

🔍 🔍 🔍 課題問題 中間期（6月） 🔍 🔍 🔍

★提出期日 2019 年 6 月 10 日、工学系は「工学」問題を含めすべての解答をすること、文系はこの「工学」問題を含めてもよいとします。解答には「別紙」を用いること。

問題 1（期待値と分散） 箱のなかには 4 個の白ボールと 3 個の赤ボールがあり、ここからボールの「非復元抽出」をおこなう。取り出しの結果、白ボールであれば、 w と表し、赤ボールでは、 r と表すとき、この取り出しの事象系列についてつぎの問いに答えよ。

- (i) 事象系列が $rrwr$ (赤赤白赤の順) となる結果の確率をもとめよ。
- (ii) 確率変数 X を初めて白ボールが出るまでの回数とすると、この確率分布をもとめよ。
- (iii) 平均 $E(X)$ と分散 $V(X)$ をもとめよ。

工学 (iv) もし取り出しを「復元抽出」であるとすれば、この場合の確率分布とその平均を分散をもとめよ。

ヒント : p.37 離散型確率分布の平均と分散、(ii) p.53 幾何分布、回数の値 $k = 1, 2, 3, \dots$ とその確率 $P(X = k)$ の対応表。初めて起こる事象の確率は幾何分布となる。一般に取り出し回数を固定したときの取り出し確率が変わらないから、復元抽出では 2 項分布であり、非復元抽出では抽出結果により、次時点の生起確率が異なる超幾何分布となる。

問題 2（結合分布の期待値） つぎの 2 元表は離散型確率変数の結合分布を表す。2 つの確率変数 X, Y の取り得る値とその確率 $P(X = i, Y = j) = p_{ij}$ を行列で表したもの。

" $X = i$ " \ " $Y = j$ "	10	11	12	計
20	0.15	0.25	0.15	0.55
30	0.15	0.10	0.20	0.45
計	0.30	0.35	0.35	1

- (i) $E(X)$ と $E(Y)$ をもとめよ。
- (ii) $E(X) + 2E(Y)$ をもとめよ。
- (iii) 確率変数 $Z = X + 2Y$ の確率分布をもとめ、等式 $E(X + 2Y) = E(X) + 2E(Y)$ を確かめよ。

ヒント : p.39 結合確率分布、2 元表での X, Y の周辺和から、周辺分布をもとめる。

問題 3（2 項分布） 2 項分布でのパラメータ: 繰り返し数 n , 生起確率 (成功の確率) p とするとき、確率変数 $X \sim \text{Binom}(n, p)$ と表す。

- (i) 確率変数 $Y = n - X$ の確率分布をもとめよ。
- (ii) $E(Y)$ をもとめよ。

工学 (iii) 確率変数 X の分散と Y の分散をもとめよ。

ヒント : (i), (ii) では裏 (失敗) の出る確率分布を求める。(iii) 分散の計算は 2 次モーメント $E(X^2)$ を計算するが、直接の計算ではなく、2 項係数 ${}_nC_k = \frac{n!}{k!(n-k)!} = \binom{n}{k}$ の形から $k \times {}_nC_k = k \cdot \frac{n!}{k!(n-k)!} = n {}_{n-1}C_{k-1} = n \binom{n-1}{k-1}$ および $k(k-1) \times {}_nC_k = n(n-1) {}_{n-2}C_{k-2} = n(n-1) \binom{n-2}{k-2}$ をもちいよ。つまり $X^2 = X(X-1) + X$ の関係!

問題 4（正規分布） 5 章の問題での 5.1 (a) ~ (h) を解け。正規分布表は付表 2 正規分布表と付表 3 正規分布のパーセント点をもちいる。

追加資料の正規分布表は今後もちいるので保存してください。

(以上)

問題 1 の解(i) 第 i 回目の結果を X_i とすると、 $P(X_1 = r) = \frac{3}{7}$,

$$P(X_1 = r, X_2 = r) = P(X_2 = r | X_1 = r)P(X_1 = r) = \frac{2}{6} \times \frac{3}{7} = \frac{3 \cdot 2}{7 \cdot 6},$$

以下は $X_{12} = rr$ などと略記すると、

$$P(X_{123} = rrw) = P(X_3 = w | X_{12} = rr)P(X_{12} = rr) = \frac{4}{5} \times \frac{3 \cdot 2}{7 \cdot 6} = \frac{3 \cdot 2 \cdot 4}{7 \cdot 6 \cdot 5},$$

$$P(X_{1234} = rrwr) = P(X_4 = r | X_{123} = rrw)P(X_{123} = rrw) = \frac{1}{4} \times \frac{3 \cdot 2 \cdot 4}{7 \cdot 6 \cdot 5} = \frac{3 \cdot 2 \cdot 4 \cdot 1}{7 \cdot 6 \cdot 5 \cdot 4} = \frac{(3)_3(4)_1}{(7)_4}.$$

最後は減少数列の記号で表しています。このような事象は順序をかえても同じ値になることを確かめてみましょう。

$$P(X_{1234} = rrrw) = P(X_{1234} = rrwr) = P(X_{1234} = rwrr) = P(X_{1234} = wrrr) = \frac{(3)_3(4)_1}{(7)_4}$$

比較のために、もし元にもどすならば、取り出し確率は赤 r は $\frac{3}{7}$ で白 w は $\frac{4}{7}$ ですから、上と同じ記号で

$$P(X_{1234} = rrrw) = P(X_{1234} = rrwr) = P(X_{1234} = rwrr) = P(X_{1234} = wrrr) = \left(\frac{3}{7}\right)^3 \left(\frac{4}{7}\right)^1.$$

となりますから、2 項分布です。

(ii) と (iii) は例題 5(p.36) と同じですから略します。(iv) は復元ですから、取りつくすということがありません。初めてということですから、たとえば “r r r w” であれば、 $X = 4$ とします。ですから $P(X = 4) = \left(\frac{3}{7}\right)^3 \left(\frac{4}{7}\right)^1$ となり、この X の取り得る値は $\{1, 2, 3, \dots\}$ と続きます。これを定数 $p = \frac{4}{7}$ $q = 1 - p = \frac{3}{7}$ をもちい、 k 回目ですべて初めてならば、 $P(X = k) = q^{k-1}p$ となります。

幾何分布の平均 $\mu = E(X) = \frac{1}{p}$ と分散 $\sigma^2 = V(X) = E(X^2) - \mu^2 = \frac{q}{p^2}$ は

$$E(X) = \sum k \times pq^{k-1} = p + 2pq + 3pq^2 + 4pq^3 + \dots = p(1 + 2q + 3q^2 + \dots) = p \frac{1}{(1-q)^2} = \frac{1}{p} \text{ となります。}$$

ここで最後の級数の和を計算するとき、

$$\begin{aligned} 1 + 2q + 3q^2 + 4q^3 + 5q^4 + \dots &= 1 + q + q^2 + q^3 + q^4 + \dots \\ &\quad + q + q^2 + q^3 + q^4 + \dots \\ &\quad + q^2 + q^3 + q^4 + \dots \\ &\quad + q^3 + q^4 + \dots \\ &\quad + q^4 + \dots \\ &= \frac{1}{1-q} + q \left(\frac{1}{1-q} \right) + q^2 \left(\frac{1}{1-q} \right) + \dots \\ &= \frac{1}{1-q} (1 + q + q^2 + q^3 + \dots) = \frac{1}{(1-q)^2} = \frac{1}{p^2} \end{aligned}$$

まず

$$\begin{aligned} E(X^2) &= \sum k^2 \times pq^{k-1} \\ &= p + 2^2pq + 3^2pq^2 + 4^2pq^3 + \dots \\ &= \sum [k(k-1)pq^{k-1} + kpq^{k-1}] \\ &= pq \sum [k(k-1)q^{k-2}] + p \sum [kq^{k-1}] \\ &= pq \times \frac{2q}{(1-q)^3} + p \times \frac{1}{(1-q)^2} = \frac{2q^2}{p^2} + \frac{1}{p} \end{aligned}$$

したがって

$$V(X) = E(X^2) - \mu^2 = \left(\frac{2q^2}{p^2} + \frac{1}{p} \right) - \left(\frac{1}{p} \right)^2 = \frac{q}{p^2}$$

ここで

$$\sum [k(k-1)q^{k-2}] = 2 \cdot 1 + 3 \cdot 2q + 4 \cdot 3q^2 + 5 \cdot 4q^3 + \cdots = \frac{2q}{(1-q)^3}$$

となることは、平均の計算で得られた関係式 $1 + 2q + 3q^2 + 4q^3 + 5q^4 + \cdots = \frac{1}{(1-q)^2} = (1-q)^{-2}$ において、両辺の微分をすることで得られます。すなわち左辺は $2 + 3 \cdot 2q + 4 \cdot 3q^2 + 5 \cdot 4q^3 + \cdots$ であり、右辺は $(-2)(1-q)^{-2-1}(-1) = 2(1-q)^{-3} = \frac{2}{p^2}$ となることから結果が得られます。幾何分布の平均と分散は p.53 にありますし、その $p = 0.4$ のグラフもあります。この幾何分布は、初めて起こるということと、その裏返しの連続して起こる回数分布にも同じことです。とり得る値が違うだけです。これは離散型分布ですが、連続型の場合は、故障が起こる時間分布やお客の到着時間、待ち時間などでよく用いられる指数分布が対応しています。その密度関数は、正の実数パラメータ $\lambda > 0$ より、時刻 t を変数として

$$f_X(t) = \frac{1}{\lambda} e^{-t/\lambda}, \quad 0 \leq t < \infty$$

と定め、平均 $\mu = \lambda$, 分散 $\sigma^2 = \lambda$ となります。

問題 2 の解 まず結合分布からそれぞれの周辺分布を求めます。

X の(周辺)分布は	X の値	20	30	計	,	Y の(周辺)分布は	X の値:	10	11	12	計
	確率	0.55	0.45	1			確率:	0.30	0.35	0.35	1

これから、期待値 $E(X)$, $E(Y)$ を求めることができます。すなわち

(i) X の期待値は $E(X) = 20 \times 0.55 + 30 \times 0.45 = 24.5$, Y の期待値は $E(Y) = 10 \times 0.30 + 11 \times 0.35 + 12 \times 0.35 = 11.05$

(ii) $E(X) + 2E(Y) = 24.5 + 2 \times 11.05 = 46.6$

(iii) $E(X + 2Y) = (20 + 2 \cdot 10)0.15 + (20 + 2 \cdot 11)0.25 + (20 + 2 \cdot 12)0.15 + (30 + 2 \cdot 10)0.15 + (30 + 2 \cdot 11)0.10 + (30 + 2 \cdot 12)0.20 = 23.1 + 23.5 = 46.6$

問題 3 の解 (i) もし $X = k$ であれば、 $Y = n - k$ に 1 対 1 に対応するから、 X の値と $Y = n - X$ の値を対応させて分布表にすると

X の値:	0	1	2	$\cdots$	k	$\cdots$	n
$Y = n - X$:	n	$n - 1$	$n - 2$	$\cdots$	$n - k$	$\cdots$	0
確率:	p^n	$\binom{n}{1} p^{n-1} q$	$\binom{n}{2} p^{n-2} q^2$	$\cdots$	$\binom{n}{k} p^{n-k} q^k$	$\cdots$	$\binom{n}{n} p^{n-n} q^n$

ここで $\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!}$ は 2 項係数を表している。

(ii) 上で求めた分布表から、

$$E(Y) = n \cdot p^n + (n-1) \binom{n}{1} p^{n-1} q + (n-2) \binom{n}{2} p^{n-2} q^2 + \cdots + 0 \binom{n}{n} p^{n-n} q^n = n(1-p) = nq$$

X は n 枚のコインを投げたとき、表の出る枚数の確率を表すが、それから定義される確率変数 Y は、裏の枚数を表す。したがって、値は計算を $E(Y) = n - E(X) = n - np = nq$ と直感的にも得られる。

(iii) 2 項分布の平均と分散の計算は p.51 から、 X に対して、 μ_X, σ_X^2 と表すと、表が p であるから、 $\mu_X = np$, $\sigma_X^2 = npq = np(1-p)$ である。問題 1 での $X^2 = X(X-1) + X$ の工夫で分散は得られる。 Y

については、“表（成功）の確率”が q であるから、平均は nq で、分散も $nq(1-q) = npq$ と得られるはずである。

問題4の解 正規分布表を正確に使うためには、平均 μ と分散（標準偏差） σ^2 と $\sigma > 0$ を図に表すことがよい。一般の分布 $N(\mu, \sigma^2)$ と標準化した分布 $N(0, 1)$ との対応は、正と負、大小関係の不等号に関する領域の把握をすれば、極めて易しい。p.68, p.69 例題1, 2（正規分布表の使い方）を復習する。

重要な関係式

$$X \sim N(\mu, \sigma^2) \Leftrightarrow Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

$$P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = P\left(Z \leq \frac{x - \mu}{\sigma}\right) = P(Z \leq z) = \Phi(z)$$

ただし $z = \frac{x - \mu}{\sigma}$ とする。

このような計算をパソコンソフトでよく用いられる表計算ソフト（Microsoft Excel, 無料総合オフィスソフトウェア Apache Openoffice）などでもちいると効率が良い。より専門性の高い統計手法のソフトでは「統計ソフト R」がよく用いられ、参考書も多数。

<https://sites.google.com/site/webtextofr/home>(名古屋市立大学)などを参照のこと。

<http://www.math.s.chiba-u.ac.jp/~yasuda/statA/stat-excel-12.pdf>

<https://support.office.com/ja-jp/article/>

◇ 正規分布に関連したコマンド名（書式引数）：正規分布でのパラメータ（平均 μ 、標準偏差 σ ）を指定し、 x 値に対する関数形式を (true/1) では確率密度関数 $N(\mu, \sigma^2)$ における $\phi(z)$, ($z = \frac{x - \mu}{\sigma}$) の値や 関数形式を (false/0) では累積確率値 $P(X \leq x) = \Phi(x) = \int_{-\infty}^x \phi(t)dt$ が出力される。

NORM.DIST($x, \mu, \sigma, 1$)	$N(\mu, \sigma^2)$ の $\phi(z)$, ($z = \frac{x - \mu}{\sigma}$) の値（高さ）
NORM.DIST($x, \mu, \sigma, 0$)	$N(\mu, \sigma^2)$ の $\Phi(z) = \int_{-\infty}^z \phi(t)dt$, ($z = \frac{x - \mu}{\sigma}$) の値（確率、積分値）
NORM.INV(α, μ, σ)	累積正規分布の逆関数の値 x ; $\alpha = \Phi(z)$, $z = \frac{x - \mu}{\sigma}$ を求める
NORM.S.DIST($z, 1$)	標準正規分布 $N(0, 1)$ の密度関数値（高さ） $\phi(z)$ を求める
NORM.S.DIST($z, 0$)	標準正規分布 $N(0, 1)$ の累積確率（積分値） $\Phi(z)$ を求める
NORM.S.INV(α)	標準正規分布 $N(0, 1)$ の逆関数値 z ; $\alpha = \Phi(z)$
PHI(x)	標準正規分布の確率 $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$ を求める (2013) =NORM.DIST($x, 0, 1, 1$) と同じ
GAUSS(z)	[例] ‘=PHI(0.75)’ 標準正規分布の密度関数の値。0.301137432 標準正規分布で平均から累積確率を求める (2013) [例] ‘=GAUSS(2)’ 結果は 0.47725 標準正規母集団において平均と平均から標準偏差の 2 倍の範囲になる確率。両側パーセント点の値