

13.4 Mean or Median?

WORKSHEET

The question

The random variables $X_1, X_2, \dots, X_{2n+1}$ are independently and uniformly distributed on the interval $0 \leq x \leq 1$. The random variable Y is defined to be the median of $X_1, X_2, \dots, X_{2n+1}$. Given that the probability density function of Y is $g(y)$, where

$$g(y) = \begin{cases} ky^n(1-y)^n & \text{if } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

use the result

$$\int_0^1 y^r (1-y)^s dy = \frac{r!s!}{(r+s+1)!}$$

to show that $k = (2n+1)!/(n!)^2$, and evaluate $E(Y)$ and $\text{Var}(Y)$. Hence show that, for any given positive number d , the inequality

$$P\left(|Y - \frac{1}{2}| < \frac{d}{\sqrt{n}}\right) < P\left(|\bar{X} - \frac{1}{2}| < \frac{d}{\sqrt{n}}\right)$$

holds provided n is large enough, where $\bar{X}$ is the mean of $X_1, X_2, \dots, X_{2n+1}$.

[You may assume that Y and $\bar{X}$ are normally distributed for large n .]

Hints

Try the question on your own first. If you get stuck, read the hints *one at a time*, and go back to the question after each one. Hints marked *Pause and try* are the key steps.

- ? **Ask yourself** Why is the density of the median proportional to $y^n(1-y)^n$? (For Y to be near y , how many of the $2n+1$ values must be below y , and how many above?)

STAGE 2 Mean and variance

- ? **Ask yourself** Compute $E(Y^2) = k \int_0^1 y^{n+2}(1-y)^n dy$ using the beta integral, and simplify the factorials.

STAGE 3 Comparing with the mean

- ? **Ask yourself** What are the mean and variance of $\bar{X}$, the average of $2n+1$ independent uniform variables?

- ? **Ask yourself** If Y and $\bar{X}$ are normal with mean $\frac{1}{2}$ and standard deviations σ_Y and $\sigma_{\bar{X}}$, write both probabilities in terms of the standard normal distribution function Φ .

Try this next

- 1 The median's density.** Derive the density $\frac{(2n+1)!}{(n!)^2} y^n(1-y)^n$ by counting: one value in $[y, y+\delta y]$, n below and n above.
- 2 Three observations.** For $n=1$ (three uniform observations), find $\text{Var}(Y)$ and $\text{Var}(\bar{X})$ exactly. Which is smaller?
- 3 The midrange.** Another estimator of the centre is the midrange $\frac{1}{2}(\max + \min)$. Show that for uniform data its variance is $\frac{1}{2(m+1)(m+2)}$ with $m=2n+1$ observations. How does it compare with the mean? (Compare Section 13.3.)
- 4 An outlier.** Eleven measurements are 9.8, 9.9, 10.0, 10.0, 10.1, 10.1, 10.2, 10.2, 10.3, 10.4 and 104 (a typing error). Compare the mean and median.

Further reading

- **John Rice, *Mathematical Statistics and Data Analysis*** (Duxbury), on order statistics and efficiency.

- ▶ **Rand Wilcox**, *Fundamentals of Modern Statistical Methods* (Springer). Why robust methods matter, for non-specialists.
- ▶ **Stephen Stigler**, *The Seven Pillars of Statistical Wisdom* (Harvard University Press). Aggregation, the mean and the median, in historical perspective.

The full guided solution, and what the question leads to at university, is in Section 13.4:
https://keithmcnulty.github.io/step_book_mcnulty/ch13/mean-or-median.html