詳細解答

STEP III 2001，第 14 題

概率與統計 • 方差的無偏估計量

> **STEP III 2001，第 14 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/01-s3-q14

關鍵思路

整道題的關鍵只有一點：「樣本的方差」所用的除數是 $n = 2$，因此會低估總體方差（population variance），而把它乘以二便可修正過來。只有兩個值時，樣本平均值（sample mean）就是兩者的中點，所以每個值與它相距都是兩值差距的一半。於是 V 只是單一個差 $X_1 - X_2$ 的函數，而最後要求的概率就變成正方形 $[0, a]^2$ 內的一塊面積。

解答

X 的方差

X 的密度函數（density）在 $[0, a]$ 上是 $f(x) = \frac{1}{a}$。由對稱性可知 $\mathrm{E}(X) = \frac{1}{2}a$，而

$$\mathrm{E}(X^2) = \int_0^a \frac{x^2}{a}\,\mathrm{d}x = \frac{a^2}{3}.$$

所以

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \frac{a^2}{3} - \frac{a^2}{4} = \frac{a^2}{12},$$

正如題目所求。

樣本方差

對於樣本 X_1, X_2，樣本平均值是 $\bar{X} = \frac{1}{2}(X_1 + X_2)$，而樣本的方差就是各值與它的偏差平方的平均值：

$$V = \tfrac{1}{2}[(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2].$$

由於 $X_1 - \bar{X} = \frac{1}{2}(X_1 - X_2)$ 及 $X_2 - \bar{X} = -\frac{1}{2}(X_1 - X_2)$，兩個偏差平方都等於 $\frac{1}{4}(X_1 - X_2)^2$。因此

$$V = \tfrac{1}{2} \cdot 2 \cdot \tfrac{1}{4}(X_1 - X_2)^2 = \tfrac{1}{4}(X_1 - X_2)^2,$$

這正是題目給出的形式。

想一想 怎樣不用二重積分（double integral）而求出 $\mathrm{E}((X_1 - X_2)^2)$？

由於 X_1 和 X_2 的分佈相同，$\mathrm{E}(X_1 - X_2) = 0$，所以 $\mathrm{E}((X_1 - X_2)^2) = \mathrm{Var}(X_1 - X_2)$。因為 X_1 和 X_2 獨立（independent），它們的方差可以相加（負號並無影響，因為 $\mathrm{Var}(-X_2) = \mathrm{Var}(X_2)$）：

$$\mathrm{Var}(X_1 - X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = \frac{a^2}{12} + \frac{a^2}{12} = \frac{a^2}{6}.$$

因此

$$\mathrm{E}(2V) = \tfrac{1}{2}\,\mathrm{E}((X_1 - X_2)^2) = \tfrac{1}{2} \cdot \frac{a^2}{6} = \frac{a^2}{12} = \mathrm{Var}(X).$$

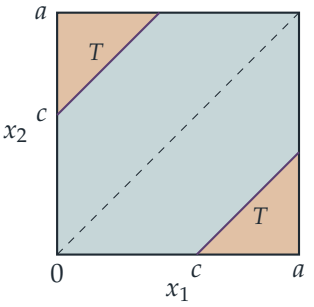
無論 a 取甚麼值，$2V$ 的期望值（expected value）都等於它所估計的量，所以 $2V$ 是 X 的方差的無偏估計量（unbiased estimator）。（相比之下，$\mathrm{E}(V) = \frac{a^2}{24}$，只有真正方差的一半。）

概率

由於 $V \geq 0$ 而 $\frac{1}{12}a^2 > 0$，我們可以取平方根：

$$V < \frac{a^2}{12} \iff (X_1 - X_2)^2 < \frac{a^2}{3} \iff |X_1 - X_2| < \frac{a}{\sqrt{3}}.$$

因為 X_1 和 X_2 獨立，而且各自服從均勻分佈（uniform distribution），所以數對 (X_1, X_2) 均勻分佈於正方形 $[0, a]^2$ 上。因此所求概率就是正方形面積中落在帶狀區域 $|x_1 - x_2| < c$ 內的比例，其中 $c = a/\sqrt{3} < a$。



餘下的部分由角落兩個着色的三角形 T 組成，每個都是等腰直角三角形，兩條直角邊的

長度均為 $a-c$。它們的總面積是 $2\cdot\frac{1}{2}(a-c)^2=(a-c)^2$，所以

$$\mathrm{P}\left(V<\frac{a^2}{12}\right)=1-\frac{(a-c)^2}{a^2}=1-\left(1-\frac{1}{\sqrt{3}}\right)^2=1-\left(\frac{4}{3}-\frac{2}{\sqrt{3}}\right)=\frac{2}{\sqrt{3}}-\frac{1}{3}.$$

也就是說，概率的準確值是

$$\mathrm{P}\left(V<\frac{a^2}{12}\right)=\frac{2\sqrt{3}-1}{3}.$$

取 $\sqrt{3}\approx 1.732$，這個值是 $\frac{2.464}{3}\approx 0.821$，所以準確至一位有效數字，概率是 0.8。

驗算

- 答案與 a 無關，這是理所當然的：把 X 按某個倍數縮放，V 和 $\frac{1}{12}a^2$ 都會按該倍數的平方縮放。
- 取 $a=3$ 模擬 4×10^5 個數對，得出 $2V$ 的平均值為 0.751（目標值 0.75），而 $V<\frac{1}{12}a^2$ 所佔的比例為 0.821，與 $\frac{1}{3}(2\sqrt{3}-1)=0.8214$ 吻合。
- 合理性檢查：概率落在 $(0,1)$ 之內；若把 $\frac{1}{12}a^2$ 換成 $\frac{1}{4}a^2$（V 可取的最大值），則 $c=a$，而公式 $1-(1-c/a)^2$ 給出 1。

常見錯誤

- 計算「樣本的方差」時用了除數 $n-1$：這樣會得到 $\frac{1}{2}(X_1-X_2)^2$，而不是題目所述的 $\frac{1}{4}(X_1-X_2)^2$。題目要的是除數 n。
- 忘記了絕對值，結果只計算了兩個三角形的其中一個（這樣得出的概率會小於 $\frac{1}{2}$）。
- 寫成 $\mathrm{Var}(X_1-X_2)=\mathrm{Var}(X_1)-\mathrm{Var}(X_2)=0$。

STEP 之外

這裏的結果是一條一般定理在 $n=2$ 時的情況：從方差為 σ^2 的分佈中抽取任何大小為 n 的樣本，$\sum(X_i-\bar{X})^2$ 的平均值都是 $(n-1)\sigma^2$。這就是為甚麼用除數 $n-1$（貝塞爾校正，Bessel's correction）會得到無偏估計量。大學一年級的統計學會進一步探討：在各個無偏估計量之中，哪一個的方差最小。Grimmett 與 Welsh 的 *Probability: An Introduction*，以及 Blitzstein 與 Hwang 的 *Introduction to Probability*，都以本文所用的方式講解和的期望值與方差。

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。