $F(\tilde{y})$ and $\tilde{y}$ itself, since $F(y)$ is a cumulative function of density.

If we let $\tilde{y} \leqslant \tilde{\mu} - c$,

$$([F(\tilde{y})][1-F(\tilde{y})])^m \leqslant ([F(\tilde{\mu}-c)][1-F(\tilde{\mu}-c)])^m < ([F(\tilde{\mu})][1-F(\tilde{\mu})])^m = \frac{1}{4^m}$$

and if we let $[F(\tilde{\mu}-c)][1-F(\tilde{\mu}-c)] = \alpha/4$, where $\alpha < 1$,

$$([F(\tilde{y})][1-F(\tilde{y})])^m \leqslant \left(\frac{\alpha}{4}\right)^m \leqslant \frac{1}{4^m}$$

We now look at the probability that $\tilde{y}$ is less than or at most $\tilde{\mu} - c$,

$$\begin{aligned}
\text{Prob}(\tilde{y} \leqslant \tilde{\mu} - c) &= \int_{\infty}^{\tilde{\mu}-c} g(\tilde{y})\, d\tilde{y} \\
&= \int_{\infty}^{\tilde{\mu}-c} \frac{(2m+1)4^m f(\tilde{y})}{\sqrt{\pi m}} [F(\tilde{y})]^m [1-F(\tilde{y})]^m \, d\tilde{y} \\
&< \frac{(2m)4^m}{\sqrt{m}} \int_{\infty}^{\tilde{\mu}-c} f(\tilde{y}) \left(\frac{\alpha}{4}\right)^m d\tilde{y} && \triangleright : 2m < (2m+1)/\sqrt{\pi} \\
&< 2\alpha^m \sqrt{m} \int_{\infty}^{\tilde{\mu}} f(\tilde{y})\, d\tilde{y} \\
&< \alpha^m \sqrt{m}
\end{aligned}$$

Since $\alpha < 1$, when $m \to \infty$, we have $\text{Prob}(\tilde{y} \leqslant \tilde{\mu} - c)$ go to zero. This proves that $\tilde{y}$ converges to $\tilde{\mu}$ when m goes large.

4 Final Step

Now let's go back to equation (2). We use the Taylor series expansion on $F(\tilde{y})$ about median point $\tilde{\mu}$, and with a number c that $\tilde{\mu} < c < \tilde{y}$:

$$F(\tilde{y}) = F(\tilde{\mu}) + f(\tilde{\mu})(\tilde{y} - \tilde{\mu}) + \frac{f'(c)}{2}(\tilde{y} - \tilde{\mu})^2$$

We don't know the value of $f'(c)$, but we do know that when m goes large, $\tilde{y}$ converges to $\tilde{\mu}$, which makes $(\tilde{y} - \tilde{\mu})^2$ so small that the last part of the above equation can be neglected without losing generosity.

$$F(\tilde{y}) \circeq \frac{1}{2} + f(\tilde{\mu})(\tilde{y} - \tilde{\mu})$$

Substitute it into equation (2), we get

$$g(\tilde{y}) = \frac{(2m+1)4^m f(\tilde{y})}{\sqrt{\pi m}} \left[\frac{1}{2} + f(\tilde{\mu})(\tilde{y} - \tilde{\mu})\right]^m \left[1 - \frac{1}{2} - f(\tilde{\mu})(\tilde{y} - \tilde{\mu})\right]^m$$

$$g(\tilde{y}) = \frac{(2m+1)4^m f(\tilde{y})}{\sqrt{\pi m}} \left[\frac{1}{4} - (f(\tilde{\mu})(\tilde{y} - \tilde{\mu}))^2\right]^m$$

$$g(\tilde{y}) = \frac{(2m+1) f(\tilde{y})}{\sqrt{\pi m}} \left[1 - \frac{4m(f(\tilde{\mu})(\tilde{y} - \tilde{\mu}))^2}{m}\right]^m$$

The bracket part looks familiar, it's a variant of exponential function indeed,

$$\lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n = e^x$$

So we have the exponential approximation of the bracket part when m goes large enough:

$$\left[1 - \frac{4m(f(\tilde{\mu})(\tilde{y} - \tilde{\mu}))^2}{m}\right]^m \circeq \exp\Big(-4mf(\tilde{\mu})^2(\tilde{y} - \tilde{\mu})^2\Big) = \exp\left(-\frac{(\tilde{y} - \tilde{\mu})^2}{1/4mf(\tilde{\mu})^2}\right)$$

Again, we use the feature that $\tilde{y} - \tilde{\mu} = 0$ when $m \to \infty$, we are safe to say $f(\tilde{y}) \circeq f(\tilde{\mu})$. Have $f(\tilde{\mu})$ as a substitution, the density function of sample median becomes:

$$g(\tilde{y}) = \frac{(2m+1) f(\tilde{\mu})}{\sqrt{\pi m}} \exp\left(-\frac{(\tilde{y} - \tilde{\mu})^2}{1/4mf(\tilde{\mu})^2}\right)$$

This function is quite similar to the normal distribution function:

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\frac{(y_i - \mu)^2}{\sigma^2}\right]$$

We can conclude now that the $g(\tilde{y})$ function is almost normal, with the mean $\tilde{\mu}$, and variance $\sigma^2 = 1/(8mf(\tilde{\mu})^2)$. If we compare the non-exponential part, we will get another version of σ:

$$\frac{(2m+1) f(\tilde{\mu})}{\sqrt{\pi m}} = \frac{1}{\sqrt{2\pi}\sigma}$$

$$\sigma^2 = \left(\frac{\sqrt{\pi m}}{(2m+1) f(\tilde{\mu})\sqrt{2\pi}}\right)^2$$

$$= \frac{m}{2(2m+1)^2 f(\tilde{\mu})^2}$$

Apparently, when m goes large, two versions of σ converges. Proof completed.