解答例

STEP III 2001、第 14 問

確率・統計 ・ 不偏分散

> **STEP III 2001、問 14**
>
> 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/01-s3-q14

鍵となる考え方

ここでの議論はすべて一つの事実にかかっている。「標本の分散」は除数 $n = 2$ を用いるので母分散（population variance）を過小評価し、それを 2 倍することでその偏りが正される、ということである。2 つの値に対しては標本平均（sample mean）は中点であるから、各値は標本平均から差のちょうど半分だけ離れている。これにより V はただ一つの差 $X_1 - X_2$ の関数となり、最後の確率は正方形 $[0, a]^2$ の中の面積となる。

解答

X の分散

X の確率密度関数（density）は $[0, a]$ 上で $f(x) = \frac{1}{a}$ である。対称性より $\mathrm{E}(X) = \frac{1}{2}a$ であり、

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

したがって

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

となり、示すべき式が得られた。

標本分散

標本（sample）X_1, X_2 に対して標本平均は $\bar{X} = \frac{1}{2}(X_1 + X_2)$ であり、標本の分散（variance）はそこからの偏差の 2 乗の平均である：

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

ここで $X_1 - \bar{X} = \frac{1}{2}(X_1 - X_2)$、$X_2 - \bar{X} = -\frac{1}{2}(X_1 - X_2)$ であるから、偏差の 2 乗はどちらも $\frac{1}{4}(X_1 - X_2)^2$ に等しい。よって

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

となり、これは与えられた形である。

考えてみよう 2 重積分を使わずに $\mathrm{E}((X_1 - X_2)^2)$ を求めるにはどうすればよいか。

X_1 と X_2 は同じ分布に従うので $\mathrm{E}(X_1 - X_2) = 0$ であり、したがって $\mathrm{E}((X_1 - X_2)^2) = \mathrm{Var}(X_1 - X_2)$ である。X_1 と X_2 は独立（independent）であるから、それらの分散は足し合わせられる（$\mathrm{Var}(-X_2) = \mathrm{Var}(X_2)$ なので、マイナス符号は影響しない）：

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

したがって

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

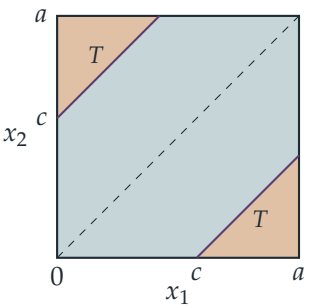
$2V$ の期待値（expected value）は、a の値によらず、それが推定する量に等しい。よって $2V$ は X の分散の不偏推定量（unbiased estimator）である。（これに対して $\mathrm{E}(V) = \frac{a^2}{24}$ であり、真の分散の半分にすぎない。）

確率

$V \geq 0$ かつ $\frac{1}{12}a^2 > 0$ であるから、平方根をとってよい：

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

X_1 と X_2 は独立で、それぞれが一様分布（uniform distribution）に従うので、組 (X_1, X_2) は正方形 $[0, a]^2$ 上で一様に分布する。したがって求める確率は、正方形の面積のうち帯状領域 $|x_1 - x_2| < c$ に含まれる部分の割合である。ただし $c = a/\sqrt{3} < a$ とする。



余事象（complement）は、影をつけた隅の 2 つの三角形 T からなる。それぞれは等しい 2 辺の長さが $a-c$ の直角二等辺三角形である。その面積の合計は $2\cdot\frac{1}{2}(a-c)^2=(a-c)^2$ であるから、

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

すなわち、正確な確率は

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

である。$\sqrt{3}\approx 1.732$ とすると、これは $\frac{2.464}{3}\approx 0.821$ であり、有効数字 1 桁で確率は 0.8 である。

検算

- 答えは a に依存しないが、これは当然である。X を定数倍すると、V と $\frac{1}{12}a^2$ はどちらもその定数の 2 乗倍になるからである。
- $a=3$ として 4×10^5 組をシミュレーションしたところ、$2V$ の平均は 0.751（目標値 0.75）、$V<\frac{1}{12}a^2$ となる割合は 0.821 となり、$\frac{1}{3}(2\sqrt{3}-1)=0.8214$ と一致した。
- 妥当性の確認：確率は $(0,1)$ に入っている。また、$\frac{1}{12}a^2$ を $\frac{1}{4}a^2$（V のとりうる最大値）に置き換えると $c=a$ となり、公式 $1-(1-c/a)^2$ は 1 を与える。

よくある間違い

- 「標本の分散」に除数 $n-1$ を用いること。そうすると $\frac{1}{2}(X_1-X_2)^2$ となり、問題に書かれた $\frac{1}{4}(X_1-X_2)^2$ にはならない。この問題では除数 n が求められている。
- 絶対値を忘れ、2 つの三角形のうち一方しか数えないこと（その場合、確率は $\frac{1}{2}$ 未満になってしまう）。
- $\mathrm{Var}(X_1-X_2)=\mathrm{Var}(X_1)-\mathrm{Var}(X_2)=0$ と書くこと。

STEP のその先へ

ここでの結果は、ある一般的な定理の $n=2$ の場合である。分散 σ^2 の分布からとった大きさ n の任意の標本に対して、$\sum(X_i-\bar{X})^2$ の平均は $(n-1)\sigma^2$ である。除数 $n-1$（ベッセルの補正（Bessel's correction））が不偏推定量を与えるのはこのためである。大学 1 年の統計学では、さらに、どの不偏推定量が最小の分散をもつかという問題へと進む。Grimmett と Welsh の *Probability: An Introduction*（確率論入門）、および Blitzstein

と Hwang の *Introduction to Probability*（確率論入門）は、どちらもここで用いたスタイルで和の期待値と分散を扱っている。

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。