

9 9回目講義の補足プリント

9.1 基本定理

定理 1. $X_1, \dots, X_n$ は互いに独立に同一の確率分布に従うものとする。

$$E(X_1) = \dots = E(X_n) \equiv \mu, \quad V(X_1) = \dots = V(X_n) \equiv \sigma^2$$

と置く。このとき、 $X_1, \dots, X_n$ の平均 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ は次の 2 式を満足する。

$$E(\bar{X}) = \mu, \quad V(\bar{X}) = \frac{\sigma^2}{n}. \quad (1)$$

例 9.1. (典型例) n 個の確率変数 $X_1, \dots, X_n$ が互いに独立に同一の分布 F に従っているとする。分布 F の平均は μ 、分散は σ^2 であるとする。 $T = \sum_{i=1}^n X_i$ とおく。このとき、次の結果が得られる。

1. ベルヌーイ分布 $Ber(p)$ の場合。この場合、 $\mu = p$, $\sigma^2 = p(1-p)$ であるから、

$$E(\bar{X}) = p, V(\bar{X}) = \frac{p(1-p)}{n} \quad (2)$$

$$E(T) = np, V(T) = np(1-p) \quad (3)$$

2. ポアソン分布 $P_O(\lambda)$ の場合。この場合、 $\mu = \lambda$, $\sigma^2 = \lambda$ であるから、

$$E(\bar{X}) = \lambda, V(\bar{X}) = \frac{\lambda}{n} \quad (4)$$

$$E(T) = n\lambda, V(T) = n\lambda \quad (5)$$

3. 正規分布 $N(\mu, \sigma^2)$ の場合。この場合、

$$E(\bar{X}) = \mu, V(\bar{X}) = \frac{\sigma^2}{n} \quad (6)$$

$$E(T) = n\mu, V(T) = n\sigma^2 \quad (7)$$

4. 指数分布 $E_X(\lambda)$ の場合。この場合、 $\mu = 1/\lambda$, $\sigma^2 = 1/\lambda^2$ であるから、

$$E(\bar{X}) = 1/\lambda, V(\bar{X}) = \frac{1}{n\lambda^2} \quad (8)$$

その他の確率分布についても同様である。

上式とチェビシェフの不等式を組み合わせ、 $\bar{X}$ のばらつきを調べよう。チェビシェフの不等式を復習すると次の通り：確率変数 Z は、平均 $E(Z) = \alpha$ 、分散 $V(Z) = \beta^2$ であるとする。このとき、

$$P(\alpha - k\beta \leq Z \leq \alpha + k\beta) \geq 1 - 1/k^2.$$

- (i) あるテレビ番組の視聴率が 20%であったとする（通常この数値は未知だが、今は全知の立場に立つことにする）。400 世帯に視聴の有無について尋ねるとする。回答を X_i ($i = 1, 2, \dots, 400$) と表し、視聴していたならば $X_i = 1$ 、そうでなければ $X_i = 0$ と記録するとする。このとき、 $X_1, \dots, X_{400}$ は互いに独立に同一のベルヌーイ分布 $Ber(0.2)$ に従っていると見なしてよい。このとき、 $\bar{X} = (X_1 + X_2 + \dots + X_{400})/400$ はどれくらいばらつくか。

上の例の 1. において $p = 0.2$ として、

$$E(\bar{X}) = 0.2, \quad V(\bar{X}) = \frac{0.2 \times 0.8}{400} = 0.0004, \quad D(\bar{X}) = \sqrt{0.0004} = 0.02. \quad (9)$$

従って、チェビシェフの不等式より、

$$P(0.2 - 2 \times 0.02 \leq \bar{X} \leq 0.2 + 2 \times 0.02) = P(0.16 \leq \bar{X} \leq 0.24) \geq 1 - 1/2^2 = 3/4 \quad (10)$$

- (ii) ある大学の男子学生の身長分布は $N(170, 6^2)$ であるとする。100 人を無作為に選び、その平均 $\bar{X}$ を計算するものとする。上の例の 3. において $\mu = 170, \sigma^2 = 6^2$ として、

$$E(\bar{X}) = 170, \quad V(\bar{X}) = \frac{6^2}{100} = 0.36, \quad D(\bar{X}) = \sqrt{0.36} = 0.6. \quad (11)$$

従って、チェビシェフの不等式より、

$$P(170 - 2 \times 0.6 \leq \bar{X} \leq 170 + 2 \times 0.6) = P(168.8 \leq \bar{X} \leq 171.2) \geq 1 - 1/2^2 = 3/4 \quad (12)$$

この例は裏頁に続く。

9.2 再生性の証明

ベルヌーイ分布、2 項分布、ポアソン分布、正規分布など幾つかの分布は再生性という良い性質を持つ。ポアソン 分布の場合にこれを示しておこう。(一見難しそうだが、実はそれほどでもない。2 項定理が分かれば十分ついてこれる。)

証明すべき内容は、

$$X \sim P_o(\lambda_1), Y \sim P_o(\lambda_2), X \text{ と } Y \text{ は独立} \Rightarrow Z \equiv X + Y \sim P_o(\lambda_1 + \lambda_2) \quad (13)$$

である。即ち、

$$P(Z = z) = e^{-(\lambda_1 + \lambda_2)} \frac{(\lambda_1 + \lambda_2)^z}{z!} \quad (z = 0, 1, 2, \dots) \quad (14)$$

を示せばよい。

$$P(Z = z) = P(X + Y = z) \quad (Z = X + Y) \quad (15)$$

$$= P(\{X = z, Y = 0\} \cup \{X = z - 1, Y = 1\} \cup \dots \cup \{X = 0, Y = z\}) \quad (\text{明ら} \text{か}) \quad (16)$$

$$= P\left(\bigcup_{y=0}^z \{X = z - y, Y = y\}\right) \quad (\text{単なる書き換え}) \quad (17)$$

$$= \sum_{y=0}^z P(\{X = z - y, Y = y\}) \quad (\text{排反}) \quad (18)$$

$$= \sum_{y=0}^z P(\{X = z - y\})P(\{Y = y\}) \quad (\text{独立性}) \quad (19)$$

$$= \sum_{y=0}^z e^{-\lambda_1} \frac{\lambda_1^{z-y}}{(z-y)!} \times e^{-\lambda_2} \frac{\lambda_2^y}{y!} \quad (X \sim P_o(\lambda_1), Y \sim P_o(\lambda_2)) \quad (20)$$

$$= e^{-\lambda_1} e^{-\lambda_2} \sum_{y=0}^z \frac{1}{(z-y)!y!} \times \lambda_1^{z-y} \lambda_2^y \quad (\text{単に整理しただけ}) \quad (21)$$

$$= \frac{e^{-\lambda_1} e^{-\lambda_2}}{z!} \sum_{y=0}^z \frac{z!}{(z-y)!y!} \times \lambda_1^{z-y} \lambda_2^y \quad (\frac{z!}{z!} = 1 \text{ は明ら} \text{か}) \quad (22)$$

$$= \frac{e^{-(\lambda_1 + \lambda_2)}}{z!} \sum_{y=0}^z \frac{z!}{(z-y)!y!} \times \lambda_1^{z-y} \lambda_2^y \quad (e^A e^B = e^{A+B}) \quad (23)$$

$$= \frac{e^{-(\lambda_1 + \lambda_2)}}{z!} \sum_{y=0}^z {}_z C_y \lambda_1^{z-y} \lambda_2^y \quad (2 \text{ 項係数の定義}) \quad (24)$$

$$= \frac{e^{-(\lambda_1 + \lambda_2)}}{z!} (\lambda_1 + \lambda_2)^z \quad (2 \text{ 項定理}) \quad (25)$$

例 9.2. (ポアソン分布の再生性) ある会社に午前中にかかってくる電話の数 X は平均 15 件のポアソン分布 $Po(15)$ に従うとする。また、午後の件数 Y が平均 30 件のポアソン分布 $Po(30)$ に従うならば、一日にかかってくる電話の数 $Z \equiv X + Y$ は平均 45 件のポアソン分布 $Po(45)$ に従う。

また、5 日間を通してかかってくる電話の数は平均 $225 (= 45 \times 5)$ のポアソン分布 $Po(225)$ に従う。何故なら、第 i 日目にかかってくる電話の数を Z_i ($i = 1, 2, 3, 4, 5$) とおけば、これらは互いに独立に同一のポアソン分布 $Po(45)$ に従うから、再生性により和 $Z_1 + Z_2 + \cdots + Z_5 \sim Po(225)$ 。□

例 9.3. (応用例:身長計測) 東大生 n 人を無作為に選んで身長を計測するとしよう。計測結果を $X_1, \dots, X_n$ で表す。 $X_1, \dots, X_n$ は互いに独立に正規分布 $N(\mu, \sigma^2)$ に従うものとする。

$\mu = 170$ 、 $\sigma^2 = 5^2$ 、 $n = 100$ とすれば、正規分布の再生性より、標本平均 $\bar{X} = \frac{1}{100} \sum_{i=1}^{100} X_i$ は

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \left(\text{従って、} Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0, 1) \right) \quad (26)$$

となるから、

$$\bar{X} \sim N\left(170, \frac{5^2}{100}\right) \quad \left(\text{従って、} \frac{\bar{X} - 170}{\sqrt{\frac{5^2}{100}}} = \frac{\bar{X} - 170}{0.5} \sim N(0, 1) \right)$$

例えば、 $169 \leq \bar{X} \leq 171$ となる確率は次式の通り：

$$P(169 \leq \bar{X} \leq 171) = P\left(\frac{169 - 170}{0.5} \leq \frac{\bar{X} - 170}{0.5} \leq \frac{171 - 170}{0.5}\right) \quad (27)$$

$$= P(-2 \leq Z \leq 2) \quad (28)$$

$$= \Phi(2) - \Phi(-2) \quad (29)$$

$$= 0.954 \quad (30)$$

上で、 $\mu = 170$ としたが、 μ の値を知っていることは稀である。むしろ、 $\bar{X}$ の値から μ について何らかの知識を得たいというのが通常であろう。では、 $\sigma^2 = 5$ であることは分かっているが、 μ は未知として議論してみよう。即ち、 $X_1, X_2, \dots, X_n$ は互いに独立に正規分布 $N(\mu, 5^2)$ に従うとする。このとき、

$$\bar{X} \sim N\left(\mu, \frac{5^2}{100}\right) \quad \left(\text{従って、} \frac{\bar{X} - \mu}{\sqrt{\frac{5^2}{100}}} = \frac{\bar{X} - 170}{0.5} \sim N(0, 1) \right) \quad (31)$$

議論の前に次の事実を復習しておく：

$$Z \sim N(0, 1) \quad \text{ならば} \quad P(-1.96 \leq Z \leq 1.96) = 0.95. \quad (32)$$

では、身長の話に戻ろう。公式 (32) を使うと、

$$P\left(-1.96 \leq \frac{\bar{X} - \mu}{\sqrt{\frac{5^2}{100}}} \leq 1.96\right) = 0.95 \quad (33)$$

が成立する。この式を未知の μ に関して整理すると、

$$P\left(\bar{X} - 1.96\sqrt{\frac{5^2}{100}} \leq \mu \leq \bar{X} + 1.96\sqrt{\frac{5^2}{100}}\right) = 0.95 \quad (34)$$

従って、

$$[\bar{X} - 1.96\sqrt{5^2/100}, \bar{X} + 1.96\sqrt{5^2/100}] = [\bar{X} \pm 1.96 \times 0.5] \quad (35)$$

なる区間は確率 0.95 で未知の μ を含むということが分かった。このような区間を、 μ に関する信頼係数 0.95 の信頼区間と言う。仮に、 $\bar{X} = 171$ が得られたとすると、上記の信頼区間は $[171 - 1.96\sqrt{\frac{5^2}{100}}, 171 + 1.96\sqrt{\frac{5^2}{100}}] = [170.02, 171.98]$ となる。

9.3 中心極限定理・大数法則の応用例

例 9.4. (視聴率、表頁の続き) $Ber(p) = B(1, p)$ であることに注意すると、再生性より、

$$T = \sum_{i=1}^{400} X_i \sim B(400, p) = B(400, 0.2) \quad (36)$$

となる。 $0.16 \leq \bar{X} \leq 0.24$ は $64 \leq T \leq 96$ に等しいから、2 項分布を用いて、

$$P(0.16 \leq \bar{X} \leq 0.24) = P(64 \leq T \leq 96) = \sum_{x=64}^{96} {}_{400}C_x (0.2)^x (0.8)^{400-x} \quad (37)$$

として求めることが出来るが、やや煩雑である。

$n = 400$ は十分大と考えられるので、中心極限定理を利用し、正規分布で近似しよう。

$$\bar{X} \sim N(p, p(1-p)/n) = N(0.2, (0.02)^2) \quad (38)$$

であるから、

$$Z = \frac{\bar{X} - p}{\sqrt{p(1-p)/n}} = \frac{\bar{X} - 0.2}{0.02} \sim N(0, 1). \quad (39)$$

よって、

$$P(0.16 \leq \bar{X} \leq 0.24) = P\left(\frac{0.16 - 0.2}{0.02} \leq \frac{\bar{X} - 0.2}{0.02} \leq \frac{0.24 - 0.2}{0.02}\right) \quad (40)$$

$$= P(-2 \leq Z \leq 2) = \Phi(2) - \Phi(-2) = 0.954. \quad (41)$$

練習のため、 $P(0.23 \leq \bar{X})$ も求めよう。

$$P(0.23 \leq \bar{X}) = P\left(\frac{0.23 - 0.2}{0.02} \leq \frac{\bar{X} - 0.2}{0.02}\right) \quad (42)$$

$$= P(1.5 \leq Z) = 1 - \Phi(1.5) = 0.067. \quad (43)$$

例 9.5. (応用例：コイン投げ) $X_1, \dots, X_n$ は互いに独立に同一のベルヌーイ分布 $Ber(p)$ に従うものとする (長さ n のベルヌーイ試行)。このとき、既に学んだ通り、

$$E(\bar{X}) = p, \quad V(\bar{X}) = \frac{p(1-p)}{n}. \quad (44)$$

$p = 0.5$ 、 $n = 10000$ としてみる。このとき、 $E(\bar{X}) = 0.5$ 、 $V(\bar{X}) = \frac{(0.5)^2}{10000} = (0.005)^2$ である。以前の講義では、チェビシェフの不等式を用いて

$$P(0.5 - 4 \times 0.005 \leq \bar{X} \leq 0.5 + 4 \times 0.005) = P(0.48 \leq \bar{X} \leq 0.52) \geq 1 - \frac{1}{4^2} = 0.9375 \quad (45)$$

を得た。今回は中心極限定理をこの問題に応用してみよう。中心極限定理を使うと (44) に加えて次のことが分かる：

$$\bar{X} \overset{\text{近似的に}}{\sim} N\left(p, \frac{p(1-p)}{n}\right) \quad (46)$$

従って、

$$Z = \frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \overset{\text{近似的に}}{\sim} N(0, 1). \quad (47)$$

ここで、上と同様に $p = 0.5$ 、 $n = 10000$ としてやると

$$\bar{X} \overset{\text{近似的に}}{\sim} N(0.5, 0.005^2). \quad \text{従って } Z = \frac{\bar{X} - 0.5}{0.005} \overset{\text{近似的に}}{\sim} N(0, 1) \quad (48)$$

このとき、(45) は次のように精緻化出来る：

$$P\left(0.5 - 4 \times 0.005 \leq \bar{X} \leq 0.5 + 4 \times 0.005\right) = P\left(-4 \leq \frac{\bar{X} - 0.5}{0.005} \leq 4\right) \quad (49)$$

$$\approx \Phi(4) - \Phi(-4) = (1 - 0.000031671) - 0.000031671 \approx 0.99994 \quad (50)$$

同様にして、例えば、 $P\left(0.5 - 2 \times 0.005 \leq \bar{X} \leq 0.5 + 2 \times 0.005\right) \approx 0.955$ なることも分かる。