「確率とその応用」ノート、その伍

逆瀬川浩孝

5　条件付き期待値

5.1　結合分布、相関係数、共分散

例題 5.1（結合分布、独立性）確率変数 X, Y の結合確率関数 $f_{X,Y}(x, y) = P(X = x, Y = y)$ が下のように与えられているものとします。このとき、以下の問いに答えなさい。**(1)** XY の期待値を計算しなさい。**(2)** X の周辺確率関数を求め、X の期待値を計算しなさい。**(3)** Y の周辺確率関数を求め、Y の期待値を計算しなさい。**(4)** $f_{X,Y}(x, y) = f_X(x) \times f_Y(y)$ がすべての x, y に対して成り立っていることを確かめなさい。で、何が言えましたか。**(5)** $E(XY) - E(X)E(Y)$ を計算しなさい。で、何が分かりましたか。

$X\diagdown Y$	-1	1
-1	0.15	0.45
1	0.1	0.3

こたえ　**(1)**

$$\begin{aligned} P(XY = 1) &= P(X = Y = \pm 1) = 0.45 \\ P(XY = -1) &= 1 - P(XY = 1) = 0.55 \\ &\Rightarrow E(XY) = -0.1 \end{aligned}$$

(2)

$$P(X = -1) = 0.6, P(X = 1) = 0.4 \Rightarrow E(X) = -0.2$$

(3)

$$P(Y = -1) = 0.25, P(Y = 1) = 0.75 \Rightarrow E(X) = 0.5$$

(4)

$$P(X = -1, Y = -1) = 0.15 = 0.6 \times 0.25 = P(X = -1)P(Y = -1)$$

などなど、4 通りすべての組み合わせで $f_{X,Y}(x, y) = f_X(x) \times f_Y(y)$ が成り立つ。したがって、X, Y は互いに独立。

(5) $E(X) = -0.2, E(Y) = 0.5$ より、

$$E(XY) - E(X)E(Y) = 0$$

独立ならば、積の期待値は期待値の積。

練習問題 5.1 X, Y の結合確率関数が次のように与えられているものとします。このとき、**(1)** $X + Y$ の確率関数を求め、その平均（期待値）と分散を計算しなさい。**(2)** XY の確率関数を求め、その期待値を計算し、共分散を計算しなさい。**(3)** $X + Y$ の分散から X の分散と Y の分散の合計を引いたものと共分散の関係を求めなさい。

$X\diagdown Y$	0	1	2
0	0.2	0.1	0.1
1	0.2	0.3	0.1

例題 5.2（共分散）X, Y の結合確率関数が次のように与えられているとき、共分散と相関係数を計算しなさい。相関係数の大きさの違いを結合確率関数の特徴と結び付けて説明しなさい。

A タイプ

$X \setminus Y$	-1	$+1$
-1	0.25	0.25
$+1$	0.25	0.25

B タイプ

$X \setminus Y$	-1	$+1$
-1	0.4	0.1
$+1$	0.2	0.3

C タイプ

$X \setminus Y$	-1	$+1$
-1	0.2	0.4
$+1$	0.3	0.1

こたえ　$E(X)E(Y) = 0$ としたので、これくらいは暗算でも。

	$E(X)$	$V(X)$	$E(Y)$	$V(Y)$	$C(X,Y)$	$R(X,Y)$
A タイプ	0	1	0	1	0	0
B タイプ	0	1	-0.2	0.96	0.4	$1/\sqrt{6}$
C タイプ	-0.2	0.96	0	1	-0.4	$-1/\sqrt{6}$

X が大きければ Y も大きく、X が小さければ Y も小さい、という関係があると相関係数が大きくなる。逆に、X, Y が反対方向に増減するならば、相関係数は小さくなる

例題 5.3（条件付き期待値）表が出る確率が p のコインを 3 回投げる試行を行って表のでた回数を X とします。2 回目に表がでたことは分かっているものとしたとき、**(1)** X の条件付き確率関数を求めなさい。**(2)** X の条件付き期待値を期待値を計算し、条件のない場合の期待値と比較しなさい。**(3)** 2 回目に裏が出たということが分かっている場合に同様な分析をしなさい。**(4)** 2 回目に表が出るか裏が出るか、という確率を使って、X の（条件なし）期待値を計算しなさい。

こたえ　**(1)** i 回目に表が出るという事象の定義関数を $\mathbf{1}_i$ とすると、

$$\begin{aligned} P(X = 1 \mid \mathbf{1}_2 = 1) &= P(\mathbf{1}_1 + \mathbf{1}_3 = 0) = (1-p)^2 \\ P(X = 2 \mid \mathbf{1}_2 = 1) &= P(\mathbf{1}_1 + \mathbf{1}_3 = 1) = 2p(1-p) \\ P(X = 3 \mid \mathbf{1}_2 = 1) &= P(\mathbf{1}_1 + \mathbf{1}_3 = 2) = p^2 \end{aligned}$$

(2) **(1)** の結果から、2 回目に表が出たという条件の下で $X - 1$ はパラメータ $2, p$ の 2 項分布に従うことを意味する。したがって、その期待値は $2p + 1$。条件がない場合は X はパラメータ $3, p$ の 2 項分布に従うので、その期待値は $3p = 2p + p < 2p + 1$ なので、（表が出たという）情報が増えた分だけ期待値が増加している。

(3) **(2)** と同様に考えて、裏が出たという条件の下で X はパラメータ $2, p$ の 2 項分布に従う。したがって、その期待値は $2p$、これは条件なし期待値より小さい。

(4) 条件付き期待値に確率を掛けることで、条件なし期待値を計算することができる。

$$\begin{aligned} E(X) &= E(X \mid \mathbf{1}_2 = 0)P(\mathbf{1}_2 = 0) + E(X \mid \mathbf{1}_2 = 1)P(\mathbf{1}_2 = 1) \\ &= 2p(1-p) + (2p+1)p = 3p \end{aligned}$$

練習問題 5.2 確率変数 X, Y に対して、$E(X) = 2, E(Y) = 5, V(X) = 1, V(Y) = 4, R(X,Y) = -0.8$ としたとき、$X + Y$ の平均と分散を計算しなさい。

練習問題 5.3 $Y = 0$ のとき X の期待値は 10、$Y = 1$ のとき X の期待値は 5、$Y = 2$ のとき X の期待値は 1、と与えられている。Y はパラメータ $2, 0.3$ の 2 項分布に従っているとき、X の条件なしの期待値を計算しなさい。

練習問題 5.4 Y は $\{2, 3, ..., n\}$ で一様分布する確率変数、X は $Y = k$ のとき、$\{1, 2, ..., k-1\}$ で一様分布する確率変数としたとき、$E(X)$ を計算しなさい。

ヒント まず $E(X \mid Y=k)$ を計算し、$E(X)=E(E(X \mid Y))$ という公式を使いなさい。

練習問題 5.5 コインを 3 回投げる。最初の 2 回で表の出る回数を X、後の 2 回で表の出る回数を Y としたときに、X, Y の結合確率関数を書き、X, Y の共分散を計算しなさい。また、相関係数も計算しなさい。

練習問題 5.6 結合確率関数が次のように与えられている二つの確率変数 X, Y に対して、以下の問いに答えなさい。**(1)** $E(X), E(Y), E(XY)$ を計算しなさい。**(2)** X, Y の共分散 $C(X, Y)$ を計算しなさい。**(3)** 相関係数 $R(X, Y)$ が 0 になることを説明しなさい。**(4)** X, Y は独立ですか、理由を付けて答えなさい。**(5)** 相関係数と独立性の関係を説明しなさい。

$X\diagdown Y$	-1	0	1
-1	0	0.1	0
0	0.1	0.6	0.1
1	0	0.1	0

練習問題 5.7 確率変数 X, Y の結合確率関数 $f_{X,Y}(x,y)=P(X=x, Y=y)$ が下のように与えられているものとします。このとき以下の問いに答えなさい。**(1)** XY の期待値を計算しなさい。**(2)** X の周辺確率関数を求め、X の期待値を計算しなさい。**(3)** Y の周辺確率関数を求め、Y の期待値を計算しなさい。**(4)** X, Y の共分散 $C(X, Y)$ を計算しなさい。**(5)** $X=-1$ という条件の下で、$Y=-1$ の確率 $f_{Y|X}(-1 \mid -1)$ を計算しなさい。**(6)** $X=-1$ という条件の下で、$Y=1$ の確率 $f_{Y|X}(1 \mid -1)$ を計算しなさい。**(7)** $X=-1$ という条件の下で、Y の期待値 $E(Y \mid X=-1)$ を計算しなさい。**(8)** $X=1$ という条件の下で、Y の期待値 $E(Y \mid X=1)$ を計算しなさい。**(9)** $\sum_x E(Y \mid X=x) \times P(X=x)$ を計算して、$E(Y)$ の値と比べなさい。

$X\diagdown Y$	-1	0	1
-1	0.1	0.1	0
0	0.1	0.2	0.15
1	0.05	0.2	0.1

例題 5.4 二つの確率変数 X, Y の確率関数を $f_{X,Y}(x,y)$ としたとき、$X+Y$ の期待値が X の期待値と Y の期待値の和になることを証明しなさい。連続で密度関数が $f_{X,Y}(x,y)$ で与えられた場合について、同様なことを証明しなさい。「**和の期待値は期待値の和**」

こたえ 定義にしたがって、

$$
\begin{aligned}
E(X+Y) &= \sum_{x,y}(x+y) f_{X,Y}(x,y) \\
&= \sum_{x,y} x f_{X,Y}(x,y) + \sum_{x,y} y f_{X,Y}(x,y) \\
&= \sum_{x} x \sum f_{X,Y}(x,y) + \sum_{y} y \sum_{x} f_{X,Y}(x,y) \\
&= \sum_{x} x P(X=x) + \sum_{y} y P(Y=y) = E(X) + E(Y)
\end{aligned}
$$

連続の場合は $\sum$ を $\int$ に替えるだけ。

例題 5.5 連続な二つの確率変数 X, Y が互いに独立であるとき、$X \times Y$ の期待値を計算しなさい。

こたえ X, Y の結合密度関数を $f_{X,Y}(x,y)$ とし、X, Y の周辺密度関数をそれぞれ $f_X(x), f_Y(y)$ とすると、X, Y が互いに独立ならば

$$f_{X,Y}(x,y) = f_X(x) \times f_Y(y)$$

が成り立つ。したがって

$$\begin{aligned}E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \times f_{X,Y}(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \times f_X(x) \times f_Y(y)dxdy \\ &= \int_{-\infty}^{\infty} x \times f_X(x)dx \int_{-\infty}^{\infty} y \times f_Y(y)dy = E(X)E(Y)\end{aligned}$$

したがって、「**確率変数が互いに独立ならば、積の期待値は期待値の積**」

例題 5.6 $X_1, X_2, \ldots$ は互いに独立で同じ分布に従う確率変数で、その期待値は $m = E(X_1)$ とします。また、N はそれらとは独立に非負の整数値を取る確率変数とします。このとき、$S_N = X_1 + X_2 + \cdots + X_N$ によって定義された確率変数 S_N の期待値を計算しなさい。

こたえ 確率変数の期待値を条件を付けて計算する公式

$$E(X) = E(E(X \mid Y))$$

の応用問題。$N = n$ と固定して考えると、S_N は n 個の独立な確率変数の和になるので

$$E(S_N \mid N = k) = kE(X_1)$$

となる。したがって、

$$\begin{aligned}E(S_N) &= \sum_{k=0}^{n} E(S_N \mid N = k)P(N = k) = E(X_1)\sum_{k=0}^{n} kP(N = k) \\ &= E(X_1)E(N)\end{aligned}$$

練習問題 5.8 整数値を取る確率変数 X の確率関数が $P(X = k) = P(X = -k)(\forall k)$ という性質を持つものとします。$Y = X^2$ と定義します。このとき、$C(X, Y) = 0$ となることを証明しなさい。この二つの確率変数 X, Y は独立ですか？

例題 5.7 $X_1, X_2, \ldots$ は互いに独立でパラメータ p のベルヌイ分布に従う確率変数、また、N はそれらとは独立に、パラメータ n, r の 2 項分布に従う確率変数とします。このとき、$S_N = X_1 + X_2 + \cdots + X_N$ によって定義された確率変数 S_N がパラメータ n, pr の 2 項分布になることを示しなさい。

こたえ パラメータ p のベルヌイ分布は、表が出るコインが p のコイン A を投げたときに表が出れば 1、さもなければ 0 とするようなものである。パラメータ n, r の 2 項分布は n 個のベルヌイ分布の和、つまり、表が出る確率が r のコイン B を n 回投げたときに表の出る回数のようなものである。したがって、S_N はコイン B が表の時だけコイン A を投げるとしたとき、表の出る回数のようなものである。B が裏の場合でもコイン A を投げ、その結果は無視する、と考えると、S_N はただ単に、コイン A, B を投げて、両方とも表の出る回数、と言っても良い。ということは S_N はパラメータ n, pr の 2 項分布に従うことを意味する。

確率変数を使ってきちんと説明すると以下のようになる。$Y_1, Y_2, \ldots, Y_n$ は互いに独立でパラメータ r のベルヌイ分布に従う確率変数、$X_1, X_2, \ldots$ とも独立とする。$Z_n = X_1Y_1 + X_2Y_2 +$

$\cdots + X_nY_n$ と定義すると、Z_n と S_N の分布は等しい。なぜならば、$N = k$ のとき、S_k はパラメータ k, p の 2 項分布になり、N はパラメータ n, r の 2 項分布に従う。一方、$\sum_{i=1}^n Y_i = k$ のとき Z_n はパラメータ k, p の 2 項分布になり、$\sum_{i=1}^n Y_i$ はパラメータ n, r の 2 項分布に従うからである。一方 $X_1Y_1, X_2Y_2, \ldots$ は、$X_i = Y_i = 1$ の場合に限り 1、さもなければ 0 という、パラメータ pr のベルヌイ分布に従うから、Z_n はパラメータ n, pr の 2 項分布に従う。

あるいは確率母関数を使うと簡単。$N = k$ のとき、$S_N = X_1 + X_2 + \cdots + X_k$ はパラメータ k, p の 2 項分布に従うので、その確率母関数は $(1-p+pz)^k$ に等しい。

$$E(z^{S_N} \mid N = k) = (1-p+pz)^k$$

したがって、S_N の確率母関数を $G_s(z)$ とすると、

$$\begin{aligned} G_s(z) = E(E(z^{S_N} \mid N)) &= \sum_{i=0}^{n} (1-p+pz)^k \binom{n}{k} r^k (1-r)^{n-k} \\ &= ((1-p+pz)r + 1 - r)^n = (1 - pr + prz)^n \end{aligned}$$

これは S_N がパラメータ n, pr の 2 項分布に従うことを示す。

練習問題 5.9 $X_1, X_2, \ldots$ は互いに独立でパラメータ p のベルヌイ分布に従う確率変数、また、N はそれらとは独立に、パラメータ a のポワソン分布に従う確率変数とします。このとき、$S_N = X_1 + X_2 + \cdots + X_N$ によって定義された確率変数 S_N がパラメータ αp のポワソン分布になることを示しなさい。

練習問題 5.10 $X_1, X_2, \ldots$ は互いに独立でパラメータ p のベルヌイ分布に従う確率変数、$Y_1, Y_2, \ldots$ は互いに独立でパラメータ r のベルヌイ分布に従う確率変数、$X_1, X_2, \ldots$ とも独立、とします。このとき、$S_n = X_1Y_1 + X_2Y_2 + \cdots + X_nY_n$ によって確率変数 S_n を定義したとき、S_n の確率関数を求め、その期待値を計算しなさい。

練習問題 5.11 $X_1, X_2, \ldots$ は独立に平均 λ^{-1} の指数分布に従い、N はそれらとは独立にパラメータ p の幾何分布に従うものとします。このとき、$S_N = X_1 + X_2 + \cdots + X_N$ によって定義された確率変数 S_N の密度関数を計算しなさい。また、結果を見て、なぜそのようになるか、直感的に説明しなさい。ただし、平均 λ^{-1} の指数分布の密度関数は $x > 0$ で定義され、$\lambda e^{-\lambda x}$ で与えられます。

5.2　応用問題

例題 5.8 迷路に迷いこんで交差点に出たとしよう。左の道を行けば 15 分後に元に戻り、右へ行けば 30 分後に元の場所に戻る。真ん中の道を選ぶと 20 分後には外へ出られるとすると、この地点から外に出られるまでに、どれくらいかかるか計算しなさい。ただし、交差点ではそれまでに選んだ方向の記憶がなくなってしまい、毎回同じようにランダムに次に行く方向を選ぶものとする。

ヒント：最初に選ぶ道に関して条件を付けて考える。

こたえ　外に出るまでの時間を t とする。元の地点に戻るということは、さらに戻った時点から

t の時間がかかる、と考える。そうすると、

$$t = \frac{1}{3}(15+t) + \frac{1}{3}(30+t) + \frac{1}{3}(20)$$

という、t に関する方程式が得られる。これを解いて、$t = 65$

例題 5.9 保険会社の１週間の保険支払い請求件数は平均 m、分散 v の確率変数で、１件当たりの支払額は平均 q、分散 w の確率変数と表すことが出来る。請求件数と１件当たりの支払額は独立として、この保険会社の１週間の支払額の平均と分散を計算しなさい。

こたえ i 件目の請求の支払額を X_i、請求件数を N とすると、総支払額は $X_1+X_2+\cdots+X_N(\equiv S)$ と表される。その平均、分散は、N が定数ならば $NE(X), NV(X)$ のように計算できるが、それが確率変数のために、条件を付けて計算する必要がある。

$$\begin{aligned}
E(S) &= E(E(S \mid N)) = E(N)E(X_1) = mq \\
E(S^2) &= E(E(S^2 \mid N)) = E(N)E(X^2) + E(N(N-1))(E(X))^2 \\
V(S) &= E(S^2) - (E(S))^2 \\
&= E(N)E(X^2) + E(N(N-1))(E(X))^2 - (E(N))^2(E(X))^2 \\
&= E(N)V(X) + V(N)(E(X))^2 = mw + vq^2
\end{aligned}$$

例題 5.10 ある店の１日の買い物客はパラメータ 30 のポワソン分布に従い、一人の買い物金額は平均 5000 円、標準偏差 2000 円という。このとき、この店の一日の平均売り上げと、その分散を計算しなさい。ただし、パラメータ a のポワソン分布の確率関数は

$$f(i) = \frac{a^i}{i!}e^{-a}, \quad i = 0, 1, 2, ...$$

で与えられます。

こたえ 買い物客を N とすると、その平均と分散は

$$\begin{aligned}
E(N) &= \sum_{i=0}^{\infty} i \times \frac{a^i}{i!}e^{-a} = a = 30, \\
V(N) &= E(N(N-1)) + E(N) - (E(N))^2 = a^2 + a - a^2 = 30
\end{aligned}$$

したがって、一日の平均売り上げは、一人の買い物金額を X として、

$$E(N)E(X) = 150000$$

その分散は $V(X) = 4 \times 10^6$ だから

$$\begin{aligned}
E(N)V(X) + V(N)(E(X))^2 &= 30 \times 4 \times 10^6 + 30 \times 15^2 \times 10^6 \\
&= 6870 \times 10^6 = 82885^2
\end{aligned}$$

練習問題 5.12 もし明日雪が降れば雪下ろしのバイトで２万円、さもなければ普通のバイトで５千円が手に入る。天気予報の降雪確率は 30% になっている。明日の収入はいくらが見込めますか

ヒント：明日雪が降るという事象を A、明日の収入を表す確率変数を X としたとき、雪が降るという事象の下での収入（の期待値）を $E(X \mid A)$ と表すことにすると、明日の収入の期待値 $E(X)$ を計算する式を $E(X \mid A), E(X \mid A^c), P(A)$ を使って表しなさい。

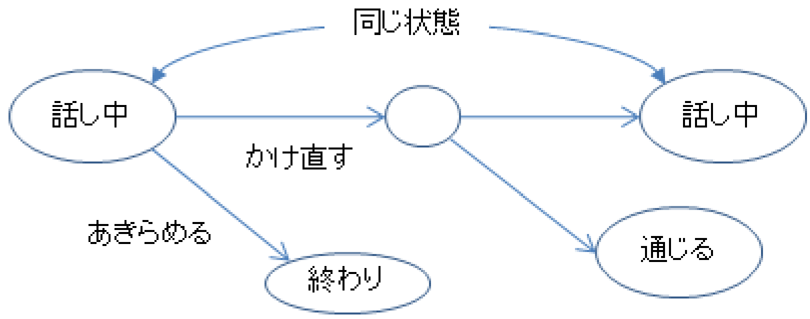
練習問題 5.13 もし明日雪が降れば雪下ろしのバイトで 2 万円、さもなければ普通のバイトで 5 千円が手に入る。しかし、あまりの大雪で交通麻痺になると動きが取れないので、バイトは休み。天気予報の降雪確率は 30% になっている。交通麻痺するような大雪はそのうちの 10% 程度、と見込まれている。明日の収入の期待値を計算しなさい。

練習問題 5.14 ある実験は確率 p で成功するか、$1-p$ で失敗に終わるかのいずれかである。成功する場合の所要時間の期待値は t_1、失敗する場合の期待値は t_2 である。実験を繰り返すとして、はじめて成功する迄の実験の所要時間の期待値を求めよ。

練習問題 5.15 携帯の 1 パケット送信する費用を c とします。一つのメールのパケット数はパラメータ p の幾何分布に従い、一日に送信するメールの数はパラメータ a のポワソン分布に従うものとすると、一ヶ月 30 日間のコストはどれくらいと見積もればよいですか。

練習問題 5.16 疑わしいメールが 1 時間当たり平均 12 のポワソン分布に従ってメールボックスに到着する。そのうち 10 分の 9 が本当の迷惑メールとしたとき、1 時間当たりに到着する迷惑メールの本数の期待値と分散を計算しなさい。

練習問題 5.17 電話をかけたとき相手が話し中だと通じるまで何回もかけ直す人がいる、そうしない人もいる。話し中の電話の後、ひとは平均何回かけ直すか、という問題。相手が話し中の確率はいつも r、相手が話し中のときかけ直す確率はいつも p とする（ちょっと不自然ですが）。このとき、相手と通じるか、あきらめてかけ直しをやめるまでに何回かけ直すか、その期待値（最初の電話はカウントしない）を計算しなさい。最初に掛けたら話し中だったという条件の下で、結局電話が通じる確率はどうなりますか。



練習問題 5.18 あるテレビ番組がレポーターを一人募集したら n 人の応募があった。順番に面接して、その場で合否を判定し、合格の人を採用するとしよう。面接は合格者が出るまで続き、合格者が出たら、残りは面接をしない。また、必ず誰かは合格にしなければいけないことにする。面接はワンパスで、一度不合格とした人は再度面接するわけにはいかないものとしよう。面接した人については相対的な順位が付けられるものとする。したがって、現在面接している人が、それまでに面接した不合格者の中の最良の候補者よりも良いか劣るかは判断できるものとする。さぁ、あなたが採用担当者だったらどういう決め方をしますか。

過去の最良の不合格者よりも良い人が現れたら合格、としたいのだが、あまり焦って決めても、面接しなかった候補者の中に最適な人が残っていたかもしれない。あるいは、後の候補者に期待して決定を先延ばしにすると、（どんなにひどくても）最後の人を合格にしなければいけなくなる。考えられる方策としては、最初の m 人は合格者を出さずに、その中の最良の人の水準

を覚えておき、$m+1$ 人目以降で最初にその水準を上回った候補者を合格にする、というやり方がもっともらしい。「最良の水準」はなるべく高めに設定したければ m を大きくする必要があるが、そうすると、最良の候補者を逃す可能性が大きくなる。m を小さくすると、最初の m 人がレベル以下だったりすると悲惨な結果になる。最適な m はどれくらいだろうか。と、ここまでは前置きで、次の問に答えなさい。

1 から n までの数が書かれたカードからランダムに m 枚を取り出し、その最小値を a とする。残りの $n-m$ 枚の中からランダムに 1 枚ずつ取り出して、a より小さいカードが取り出されるか、カードが無くなったら終わりにする。このとき、最後に取り出されたカードの数字が 1 である確率はどれくらいか，計算しなさい。

ヒント　これは「秘書選びの問題」という有名な問題の言い換え。中止したときに取り出したカードの枚数によって条件を付けて考える。最後のカードが 1 であるという事象を A、k 番目に取り出されたカードが選択されるという事象を B_k とする。全確率の公式から

$$P(A) = \sum_{k=m+1}^{n} P(A \cap B_k)$$

$A \cap B_k$ は n 枚のカードから k 枚を取り出して一列に並べる、という試行（$\binom{n}{k}k!$ 通り）の中で、条件に合う並べ方何通りあるか数え上げることによって、その確率を計算することが出来る（等可能性の原理）。k 番目は 1 で決まり、2 番目に大きい数は最初の m 枚に含まれていなければいけない、あとの $k-2$ 枚はどうでも良い、というように考える。

練習問題 5.19 (続き）$n \gg 1$ のとき、$P(A)$ を最大にするという意味で最良の m を求めなさい。計算する前に、m/n がどれくらいか、予想してみてください。

例題 5.11（負の 2 項分布）訪問販売で商品を売る。1 回の訪問先で売れる確率はいつも p とする。3 個売れるまで訪問し続けるものとしたとき、訪問先の軒数の期待値を計算しなさい。

ヒント：1 個目が売れるまでの期待値と、その直後から 2 個目が売れるまでの期待値の関係はどうなりますか。1 個目が売れるまでの訪問軒数が k ということは $k-1$ 軒では売れなかったということ。その確率は... というように考えればよい。

こたえ　1 個売れるまでの訪問軒数を X_1、その直後から 2 個目が売れるまでの訪問軒数を X_3、その直後から 3 個目が売れるまでの訪問軒数を X_3 とすると、$E(X_1+X_2+X_3)$ はいくつかという問題。例えば、「売れる」を「○」、「売れない」を「×」とすると、× × × ○ × ○ ×× ○みたいな系列が出来る。× 一つの確率が $1-p$、○が p なので、これはコイン投げの試行を記録したようなもの。したがって、X_1 も X_2, X_3 も確率規則はみんな同じにパラメータ p の幾何分布に従い、その期待値は $\frac{1}{p}$。結局

$$E(X_1+X_2+X_3) = E(X_1)+E(X_2)+E(X_3) = \frac{1}{p}+\frac{1}{p}+\frac{1}{p} = \frac{3}{p}$$

コメント　N を 3 個売れるまでに訪問する軒数とすると、$N=n$ という事象は、$n-1$ までに 2 個売れ、ちょうど n 軒目に売れる、という事象に等しい、ということを使って、N の確率関数を計算して、その期待値を計算する方法でも答えにたどり着けます。$n-1$ 軒のうち 2 軒で売れ、$n-3$ 軒で売れない、という事象の確率は、パラメータ $n-1, p$ の 2 項分布に従う確率変数がちょうど 2 に等しい確率 $\binom{n-1}{2}p^2(1-p)^{n-3}$ と同じです。ちょうど n 軒目に売れる確率は p、

したがって、

$$P(N=n) = \binom{n-1}{2} p^3 (1-p)^{n-3}$$

となりますから、その期待値を計算すればよいのです。このような確率関数は**負の 2 項分布**と呼ばれるものの一種です。

$$\sum_{n=0}^{\infty} n(n-1)(n-2)x^{n-3} = \frac{d^3}{dx^3}\sum_{n=0}^{\infty} x^n = \frac{d^3}{dx^3}\frac{1}{1-x} = \frac{6}{(1-x)^4}$$

という式を使えば、

$$G_N(z) = E(z^N) = \sum_{n=3}^{\infty} n\binom{n-1}{2} p^3 (1-p)^{n-3} = \frac{3}{p}$$

というように計算出来ます。しかし、「こたえ」のようにランダム試行の性質を利用した方が、間違いが少なく、計算量も少なくて済みます。がむしゃらに計算しないで、ちょっと考えてみるとこういう解き方が見つかるかもしれません。

例題 5.12 賃貸物件を 1 件ずつ下見をして、気に入ったものと契約したい、という場合、良い物件に巡り会うまでにどれくらい下見が必要かを調べるための確率モデルを以下のように考えました。X_n を n 件目に見たチンタイの家賃 (万円) とします。$X_1, X_2, ...$ は独立で、同じ累積分布関数 $F(x)$ を持つ確率変数とします。家賃が a（万円）以下だったら決めよう、と考えて探し回るとして、決まるまでに見る物件の数を N とします。このとき、以下の問いに答えなさい。**(1)** m を任意の自然数としたとき、$N > m$ という事象を言葉で説明し、それを確率変数 $X_1, X_2, ...$ を使って表しなさい。**(2)** $N > m$ という事象の確率を累積分布関数 $F(x)$ を使って表しなさい。**(3)** 確率変数 N の確率関数を求めなさい。それは何分布と呼ばれますか。**(4)** 累積分布関数 $F(x)$ が以下の式で与えられているとして、$a = 8$ としたときの N の平均と分散を計算しなさい。

$$F(x) = \begin{cases} 0, & x < 6 \\ \left(\frac{x-6}{4}\right)^2, & 6 \le x \le 10 \\ 1, & x > 10 \end{cases}$$

解説　部屋を借りたいとき、初めにいくつか見て大体の水準を知り、その中で最も良かったものを基準に（それが a)、それより良い（a より小さい）ものがあったら契約する、そういっているあいだに残りが無くなって、やむを得ず，最後に見たものを借りることになる (カードが無くなる)、ということもありうる、という経験を確率の問題に直したもの。たとえば、「m をいくつにしたら最良の物件を借りる確率が最大になるか」というような最適化問題が考えられる。

こたえ　**(1)** $N > m$ ということは、m 回連続して契約不成立（家賃が a より高い）ということと同じ、それは、$X_1 > a, X_2 > a, ..., X_m > a$

(2) $X_1, X_2, ..., X_m$ は互いに独立なので、

$$\begin{aligned} P(N > m) &= P(X_1 > a, X_2 > a, ..., X_m > a) \\ &= P(X_1 > a)P(X_2 > a)\cdots P(X_m > a) = (1 - F(a))^m \end{aligned}$$

(3) $P(N = m) = P(N > m-1) - P(N > m)$ という関係を使えば

$$P(N = m) = F(a)(1 - F(a))^{m-1} \equiv p(1-p)^{m-1}$$

これはパラメータ $p=F(a)$ の幾何分布に他ならない。

(4) パラメータ p の幾何分布の平均は $\frac{1}{p}$、分散は $\frac{1-p}{p^2}$。$p=F(8)=\frac{1}{4}$ なので、

$$E(N)=4, V(N)=12$$

練習問題 5.20 自作のゲームソフトを売り込むために、買い取ってくれそうな企業を回るとしましょう。訪問企業がオファーする金額は予想が付かないけれど、過去の経験から金額の確率分布が分かっているものとします。X_k を k 番目に訪問した企業がオファーする金額を表す確率変数とすると、$X_1, X_2, \ldots$ は独立で、同じ累積分布関数 $F(x)$ に従うものとします。自分であらかじめ、最低販売価格 $a(>0)$ を決めて、最初に a 以上の金額をオファーした企業の言い値で売却するものとします。一つの企業を訪問するのには C のコストがかかるものとすると、もし、C が高ければあまり贅沢を言うことは出来ませんが、C が安ければ、高い a を設定することが可能でしょう。この a を決めるのが問題です。以下の問いに答えなさい。**(1)** 契約が決まるまでに訪問する企業の数を N としたとき、$N>4$ という事象を確率変数 $X_1, X_2, \ldots$ によって表しなさい。**(2)** $P(N>n)$ を $X_1, X_2, \ldots$ の累積分布関数 $F(x)$ で表しなさい。**(3)** $P(N=n)=P(N>n-1)-P(N>n)$ という関係が成り立つことを説明しなさい。**(4)** $P(N=n)$ を $F(x)$ を使って表しなさい。**(5)** N の期待値を計算しなさい。**(6)** X_N という確率変数は何を表すのか説明しなさい。**(7)** $E(X_N)=E(X_1 \mid X_1 \geq a)$ であることを説明しなさい。

以下では、解を簡単に求めるために、オファー金額を連続値として考えます。$X_1, X_2, \ldots$ が平均 m の指数分布に従い、その累積分布関数が $F(x)=1-e^{-x/m}(x \geq 0$、それ以外は $0)$ で与えられるものとします。**(8)** $E(X_N)$ を計算しなさい。**(9)** $E(X_N - CN)$ を求めなさい。**(10)** $E(X_N - CN)$ が最大となる a を決めるのが問題です。$E(X_N - CN)$ を a の関数と考えて、微分することにより、$E(X_N - CN)$ を最大とする a の値 a^* を C, m を使って表しなさい。C あるいは m が大きくなったとき、a^* はどうなるか調べなさい。

例題 5.13（世論調査の分析）最近、某新聞社は「我が社の n 人調査によると、内閣支持率は 40% で、前回調査の 42% に比べると、やや減少傾向にある」と報じた。この報道について、確率の知識を用いて検討してみよう。**(1)** 確率を知らない人にとっては「n 人を調査する内閣支持率調査は、歪んだコインを n 回繰り返し投げて、表の相対度数を調べるようなもの」と言われても分からないかもしれません。もう少しかみ砕いた説明を考えてください。**(2)** 支持するという回答の数を X としたとき、X の確率関数を求めなさい。ただし、（未知の）支持率は p とします。**(3)** 支持するという回答の相対度数を $Z(\equiv X/n)$ としたとき、Z の確率関数を求めなさい。**(4)** Z の期待値を計算しなさい。**(5)** Z の標準偏差を計算しなさい。**(6)** 調査する数 n を十分に取らないと、調査結果は信用できない、ということを確率の言葉を使って説明しなさい。

こたえ **(1)** 一人一人の回答は予測が出来ないので、ランダムに選んだ人は確率 p で内閣を支持し、$1-p$ でそうではない、と回答すると「仮定する（せざるを得ない）」。そうすると、回答者が誰であれ、1 番目の回答も 2 番目の回答も、3 番目の回答も、確率 p で「支持」、確率 $1-p$ で「不支持」という結果になる。ということは、表の出る確率が p のコインを n 回投げて、表、裏の記録を並べたものと、調査結果の「支持」「不支持」の回答を並べたものとではそのランダムさが同じようなもの、と言っても良いことになる。

(2) X は n 回のコイン投げで表の出る回数と同じなので、パラメータ n, p の 2 項分布に従う。支持率 p は未知であっても、「パラメータ n, p の 2 項分布」は定義できるので、このように答えても良い。

(3) Z の取りうる値は、$\frac{0}{n}, \frac{1}{n}, \frac{2}{n}, ..., \frac{n}{n}$ のいずれか。

$$P\left(Z = \frac{k}{n}\right) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

(4)

$$E(Z) = E\left(\frac{X}{n}\right) = p$$

(5)

$$V(Z) = V\left(\frac{X}{n}\right) = \frac{1}{n^2}V(X) = \frac{p(1-p)}{n} \Rightarrow S(Z) = \sqrt{\frac{p(1-p)}{n}}$$

(6) 2 項分布の確率関数は np を中心に正規分布のように np から離れるほど、急速に小さくなっていく。Z の場合は 2 項分布の 0 から n までを $[0,1]$ に縮小したもので、中心は p になり、n を大きくするにつれて、小さくなり方が激しく、ほとんど p の周辺だけの値を取るようになる。その集中度合いを区間の長さで表現したものが、統計の信頼区間である。95% 信頼区間とは確率関数の合計が 0.95 になるような範囲、と考えればよい。95% 信頼区間の幅は $4\sigma(= 4S(Z))$ なので、n が大きくなると小さくなることが式の上からも確かめられる。

練習問題 5.21（ポートフォリオの収益）ポートフォリオ（分散投資）とは、例えば投資の対象となる資産が A, B 二つあったとすると、1 単位の投資額のうち、$100r\%$ を A に、$100(1-r)\%$ を B に投資するというように、分散して投資するという投資計画のことです。投資プラン A に 1 単位投資するときのもうけを X、投資プラン B に 1 単位投資するときのもうけを Y とします。これらはいずれも確率変数、それぞれの期待値と標準偏差は $\mu_X, \mu_Y; \sigma_X, \sigma_Y$、$X, Y$ の相関係数は ρ とします。A, B 二つの投資プランに投資するというポートフォリオの収益を Z で表すと、

$$Z = rX + (1-r)Y$$

と書くことが出来るでしょう。このとき、以下の問に答えなさい **(1)** Z の期待値を計算し、それを最大にする r を決定しなさい。**(2)** Z の分散を計算しなさい。**(3)** $\rho = 1$ の場合に、Z の分散を最小にする r を決定しなさい。**(4)** $\rho = 0$ の場合に、Z の分散を最小にする r を決定しなさい。**(5)** $\rho = -1$ の場合に、Z の分散を最小にする r を決定しなさい。**(6)** ρ が一般の場合、$g(r) = V(Z)$ と置いて、$g(r)$ を最小にする r を r^* とするとき、$g(r^*)$ を計算しなさい。