応用統計学 2017 確率に関する補足資料

2017 年 10 月 25 日 (水)

清 智也 sei@mist.i.u-tokyo.ac.jp

推測統計を学ぶには，確率の基礎を知っている必要がある。参考書として

- 伏見 正則「確率と確率過程」朝倉書店
- 藤澤 洋徳「確率と統計」朝倉書店

を挙げておく。

演習問題

問題 5-A1. 事象 A の確率を $P(A)$ と書く。事象 A, B に対して $P(A) = 0.9$，$P(B) = 0.7$ のとき，$P(A \cap B) \geq 0.6$ が成り立つことを示せ。

問題 5-A2. n 個の箱と n 個のボールがあり，箱には 1 から n の番号が書かれている。各ボールを独立にランダムな番号の箱に入れたとする。このとき以下の問いに答えよ。

(a) 1 番目の箱に k 個のボールが入っている確率を求めよ。ただし $0 \leq k \leq n$ とする。

(b) 問 (a) で求めた結果は，k を固定して $n \to \infty$ とするとどんな値に近づくか。

(c) 各ボールが入った箱の番号を $\{I_j\}_{j=1}^n$ とおく。また $x_1, \ldots, x_n$ を実数の定数とする。確率変数 $X = n^{-1} \sum_{j=1}^n x_{I_j}$ の期待値と分散を計算せよ。

問題 5-A3 (ベイズの定理). ある製品が故障したときの症状 S と，その原因 G について調べたところ，以下のような確率で発生していることがわかったとする。

故障原因 G_i	G_1		G_2		G_3	
確率 $P(G_i)$	0.60		0.10		0.30	
症状 S_j	S_1	S_2	S_1	S_2	S_1	S_2
条件付き確率 $P(S_j \mid G_i)$	0.40	0.60	0.10	0.90	0.80	0.20

(例：製品をプリンタとして，$S_1 =$ 紙詰まり，$G_1 =$ ローラーの汚れ，など。)

(a) 症状が S_1 という条件の下で，故障原因が G_1 である確率を求めよ。ただし答えは小数点以下第 2 位までに丸めよ。

(b) 症状が S_2 という条件の下で，故障原因として最も確率の高いのは G_1, G_2, G_3 のうちどれか。

問題 5-A4. 正の値のみを取る確率変数 X に対して，

$$\frac{\sqrt{\mathrm{V}[X]}}{\mathrm{E}[X]}$$

のことを変動係数 (coefficient of variation) と呼ぶ。ここで $\mathrm{E}[X]$ は X の期待値，$\mathrm{V}[X]$ は X の分散を表す。

(a) パラメータ $\lambda > 0$ の指数分布に対して，変動係数を計算せよ。ただしパラメータ λ の指数分布の確率密度関数は $f(x) = \lambda e^{-\lambda x}$ $(x > 0)$ と定義される。

(b) パラメータ $\alpha, \beta > 0$ のガンマ分布に対して，変動係数を計算せよ。ただしパラメータ α, β のガンマ分布の確率密度関数は $f(x) = \beta^\alpha x^{\alpha-1} e^{-\beta x}/\Gamma(\alpha)$ $(x > 0)$ と定義される。また $\Gamma(\alpha) = \int_0^\infty z^{\alpha-1} e^{-z} dz$ はガンマ関数を表す。

問題 5-A5. 累積分布関数が

$$F(x) = \frac{1}{1+e^{-x}}, \quad -\infty < x < \infty,$$

で与えられる確率変数 X の期待値と分散を求めよ。

（Hint: 分散の計算では，$u = 1/(1+e^{-x})$ と変換するとよい。）

問題 5-A6. 平均ベクトル $\boldsymbol{\mu} \in \mathbb{R}^d$，共分散行列 $\boldsymbol{\Sigma} \in \mathbb{R}^{d \times d}$ の多変量正規分布の確率密度関数は

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/2}$$

と定義される。ただし $\boldsymbol{\Sigma}$ は正定値対称行列とする。これが実際に確率密度関数になっていることを示せ。すなわち，$f(\boldsymbol{x}) \geq 0$ かつ $\int f(\boldsymbol{x}) d\boldsymbol{x} = 1$ となることを示せ。ただし公式 $\int_{-\infty}^{\infty} (2\pi)^{-1/2} e^{-x^2/2} dx = 1$ は既知としてよい。

（Hint: 行列の対角化，あるいは Cholesky 分解を用いよ。）

問題 5-A7. 確率変数 $X_1, \ldots, X_n$ は独立で，以下の確率関数を持つ分布に従うとする：

$$f(x) = (1-p)p^x, \quad x = 0, 1, \cdots.$$

ただし $0 < p < 1$ とする。このとき以下の問いに答えよ。

(a) X_1 のモーメント母関数 $M(\theta) = \mathrm{E}[e^{\theta X_1}]$ を求めよ。ただし θ の範囲は適切に設定せよ。

(b) 大数の法則を使って $\bar{X}_n = (X_1 + \cdots + X_n)/n$ の確率収束先 μ を求めよ。

(c) 中心極限定理を使って $\sqrt{n}(\bar{X}_n - \mu)$ の分布収束先を求めよ。

資料：代表的な確率分布

1 変量離散分布

ベルヌーイ分布 $\mathrm{Bernoulli}(p)$	$f(x) = p^x(1-p)^{1-x}$	$x \in \{0, 1\}$
二項分布 $\mathrm{Bin}(N, p)$	$f(x) = \binom{N}{x} p^x (1-p)^{N-x}$	$x \in \{0, 1, \ldots, N\}$
ポアソン分布 $\mathrm{Po}(\lambda)$	$f(x) = \lambda^x \mathrm{e}^{-\lambda}/x!$	$x \in \{0, 1, \ldots\}$
負の二項分布 $\mathrm{NB}(r, p)$	$f(x) = \binom{r+x-1}{x} p^r (1-p)^x$	$x \in \{0, 1, \ldots\}$

ベルヌーイ分布

$\downarrow$ 　和 $\sum_i X_i$

二項分布

$\downarrow$ 　極限

ポアソン分布

1 変量連続分布

正規分布 $\mathrm{N}(\mu, \sigma^2)$	$f(x) = (2\pi\sigma^2)^{-1/2} \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}$	$x \in \mathbb{R}$
指数分布 $\mathrm{Ex}(\lambda)$	$f(x) = \lambda \mathrm{e}^{-\lambda x}$	$x > 0$
ガンマ分布 $\mathrm{Gamma}(\alpha, \beta)$	$f(x) = \beta^\alpha x^{\alpha-1} \mathrm{e}^{-\beta x}/\Gamma(\alpha)$	$x > 0$
ベータ分布 $\mathrm{Beta}(\alpha_1, \alpha_2)$	$f(x) = x^{\alpha_1 - 1}(1-x)^{\alpha_2 - 1}/B(\alpha_1, \alpha_2)$	$0 < x < 1$

標準正規分布

$\downarrow$ 　二乗和 $\sum_i X_i^2$

カイ二乗分布

$\cap$

指数分布 $\subset$ **ガンマ分布**

$\downarrow$ 　割合 $X/(X+Y)$

$[0, 1]$ 上の一様分布 $\subset$ **ベータ分布**

$\downarrow$ 　変数変換

F 分布

$\downarrow$ 　特殊ケースの平方根

コーシー分布 $\subset$ t 分布

$\cap$ 　 $\downarrow$ 　極限

安定分布 $\supset$ **標準正規分布**

多変量分布

多項分布 $\mathrm{Multi}(N, \boldsymbol{p})$	$f(\boldsymbol{x}) = \frac{N!}{x_1! \cdots x_k!} \prod_{i=1}^k p_i^{x_i}$	$x_i \in \{0, 1, \ldots, N\}$，$\sum_{i=1}^k x_i = N$

多変量正規分布 $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	$f(\boldsymbol{x}) = \lvert 2\pi\boldsymbol{\Sigma}\rvert^{-1/2} \mathrm{e}^{-(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})/2}$	$\boldsymbol{x} \in \mathbb{R}^k$
ディリクレ分布 $\mathrm{Dir}(\boldsymbol{\alpha})$	$f(\boldsymbol{x}) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k x_i^{\alpha_i - 1}$	$x_i > 0, \sum_{i=1}^k x_i = 1$

注：ディリクレ分布の密度は，例えば $(x_1, \ldots, x_{k-1})$ に関する密度を表す．残りの 1 成分 x_k は $\sum_{i=1}^k x_i = 1$ より一意に定まる．