解答例

STEP III 2009、第 13 問

確率・統計 ・ 中心極限定理

STEP III 2009、問 13

先に進む前に、まず問題を読んでおこう。問題は STEP Support Programme の問題データベースにある：

step.maths.org/questions/09-s3-q13

鍵となる考え方

すべては一つの角 θ の関数である：$X = \cos\theta$、$Y = \sin\theta$。したがって X の分布は累積分布関数（cumulative distribution function）から求まり（$\cos\theta \le x$ となるのが円のどの弧かを調べればよい）、平均・分散・共分散は単に三角関数を一周にわたって平均したものにすぎない。小問 (ii) では、$\overline{X}$ は多数の独立同分布（independent, identically distributed）な確率変数の平均であり、これはまさに中心極限定理（central limit theorem）が適用できる状況である。

解答

小問 (i)

P は単位円上にあるので、$X = \cos\theta$、$Y = \sin\theta$ である。ここで θ は $[0, 2\pi)$ 上で密度 $\frac{1}{2\pi}$ をもつ。

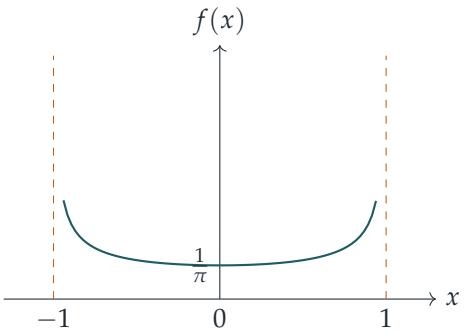
X の確率密度関数。 明らかに $-1 \le X \le 1$ である。$-1 < x < 1$ をとり、$\alpha = \arccos x \in (0, \pi)$ とおく。$[0, 2\pi)$ 上で $\cos\theta > x$ となるのはちょうど $\theta < \alpha$ または $\theta > 2\pi - \alpha$ のときであり、これは全長 2α の弧である。よって

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

$\frac{\mathrm{d}}{\mathrm{d}x}\arccos x = -\frac{1}{\sqrt{1-x^2}}$ を用いて微分すると、

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

であり、それ以外では $f(x) = 0$ である。(簡単な確認：$\int_{-1}^{1} f(x)\,\mathrm{d}x = \frac{1}{\pi}[\arcsin x]_{-1}^{1} = 1$。)



密度は $x = 0$ で最小となり、$x = \pm 1$ で発散する。点 P は水平な直径の両端付近で水平方向に最もゆっくり動くので、X はそこで大部分の時間を過ごすのである。

分散。 対称性（f は偶関数）より $\mathrm{E}(X) = 0$ である。θ を直接用いて計算すると、

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

よって

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

同じ議論により（あるいは $\theta - \frac{\pi}{2}$ もまた一周にわたって一様に分布することから）、$\mathrm{E}(Y) = 0$、$\mathrm{Var}(Y) = \frac{1}{2}$ である。

無相関であること。

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

であるから、$\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) = 0 - 0 = 0$ である。よって X と Y は無相関（uncorrelated）である。

独立でないこと。 $X^2 + Y^2 = 1$ であるから、両者は決して独立（independent）ではない。X がわかれば、Y は $\pm\sqrt{1-X^2}$ のいずれかに確定してしまう。具体的な事象で示すと、$\mathrm{P}(X > 0.8) > 0$ かつ $\mathrm{P}(Y > 0.8) > 0$ であるが、$\mathrm{P}(X > 0.8 \text{ かつ} Y > 0.8) = 0$ である。このとき $X^2 + Y^2 > 1.28$ となってしまうからである。独立であれば、この同時確率は（正の）積に等しくなければならない。相関が 0 であるということは、**線形な**関係がないことを述べているにすぎない。ここでの関係は円であり、線形な傾向がまったく現れないほど対称なのである。

小問 (ii)

各組 (X_i, Y_i) は小問 (i) の分布に従い、点は独立に選ばれるので、添字の異なる組どうしは独立である。

$\overline{X}$ と $\overline{Y}$ は無相関である。 共分散は双線形（bilinear）であるから、

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

$i \neq j$ のとき、X_i は P_i のみに、Y_j は P_j のみに依存し、これらは独立であるから $\mathrm{Cov}(X_i, Y_j) = 0$ である。$i = j$ のときは、小問 (i) より $\mathrm{Cov}(X_i, Y_i) = 0$ である。すべての項が 0 になるので $\mathrm{Cov}(\overline{X}, \overline{Y}) = 0$、すなわち $\overline{X}$ と $\overline{Y}$ は無相関である。

近似確率。 X_i は平均 0、分散 $\frac{1}{2}$ の独立同分布な確率変数である。したがって

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

ここで分散を足し合わせてよいのは、X_i が独立だからである。

> **考えてみよう** X は正規分布とはまったく似ていない（その密度を見よ）。なぜ正規近似を用いてよいのだろうか。

中心極限定理が必要とするのは、分散が有限な独立同分布の項であることだけであり、分布の形は問わない。したがって n が大きいとき、$\overline{X}$ は近似的に $\mathrm{N}(0, \frac{1}{2n})$ に従い、その標準偏差は $\sigma = \dfrac{1}{\sqrt{2n}}$ である。ここで

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

であるから、$Z \sim \mathrm{N}(0,1)$ として

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

となり、題意が示された。

検算

- 分散が $\frac{1}{2}$ であることは納得がいく。$X^2 + Y^2 = 1$ であり、対称性から $\mathrm{E}(X^2) = \mathrm{E}(Y^2)$ なので、それぞれ $\frac{1}{2}$ でなければならない。積分は不要である！

- 4×10^5 個のランダムな角によるシミュレーションでは、いくつかの x の値において $F(x) = 1 - \frac{1}{\pi}\arccos x$ と小数第 3 位程度まで一致し、$\mathrm{Var}(X) \approx 0.5007$、$\mathrm{Cov}(X, Y) \approx 0.001$ が得られた。
- $n = 50$、反復 40 000 回のシミュレーションでは、$\mathrm{Var}(\overline{X})$ は 0.00997（理論値 0.01）であり、$|\overline{X}| \leq \sqrt{2/50} = 0.2$ となった割合は 0.954 であった。これに対し $\mathrm{P}(|Z| \leq 2) = 0.9545$ である。

よくある間違い

- $1 - \frac{1}{\pi}\arccos x$ を微分する際に符号を誤り、負の「密度」を得てしまう。密度が負になることはありえないので、符号を確認すること。
- 共分散が 0 であることから独立性を結論してしまう。独立ならば共分散は 0 であるが、逆は成り立たない。
- $\mathrm{Var}(\overline{X}) = \frac{1}{2}$ や $\frac{1}{2} \cdot \frac{1}{\sqrt{n}}$ としてしまう。n 個の独立なコピーの平均の分散は、一つの分散を n で割ったものである。$1/\sqrt{n}$ 倍になるのは**標準偏差**のほうである。

STEP のその先へ

密度 $1/(\pi\sqrt{1-x^2})$ は**逆正弦分布**（arcsine distribution）であり、驚くほど頻繁に現れる。たとえば、ランダムウォーク（random walk）が正の側にいる時間の割合の極限分布として現れる。この問題の残りの部分は、無相関と独立の間の隔たり、そして中心極限定理に関するものである。中心極限定理は、大学 1 年次の確率論の講義で、積率母関数（moment generating function）や特性関数（characteristic function）を用いて証明することになる。実は $(\overline{X}, \overline{Y})$ は近似的に 2 変量正規分布（bivariate normal distribution）に従い、同時正規分布に従う確率変数については、相関が 0 ならば独立である。Grimmett と Welsh による *Probability: An Introduction*（確率論入門）は、これらすべてをよく扱っている。

Beyond STEP 解答集：https://beyondstep.keithmcnulty.org/ja/。解答本文は CC0 1.0 である。問題そのものは © Cambridge University Press & Assessment に帰属し、ここには掲載していない。英語版から機械翻訳したものであり、まだ校閲を経ていない。