詳細解答

STEP III 1989，第 15 題

概率與統計 • 中位數的分佈

> **STEP III 1989，第 15 題**
>
> 繼續閱讀之前，請先看題目（原題為英文）。題目載於 STEP Support Programme 的試題資料庫：
>
> step.maths.org/questions/89-s3-q15

關鍵思路

$2n+1$ 個數值的中位數（median）落在 z 附近，當且僅當其中一個數值落在 z 處的一個極小區間內，另有 n 個數值在它之下、n 個在它之上。第 (i) 及第 (ii) 部分給出了「全部在下」和「全部在上」這兩部分的概率，再數一數有多少種方法選定哪些數值擔當哪個角色，便得到密度。之後，利用密度的積分等於 1 這個事實，不費吹灰之力就能求出該積分；把同一技巧用於以 $n+1$ 代替 n 的情況，便得到方差。

解答

在整個解答中，X 在 $[-c,c]$ 上的密度為 $\frac{1}{2c}$，所以對於 $-c \le k \le c$，

$$\mathrm{P}(X > k) = \frac{c-k}{2c}, \qquad \mathrm{P}(X < k) = \frac{c+k}{2c}.$$

第 (i) 部分

這 n 個數值互相獨立（independent），所以把概率相乘：

$$\mathrm{P}(\text{全部 } n \text{ 個數值均大於} k) = \left(\frac{c-k}{2c}\right)^n.$$

第 (ii) 部分

同理，

$$\mathrm{P}(\text{全部 } n \text{ 個數值均小於} k) = \left(\frac{c+k}{2c}\right)^n.$$

中位數的密度

設 $X_1, \ldots, X_{2n+1}$ 為樣本，Z 為其中位數，即第 $(n+1)$ 小的數值。由於 X 是連續的，出現相同數值的概率為 0，所以我們不予理會。每個 X_i 都在 $[-c, c]$ 內，所以 Z 也在其中。

取 $(-c, c)$ 內的 z 和一個小的 $\delta z > 0$。略去 $(\delta z)^2$ 階的項後，事件 $z < Z \leq z + \delta z$ 發生，當且僅當

- 其中一個數值落在 $(z, z+\delta z]$ 內，
- 其餘數值中有 n 個小於 z，並且
- 餘下的 n 個數值大於 $z + \delta z$。

（有兩個或以上數值落在這個小區間內的情況，其概率為 $(\delta z)^2$ 階，在取極限時無關重要。）

想一想 把這 $2n+1$ 個數值分配到這三個角色，共有多少種方法？

落在小區間內的數值有 $2n+1$ 種選擇，然後從其餘數值中選出 n 個放在 z 之下，有 $\binom{2n}{n}$ 種方法。於是共有

$$(2n+1)\binom{2n}{n} = \frac{(2n+1)!}{(n!)^2}$$

種分配方法。這些分配互斥（mutually exclusive），而且每一種的概率相同。對於一種固定的分配，由獨立性以及第 (i) 和第 (ii) 部分（取 $k = z$），其概率為

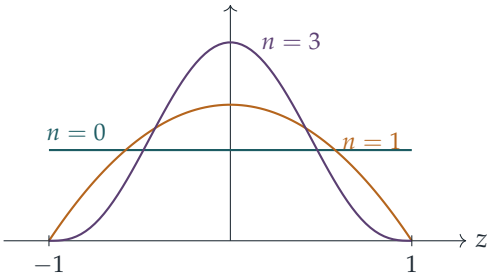
$$\frac{\delta z}{2c}\left(\frac{c+z}{2c}\right)^n \left(\frac{c-z-\delta z}{2c}\right)^n.$$

對所有分配求和，除以 δz，再令 $\delta z \to 0$，便得 Z 的密度為

$$\begin{aligned} g(z) &= \frac{(2n+1)!}{(n!)^2} \cdot \frac{1}{2c} \cdot \frac{(c+z)^n (c-z)^n}{(2c)^{2n}} \\ &= \frac{(2n+1)!}{(n!)^2 (2c)^{2n+1}} (c^2 - z^2)^n, \qquad -c \leq z \leq c, \end{aligned}$$

其中用到 $(c+z)(c-z) = c^2 - z^2$。這就是所要求的結果。

下圖顯示 $c = 1$ 時的密度。當 $n = 0$ 時，「中位數」只是單獨一個數值，服從均勻分佈（uniform distribution）。隨着 n 增大，中位數愈來愈集中在 0 附近。



積分

概率密度（probability density）在其取值範圍上的積分等於 1，所以

$$\frac{(2n+1)!}{(n!)^2(2c)^{2n+1}}\int_{-c}^{c}(c^2-z^2)^n\,\mathrm{d}z = 1,$$

因此

$$\int_{-c}^{c}(c^2-z^2)^n\,\mathrm{d}z = \frac{(n!)^2(2c)^{2n+1}}{(2n+1)!}.$$

這對每個整數 $n \geq 0$ 都成立，稍後我們就會用到這一點。

平均值與方差

密度 g 是 z 的偶函數（even function），所以 $zg(z)$ 是奇函數，它在對稱區間 $[-c,c]$ 上的積分為零：

$$\mathrm{E}(Z) = 0.$$

因此 $\mathrm{var}(Z) = \mathrm{E}(Z^2)$。記 $K_n = \dfrac{(2n+1)!}{(n!)^2(2c)^{2n+1}}$ 及 $I_n = \displaystyle\int_{-c}^{c}(c^2-z^2)^n\,\mathrm{d}z = \frac{1}{K_n}$。

? **想一想** 怎樣才能不展開而求出 $z^2(c^2-z^2)^n$ 的積分？

寫成 $z^2 = c^2 - (c^2-z^2)$。於是

$$\begin{aligned}\mathrm{E}(Z^2) &= K_n\int_{-c}^{c}\big(c^2-(c^2-z^2)\big)(c^2-z^2)^n\,\mathrm{d}z\\ &= K_n\big(c^2I_n - I_{n+1}\big) = c^2 - \frac{K_n}{K_{n+1}},\end{aligned}$$

其中用到 $K_nI_n = 1$ 和 $I_{n+1} = 1/K_{n+1}$，後者就是以 $n+1$ 代替 n 的積分結果。現在

$$\frac{K_n}{K_{n+1}} = \frac{(2n+1)!}{(2n+3)!}\cdot\frac{((n+1)!)^2}{(n!)^2}\cdot\frac{(2c)^{2n+3}}{(2c)^{2n+1}} = \frac{(n+1)^2\cdot 4c^2}{(2n+2)(2n+3)} = \frac{2(n+1)c^2}{2n+3}.$$

因此

$$\mathrm{var}(Z) = c^2\left(1-\frac{2n+2}{2n+3}\right) = \frac{c^2}{2n+3}.$$

驗算

- 當 $n = 0$ 時，中位數就是 X 的一個數值。密度變成 $\frac{1}{2c}$，積分為 $2c$，方差為 $\frac{c^2}{3}$，正是 $[-c, c]$ 上均勻分佈的方差。
- 當 $n = 1$ 及 $c = 1$ 時，積分公式給出 $\int_{-1}^{1}(1 - z^2)\,\mathrm{d}z = \frac{1 \cdot 8}{6} = \frac{4}{3}$，結果正確。
- 以 10^5 個大小為 5 的樣本（$n = 2$）並取 $c = 1.7$ 進行模擬，得到中位數的方差為 0.415，而 $c^2/7 \approx 0.413$。

常見錯誤

- 忘記了選擇中間數值所帶來的因子 $2n + 1$，只用了 $\binom{2n}{n}$。
- 漏掉中間數值本身的密度所帶來的因子 $\frac{1}{2c}$，結果得到的冪是 $(2c)^{2n}$ 而不是 $(2c)^{2n+1}$。
- 用二項式展開 $(c^2 - z^2)^n$ 來求 $\mathrm{E}(Z^2)$。這樣做是可行的，但所得的和很難化簡。

STEP 之外

中位數是樣本的順序統計量（order statistics）之一。同樣的計數論證可以給出任何連續分佈的 m 個數值中第 r 小者的密度，即 $\frac{m!}{(r-1)!\,(m-r)!}F^{r-1}(1 - F)^{m-r}f$。對於均勻分佈，這些都是貝塔分佈（beta distributions），而你所推導出的積分正是貝塔函數 $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ 的一個特例。方差 $c^2/(2n + 3)$ 以 $1/n$ 的速度縮小，這是邁向樣本中位數的漸近正態性（asymptotic normality）的第一步。Blitzstein 與 Hwang 合著的 *Introduction to Probability*（概率論入門）對順序統計量和貝塔分佈有很好的講解。

Beyond STEP 詳細解答：https://beyondstep.keithmcnulty.org/zh-hant/。解答文字採用 CC0 1.0；題目本身版權屬 Cambridge University Press & Assessment 所有（©），本文件不予轉載。由機器從英文翻譯而成；尚未經審閱。