詳細解答

STEP III 2009，第 13 題

概率與統計 • 中心極限定理

> **STEP III 2009，第 13 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/09-s3-q13

關鍵思路

一切都是單一角度 θ 的函數：$X = \cos\theta$ 及 $Y = \sin\theta$。因此，X 的分佈可由累積分佈函數（cumulative distribution function）求得（找出圓上哪一段弧使 $\cos\theta \le x$），而平均值、方差（variance）和協方差（covariance）不過是三角函數在一整圈上的平均。在第 (ii) 部分，$\overline{X}$ 是許多獨立同分佈（independent, identically distributed）變量的平均，這正是中心極限定理（central limit theorem）適用的情況。

解答

第 (i) 部分

由於 P 在單位圓上，$X = \cos\theta$ 及 $Y = \sin\theta$，其中 θ 在 $[0, 2\pi)$ 上的密度為 $\frac{1}{2\pi}$。

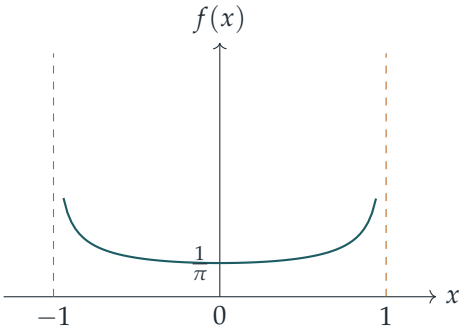
X 的密度。 顯然 $-1 \le X \le 1$。取 $-1 < x < 1$，並設 $\alpha = \arccos x \in (0, \pi)$。在 $[0, 2\pi)$ 上，$\cos\theta > x$ 當且僅當 $\theta < \alpha$ 或 $\theta > 2\pi - \alpha$，這是一段總長度為 2α 的弧。因此

$$F(x) = \mathrm{P}(X \le x) = 1 - \frac{2\alpha}{2\pi} = 1 - \frac{\arccos x}{\pi}.$$

利用 $\dfrac{\mathrm{d}}{\mathrm{d}x}\arccos x = -\dfrac{1}{\sqrt{1-x^2}}$ 求導，得

$$f(x) = \frac{1}{\pi\sqrt{1-x^2}} \qquad (-1 < x < 1),$$

而在其他情況下 $f(x) = 0$。（簡單驗算一下：$\int_{-1}^{1} f(x)\,\mathrm{d}x = \frac{1}{\pi}[\arcsin x]_{-1}^{1} = 1$。）



密度在 $x = 0$ 處最小，在 $x = \pm 1$ 處趨向無窮大：點 P 在水平直徑兩端附近的水平移動最慢，所以 X 大部分時間都停留在那裏。

方差。 由對稱性（f 是偶函數），$\mathrm{E}(X) = 0$。直接用 θ 計算，

$$\mathrm{E}(X^2) = \frac{1}{2\pi}\int_0^{2\pi} \cos^2\theta \,\mathrm{d}\theta = \frac{1}{2\pi}\int_0^{2\pi} \tfrac{1}{2}(1+\cos 2\theta)\,\mathrm{d}\theta = \frac{1}{2},$$

所以

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \tfrac{1}{2}.$$

同理（或因為 $\theta - \frac{\pi}{2}$ 同樣在一整圈上均勻分佈），$\mathrm{E}(Y) = 0$ 及 $\mathrm{Var}(Y) = \frac{1}{2}$。

不相關（uncorrelated）。 我們有

$$\mathrm{E}(XY) = \frac{1}{2\pi}\int_0^{2\pi} \cos\theta \sin\theta \,\mathrm{d}\theta = \frac{1}{4\pi}\int_0^{2\pi} \sin 2\theta \,\mathrm{d}\theta = 0,$$

所以 $\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) = 0 - 0 = 0$。因此 X 與 Y 不相關。

並非獨立。 它們肯定不是獨立的，因為 $X^2 + Y^2 = 1$：一旦知道 X，Y 就只能是 $\pm\sqrt{1 - X^2}$ 其中之一。舉一個具體事件為例：$\mathrm{P}(X > 0.8) > 0$ 及 $\mathrm{P}(Y > 0.8) > 0$，但 $\mathrm{P}(X > 0.8 \text{ 及 } Y > 0.8) = 0$，因為那樣的話 $X^2 + Y^2 > 1.28$。若兩者獨立，這個聯合概率便須等於那個（正的）乘積。相關係數為零只表示沒有線性關係；這裏的關係是一個圓，它足夠對稱，以致完全沒有線性趨勢。

第 (ii) 部分

每一對 (X_i, Y_i) 都服從第 (i) 部分的分佈，而由於各點是獨立選取的，下標不同的各對互相獨立。

$\overline{X}$ 與 $\overline{Y}$ 不相關。 協方差是雙線性（bilinear）的，所以

$$\operatorname{Cov}(\overline{X},\overline{Y}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{Cov}(X_i,Y_j).$$

當 $i \neq j$ 時，X_i 只取決於 P_i，Y_j 只取決於 P_j，而這兩點互相獨立，所以 $\operatorname{Cov}(X_i,Y_j)=0$。當 $i=j$ 時，由第 (i) 部分得 $\operatorname{Cov}(X_i,Y_i)=0$。每一項都等於零，所以 $\operatorname{Cov}(\overline{X},\overline{Y})=0$：$\overline{X}$ 與 $\overline{Y}$ 不相關。

近似概率。 各 X_i 獨立同分佈，平均值為 0，方差為 $\frac{1}{2}$。所以

$$\mathrm{E}(\overline{X})=0, \qquad \operatorname{Var}(\overline{X}) = \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{Var}(X_i) = \frac{1}{n^2}\cdot\frac{n}{2} = \frac{1}{2n},$$

這裏可以把方差相加，是因為各 X_i 互相獨立。

> **想一想** X 一點也不像正態分佈（normal distribution）（看看它的密度）。為甚麼我們可以用正態近似？

中心極限定理只要求各項獨立同分佈且方差有限，不論其分佈形狀如何。所以當 n 很大時，$\overline{X}$ 近似服從 $\mathrm{N}(0,\frac{1}{2n})$，標準差（standard deviation）為 $\sigma = \dfrac{1}{\sqrt{2n}}$。現在

$$\sqrt{\frac{2}{n}} = \frac{2}{\sqrt{2n}} = 2\sigma,$$

所以設 $Z \sim \mathrm{N}(0,1)$，則

$$\begin{aligned}\mathrm{P}\left(|\overline{X}| \le \sqrt{\tfrac{2}{n}}\right) = \mathrm{P}\left(\frac{|\overline{X}|}{\sigma} \le 2\right) &\approx \mathrm{P}(|Z|\le 2)\\ &= 2\Phi(2)-1 = 2(0.9772)-1 \approx 0.954 \approx 0.95,\end{aligned}$$

正如題目所求。

驗算

- 方差 $\frac{1}{2}$ 是合理的：$X^2+Y^2=1$，而由對稱性 $\mathrm{E}(X^2)=\mathrm{E}(Y^2)$，所以兩者都必定是 $\frac{1}{2}$。根本不用積分！
- 用 4×10^5 個隨機角度進行模擬，結果在多個 x 值上與 $F(x)=1-\frac{1}{\pi}\arccos x$ 吻合至約小數點後三位，並得出 $\operatorname{Var}(X)\approx 0.5007$ 及 $\operatorname{Cov}(X,Y)\approx 0.001$。
- 取 $n=50$ 並重複 40 000 次，模擬所得的 $\operatorname{Var}(\overline{X})$ 為 0.00997（理論值 0.01），而 $|\overline{X}| \le \sqrt{2/50}=0.2$ 所佔的比例為 0.954，對比 $\mathrm{P}(|Z|\le 2)=0.9545$。

常見錯誤

- 對 $1 - \frac{1}{\pi}\arccos x$ 求導時弄錯正負號，得出一個負的「密度」。密度永遠不可能是負數，所以要檢查正負號。
- 由協方差為零推斷兩者獨立。獨立可推出協方差為零，但反過來並不成立。
- 使用 $\mathrm{Var}(\overline{X}) = \frac{1}{2}$ 或 $\frac{1}{2} \cdot \frac{1}{\sqrt{n}}$：$n$ 個獨立副本的平均值，其方差是單個方差除以 n；按 $1/\sqrt{n}$ 的比例變化的是標準差。

STEP 之外

密度 $1/(\pi\sqrt{1-x^2})$ 就是反正弦分佈（arcsine distribution），它出現的頻繁程度出人意表，例如它是隨機遊走（random walk）處於正值的時間所佔比例的極限分佈。這道題的其餘部分，談的是不相關與獨立之間的差距，以及中心極限定理；在大學一年級的概率課程中，你會用矩母函數（moment generating function）或特徵函數（characteristic function）來證明這個定理。事實上，$(\overline{X}, \overline{Y})$ 近似服從二元正態分佈（bivariate normal distribution），而對於聯合正態的變量，相關係數為零確實可推出獨立。Grimmett 與 Welsh 的 *Probability: An Introduction*（概率入門教材）對以上各點都有很好的講解。

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。