

問 1

ある党の議員一人ひとりの選挙の当選確率を p とし、議会の議席数を n としたとき、その党の獲得議席数の上限と下限を求めることにする。

(a) 当選した場合を 1, 落選した場合を 0 とし、当落の確率変数を X としたとき、 X の平均と分散を求めなさい【10】

(b) 新しい確率変数 Y_n を以下のように定義する。

$$Y_n = X_1 + X_2 + \cdots + X_n$$

この変数は獲得議席数を表しており、独立な確率変数 $X_1, X_2, \dots, X_n$ は X と同じ分布に従う。このとき、 Y_n の平均と分散を求めなさい。【10】

(c) 中心極限定理と裏面の標準正規分布表を用いて以下の等式を満たす U を求めなさい。

$$\Pr(|Y_n - m| \leq U) = 0.9974$$

ただし、 $m = E[Y_n]$ である。【20】

(d) $n = 480, p = 0.3$ として、(c) で求めた U を用いて上限と下限を求めなさい。【10】

問 2

区間 $[-1/2, 1/2]$ 上の一様分布に従う確率変数を X とする。以下の問に答えなさい。

(a) 確率変数 X の平均と分散を求めなさい【10】

(b) 新しい確率変数 Y を以下のように定義する。

$$Y = (X_1 + X_2) / \sqrt{2}$$

独立な確率変数 X_1, X_2 は X と同じ分布に従う。確率変数 Y の平均と分散を求めよ【10】

(c) 確率変数 Y の確率密度関数を求めなさい。【20】

(d) 確率変数 X と Y のどちらが正規分布に近いのかを定量的に評価したい。その評価を行うために 2 つの分布の間の距離のようなものを以下の式で定義する。

$$D(p \parallel q) = \int p(x) \ln \left(\frac{p(x)}{q(x)} \right) dx$$

この式は確率密度関数 $p(x)$ と $q(x)$ の類似性を表すカルバック情報量と呼ばれるものである。このとき、以下のように $q(x)$ を定義する。

$$q(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

ただし、 $m = E[X], \sigma^2 = V[X]$ である。 $p(x)$ が一様分布のときと (b) で求めた分布のときの 2 通りの場合に関してカルバック情報量を求めなさい。また、必要であれば以下の式を用いて良い【20】

$$\lim_{x \rightarrow 0} x^2 \ln(x) = 0, \ln(\pi/3) = 0.0461, \ln(\pi/6) = -0.647$$

問 1

ある党の議員一人ひとりの選挙の当選確率を p とし、議会の議席数を n としたとき、その党の獲得議席数の上限と下限を求めることにする。

(a) 当選した場合を 1, 落選した場合を 0 とし、当落の確率変数を X としたとき、 X の平均と分散を求めなさい【10】

(b) 新しい確率変数 Y_n を以下のように定義する。

$$Y_n = X_1 + X_2 + \cdots + X_n$$

この変数は獲得議席数を表しており、独立な確率変数 $X_1, X_2, \dots, X_n$ は X と同じ分布に従う。このとき、 Y_n の平均と分散を求めなさい。【10】

(c) 中心極限定理と裏面の標準正規分布表を用いて以下の等式を満たす U を求めなさい。

$$\Pr(|Y_n - m| \leq U) = 0.9974$$

ただし、 $m = E[Y_n]$ である。【20】

(d) $n = 480, p = 0.3$ として、(c) で求めた U を用いて上限と下限を求めなさい。【10】

解答

(a) $E[X] = 1 \cdot p + 0 \cdot (1 - p) = p$

$$\begin{aligned} V[X] &= (1 - p)^2 \cdot p + (0 - p)^2 \cdot (1 - p) \\ &= p - p^2 \end{aligned}$$

(b) $E[Y_n] = E[X_1] + \cdots + E[X_n]$
 $= np$

$$\begin{aligned} V[Y_n] &= V[X_1] + \cdots + V[X_n] \\ &= n(p - p^2) \end{aligned}$$

(c) 中心極限定理より確率変数 Y_n は (b) の平均と分散をもつ正規分布に従う。このため、

$$\begin{aligned} \Pr(|Y_n - np| \leq t\sqrt{n(p - p^2)}) &= 0.9974 \text{ とする。以下の標準正規分布表において、} \\ \Pr(|X| \leq 3) &= \Pr(X \leq 3) - \Pr(X < -3) \\ &= \Pr(X \leq 3) - \Pr(X > 3) \\ &= \Pr(X \leq 3) - (1 - \Pr(X \leq 3)) \\ &= 2\Pr(X \leq 3) - 1 = 0.9974 \end{aligned}$$

となるため、 $t = 3$ のときに等式が成立する。よって、 $U = 3\sqrt{n(p - p^2)}$ である。

(d) (c) より、上限と下限は
 $np - 3\sqrt{n(p - p^2)} \leq Y_n \leq np + 3\sqrt{n(p - p^2)}$
 となるため、下限: 114 名、上限: 174 名

以上

・標準正規分布の累積分布と値は次式と表を参照せよ。

$$P(x < z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

$P(x < z)$	0.500	0.6915	0.8413	0.9332	0.9772	0.9938	0.9987	0.9998	0.9999	1
z	0	0.50	1.00	1.50	2.00	2.50	3.00	3.50	4.00	> 4

問 2

区間 $[-1/2, 1/2]$ 上の一様分布に従う確率変数を X とする。以下の問に答えなさい。

(a) 確率変数 X の平均と分散を求めなさい 【10】

(b) 新しい確率変数 Y を以下のように定義する。

$$Y = (X_1 + X_2) / \sqrt{2}$$

独立な確率変数 X_1, X_2 は X と同じ分布に従う。確率変数 Y の平均と分散を求めよ。【10】

(c) 確率変数 Y の確率密度関数を求めなさい。【20】

(d) 確率変数 X と Y のどちらが正規分布に近いのかを定量的に評価したい。その評価を行うために 2 つの分布の間の距離を以下の式で定義する。

$$D(p \parallel q) = \int p(x) \ln \left(\frac{p(x)}{q(x)} \right) dx$$

この式は確率密度関数 $p(x)$ と $q(x)$ の類似性を表すカルバック情報量と呼ばれるものである。このとき、以下のように $q(x)$ を定義する。

$$q(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

ただし、 $m = E[X], \sigma^2 = V[X]$ である。 $p(x)$ が一様分布のときと (b) で求めた分布のときの 2 通りの場合に関してカルバック情報量を求めなさい。また、必要であれば以下の式を用いて良い 【20】

$$\lim_{x \rightarrow 0} x^2 \ln(x) = 0, \ln(\pi/3) = 0.0461, \ln(\pi/6) = -0.647$$

解答

(a) $E[X] = 0, V[X] = 1/12$

(b) $E[Y] = (E[X_1] + E[X_2]) / 2 = 0$

$$V[Y] = (V[X_1] + V[X_2]) / 2 = 1/12$$

(c) 先に $Z = X_1 + X_2$ として Z の分布を求める。

$$\begin{aligned} p_Z(z) &= \iint \delta(z - x_1 - x_2) p(x_1) p(x_2) dx_1 dx_2 \\ &= \int p(z - x_2) p(x_2) dx_2 \\ &= \int_{-1/2}^{1/2} p(z - x_2) dx_2 \end{aligned}$$

このとき、 $t = z - x_2$ と置き換える。積分区間が $[-1/2, 1/2]$ である点に注意すると

$$p_Z(z) = \int_{z-1/2}^{z+1/2} p(t) dt$$

$p(t)$ もまた、 X と同じ分布に従うため、場合分けが必要である。

(i) $-1 \leq z < 0$ のとき

$$p_Z(z) = \int_{-1/2}^{z+1/2} p(t) dt = 1 + z$$

(ii) $0 \leq z < 1$ のとき

$$p_Z(z) = \int_{z-1/2}^{1/2} p(t) dt = 1 - z$$

(iii) $|z| \geq 1$ のとき

$$p_Z(z) = 0$$

最後に $Y = Z / \sqrt{2}$ と変換する

$$\begin{aligned} p_Y(y) &= \int \delta(y - z / \sqrt{2}) p_Z(z) dz \\ &= p_Z(\sqrt{2}y) \sqrt{2} \\ &= \sqrt{2}(1 - \sqrt{2}|y|), (|y| < 1/\sqrt{2}) \end{aligned}$$

(d) $p(x)$ と $q(x)$ は共に偶関数であるため、カルバック情報量の被積分関数も偶関数である。また、正規分布の対数は以下のように表現し直すことができる。

$$\ln(q(x)) = -x^2 / 2\sigma^2 - \ln(2\pi\sigma^2) / 2$$

また、(a) より $m = 0, \sigma^2 = 1/12$ である。

・ 一様分布の場合

$$\begin{aligned} D(p \parallel q) &= 2 \left(\int_0^{1/2} p(x) \ln p(x) dx - \int_0^{1/2} p(x) \ln q(x) dx \right) \\ &= -2 \int_0^{1/2} p(x) \ln q(x) dx \\ &= -2 \int_0^{1/2} [-x^2 / 2\sigma^2 - \ln(2\pi\sigma^2) / 2] dx \\ &= (1 + \ln(\pi/6)) / 2 = 0.1765 \end{aligned}$$

・ (b) の分布の場合

$$D(p \parallel q) = 2 \left(\int_0^{1/\sqrt{2}} p(x) \ln p(x) dx - \int_0^{1/\sqrt{2}} p(x) \ln q(x) dx \right)$$

<右辺第 1 項>

$$\begin{aligned} \int_0^{1/\sqrt{2}} p(x) \ln p(x) dx &= \frac{1}{2} \int_0^{\sqrt{2}} t \ln t dt \\ &= \frac{1}{2} \{ [t^2 \ln(t) / 2]_{t=0}^{\sqrt{2}} - [t^2 / 4]_{t=0}^{\sqrt{2}} \} \\ &= (\ln(2) - 1) / 4 \end{aligned}$$

途中で $t = p(x)$ と置換。

<右辺第 2 項>

$$\begin{aligned} \int_0^{1/\sqrt{2}} p(x) \ln q(x) dx &= -\sqrt{2} \int_0^{1/\sqrt{2}} (1 - \sqrt{2}x) [x^2 / 2\sigma^2 + \ln(2\pi\sigma^2) / 2] dx \\ &= -\sqrt{2} \int_0^{1/\sqrt{2}} \left[-\sqrt{2} \frac{x^3}{2\sigma^2} + \frac{x^2}{2\sigma^2} - \frac{\sqrt{2} \ln(2\pi\sigma^2)}{2} x + \frac{\ln(2\pi\sigma^2)}{2} \right] dx \\ &= -\sqrt{2} \int_0^{1/\sqrt{2}} \left[-6\sqrt{2}x^3 + 6x^2 - \frac{\sqrt{2} \ln(\pi/6)}{2} x + \frac{\ln(\pi/6)}{2} \right] dx \\ &= -\sqrt{2} \left(-\frac{3\sqrt{2}}{8} + \frac{1}{\sqrt{2}} - \frac{\sqrt{2} \ln(\pi/6)}{8} + \frac{\ln(\pi/6)}{2\sqrt{2}} \right) \\ &= -\frac{1 + \ln(\pi/6)}{4} \end{aligned}$$

以上より、

$$\begin{aligned} D(p \parallel q) &= \frac{\ln(2) + \ln(\pi/6)}{2} \\ &= \ln(\pi/3) / 2 = 0.0231 \end{aligned}$$

- 標準正規分布の累積分布と値は次式と表を参照せよ。

$$P(x < z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx$$

$P(x < z)$	0.500	0.6915	0.8413	0.9332	0.9772	0.9938	0.9987	0.9998	0.9999	1
z	0	0.50	1.00	1.50	2.00	2.50	3.00	3.50	4.00	> 4