確率変数の期待値について

□□□□□ □□□□□ □□□□□ □□□□□ □□□□□ □□□□□　2009 年 6 月 10 日版

目次

1　賭けと期待値

偶然をともなうゲームを考えよう。このゲームに参加するには参加料 c 円を支払う必要がある。ゲームの結果、ある金額をもらうものとし、それを X と表す。必ずしも正の値とは限らず、負の値（このときには負けてさらに支払わねばならない）かもしれない。簡単のために、この値は r 通りの $x_1, x_2, \cdots, x_r$ という値をとるものとする。このゲームに参加したほうがよいだろうか、あるいは参加しないほうが賢明なのであろうか。もしこのようなゲームをかなり多くの回数繰り返すことができるとすれば、とり得る値の頻度（度数）から、参加の決定により得られる利得を判断できる。n 回ゲームをすれば、nc 円支払うが、$X_1 + X_2 + \cdots + X_n$ を手にする。各々の値は互いに独立な確率変数である。その密度関数は $P(X = x_i) = f_X(x_i), i = 1, 2, \cdots, r$ とする。いま $N_n(x)$ で値 x をとったゲームの回数、つまり $X_1, X_2, \cdots, X_n$ のうち、値 x となっているものの個数 $N_n(x) = \sum_i \mathbf{1}_{\{X_i = x\}}$ とし、これを総回数で割れば、その比率（割合）が求まる。$X_1 + X_2 + \cdots + X_n = \sum_{i=1}^r x_i N_n(x_i)$, すなわち

$$\frac{X_1 + X_2 + \cdots + X_n}{n} = \sum_{i=1}^r x_i \frac{N_n(x_i)}{n} \tag{1.1}$$

大数の法則[*1]から、$N_n(x)/n \to f_X(x) = P(X = x), (n \to \infty)$ であるから、

X の値	x_1	x_2	$\cdots$	x_r	計
比率 $N_n(x)/n$	$N_n(x_1)/n$	$N_n(x_2)/n$	$\cdots$	$N_n(x_r)/n$	1
離散密度 $f_X(x) = P(X = x)$	$f_X(x_1)$	$f_X(x_2)$	$\cdots$	$f_X(x_r)$	1

[*1] 大数の弱法則とは、分布が近づくことであり、大数の強法則とは、確実に（除外される可能性はゼロの確率）近づくという概念である。証明は平均と分散に関する確率の関係式を述べるチェビシェフの不等式により弱法則についての確率収束（あるいは分布収束ともよばれる）は簡単に証明されるが、一方強法則のいう概収束は多少の確率に関する予備知識が必要となる。

極限として $\mu = \sum x_i f_X(x_i)$ に近づく。

左辺はゲーム 1 回あたりの参加費用 c であるから、こうして $\mu > c$ ならば、利益が得られ、逆の不等式ならば、損失をこうむる。等号の $\mu = c$ ならば、利益、損失もない状態である。このような閾値が確率変数 X の期待値

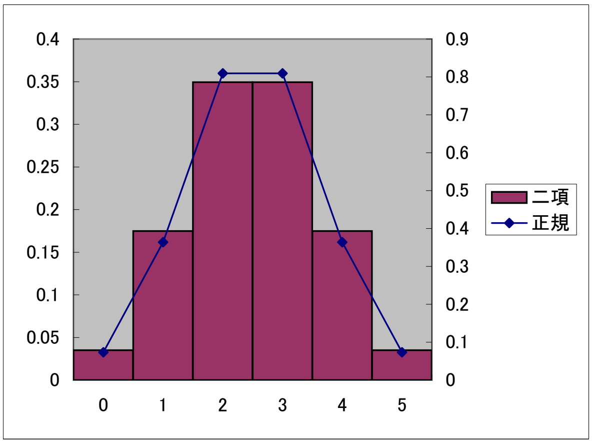
$$\mu = \sum_i x_i P(X = x_i) = \sum_i x_i f_X(x_i) \tag{1.2}$$

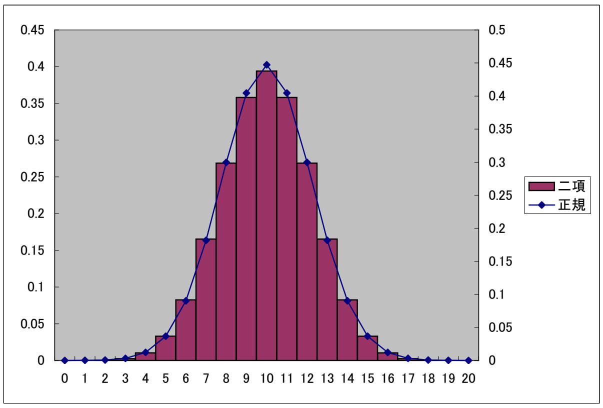
である。

以上は離散型確率変数であるが、連続型確率変数で密度関数を $f_X(x) = \dfrac{d}{dx}F_X(x) = \dfrac{d}{dx}P(X \le x)$ とすると、$f_X(x)\delta x \sim P(x < X < x + \delta x)$ とみなせるから

$$\mu = \int_0^\infty x f_X(x) dx \tag{1.3}$$

という形になる。つまり、分布の期待値（平均値）は「とり得る値×確率の総計」で与えられる。式 (1.1) にみられるようデータ値の算術平均と分布の期待値は、極めて密接な関係をもち、平均値とか期待値などの用語も注意して用いなければならない。





2　一般的な定義からの計算法

期待値は確率空間 $(\Omega, \mathcal{F}, \mathcal{P})$ での抽象的な積分で定義される。

まず非負値のばあいには, ある整数 $n > 1$ をもちいて、区間を重複のないように分割し $\left(\dfrac{k}{2^n}, \dfrac{k+1}{2^n}\right], k = 0, 1, 2, \cdots, n2^n - 1$ に対して、つぎの極限値が存在するとき、記号 EX で

$$EX = \lim_n \sum_k \frac{k}{2^n} P\left(\frac{k}{2^n} < X \le \frac{k+1}{2^n}\right) = \lim_n \sum_k \frac{k}{2^n}\left\{F_X\left(\frac{k+1}{2^n}\right) - F_X\left(\frac{k}{2^n}\right)\right\}$$

とする。非負の値であるから、極限値が有限であるときに、「確率変数 X は期待値 EX をもつ」という。負の値ももつときには、$X^+ = max\{X, 0\}, X^- = min\{-X, 0\}$ とすると、両方とも非負値で $X = X^+ - X^-$ とかけるから、共に有限 $EX^+ < \infty, EX^- < \infty$ であるときに、$EX = EX^+ - EX^-$ で定める。

確率変数 X の関数 $h(X)$ は確率変数で X の分布から計算される。逆関数により、$h(X)$ の範囲と X の範囲を対応させて

$$\sum_k \frac{k}{2^n} P\left(\frac{k}{2^n} < h(X) \le \frac{k+1}{2^n}\right) = \sum_k h\left(\frac{k}{2^n}\right) P\left(\frac{k}{2^n} < X \le \frac{k+1}{2^n}\right)$$

しかしより具体的な偶然事象に対して期待値を計算するときには積分の変数変換をおこなって確率変数の確率分布で積分するのが簡単である。

確率変数 X の分布を $F_X(x) = P(X \leq x)$ とすると、これから定められる測度 $F_X(dx)$ について、任意の可測関数（やや不正確にいえば、階段関数や連続関数の極限形）$y = h(x)$ にたいして, 確率変数 Y を $Y(\omega) = h(X(\omega))$ で定められるから、その Y の期待値は

$$E[Y] = E[h(X)] = \int_\Omega h(X(\omega))P(d\omega) = \int_\mathcal{R} h(x)F_X(dx) \tag{2.4}$$

と計算される。

もし F_X が確率密度関数 f_X をもつ場合、あるいは離散型密度 p_X の場合には、$F_X(dx) = f_X(x)dx$, あるいは $F_X(dx) = p_X(x)$, $A = \{X$ のとり得る値 $\}$ とすれば、これらの両方の場合を同時に定義している。

$$E[h(X)] = \begin{cases} \displaystyle\int_{x\in A} h(x)f_X(x)dx & \text{離散型} \\ \displaystyle\sum_{i\in A} h(x_i)p_X(x_i) & \text{連続型} \end{cases} \tag{2.5}$$

で計算できるようになる。変数変換も $h(X)$ の分布を求めることよりも、X の分布と関数 $h(X)$ のとる値から、上の式から求めればよい。

さらに X, Y の同時（結合）分布 $F_{X,Y}$ が与えられたとき、関数 $z = h(x, y)$ に対しての期待値もつぎのように計算できる。$Z(\omega) = h(X(\omega), Y(\omega))$ であるから、集合 B を組 (x, y) あるいは (i, j) のとりえる値の集合として、

$$E[h(X,Y)] = \begin{cases} \displaystyle\int\int_{(x,y)\in B} h(x,y)f_{X,Y}(x,y)dxdy & \text{離散型} \\ \displaystyle\sum\sum_{(i,j)\in B} h(x_i,y_j)p_{X,Y}(x_i,y_j) & \text{連続型} \end{cases} \tag{2.6}$$

関数 $h(x, y)$ の例としては和 $h(x, y) = x + y$ や積 $h(x, y) = xy$ などがあるが、和の場合には積分の線形性と周辺分布の性質から、独立でなくてもそれぞれの和で計算できるが、積のほうは独立であれば、積で与えられる。

定義 2.1 共分散

確率変数の共分散（きょうぶんさん,*covariance*）とは、2 組の対応する確率変数での、各変数の平均からの偏差の積に対するの期待値（平均値）と定める。記号 $cov(X, Y)$ などと表す。

$$\begin{aligned} cov(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] \\ &= \begin{cases} \displaystyle\int_R\int_R (x-\mu_X)(y-\mu_Y)f_{X,Y}(x,y)dxdy \\ \displaystyle\sum_i\sum_j (x_i-\mu_X)(y_j-\mu_Y)f_{X,Y}(x_i,y_j) \end{cases} \end{aligned} \tag{2.7}$$

定理 2.1 *(1)* 線形変換した平均と分散：$Y = aX + b$,*(a,b* は定数*)* のとき、X の平均 μ_X, 分散 σ_X^2 から Y に対する平均 μ_Y, 分散 σ_Y^2 が次の式で与えられる。

$$\begin{aligned} &(i) \quad \mu_Y = a\mu_X + b \\ &(ii) \quad \sigma_Y^2 = a^2\sigma_X^2 \qquad \text{（定数}b\text{にはよらない）} \end{aligned} \tag{2.8}$$

(2) 2 つの確率変数の平均と分散 $Z = aX + bY$,*(a,b* は定数*)* のとき、X, Y の平均 μ_X, μ_Y, 分散 σ_X^2, σ_Y^2、共分散 $var(X, Y)$ から Z に対する平均 μ_Z, 分散 σ_Z^2 が次の式で与えられる。

$$\begin{aligned} &(i) \quad \mu_Z = a\mu_X + b\mu_Y \\ &(ii) \quad \sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab \cdot cov(X, Y) \end{aligned} \tag{2.9}$$

もし X, Y が独立ならば、$cov(X, Y) = 0$ であり、$\sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2$. さらにもし Y が変動していない、つまり確率変数ではなく、一定の値しかとらないならば、分散はゼロ、$var(Y) = \sigma_Y^2 = 0$ であるから、上の結果 (2.9) は (2.8) へ帰着される。

また線形変換 $Z = aX + b$ において $a = \dfrac{1}{\sigma_X}, b = -\dfrac{\mu_X}{\sigma_X}$ とおけば、つまり $Z = \dfrac{X - \mu_X}{\sigma_X}$ とすれば

$$X: \begin{array}{l} \text{平均}: \mu_X \\ \text{分散}: \sigma_X^2 \end{array} \xrightarrow{Z = \frac{X - \mu_X}{\sigma_X}} Z: \begin{array}{l} \text{平均}: \mu_Z = 0 \\ \text{分散}: \sigma_Z^2 = 1 \end{array} \tag{2.10}$$

このような新たに平均を 0, 分散を 1 にするような変換を確率変数 X の基準化あるいは標準化とよぶ。

3　より一般な期待値の性質

期待値の性質

期待値は積和あるいは積分によって定義されるので、和あるいは積分のもつ性質をもっている。

- 線形性：$E(X + Y) = E(X) + E(Y)$ (2 つの確率変数は独立でなくても成立する)
- 単調性: もし $X \leq Y$ という条件[*2]ならば、$E(X) \leq E(Y)$

これから、多くの関数について期待値の計算が定められている。平均、分散、モーメント：

1. 平均：$\mu = E(X)$
2. 分散：$\sigma^2 = var(X) = E(X - \mu)^2 = E(X^2) - \mu^2$
3. r 次の原点モーメント：$E(X^r)$
4. 平均 μ のまわりの r 次モーメント：$E[(X - \mu)^r]$
5. 階乗モーメント：$E\left[X(X-1)\cdots(X-r+1)\right]$

例題 3.1 2 項分布 $X \sim Binom(n, p)$ の期待値：

とりえる値の集合 $\{0, 1, 2, \cdots, n\}$, 確率密度 $\dbinom{n}{k} p^k (1-p)^{n-k}, k \in \{0, 1, 2, \cdots, n\}$

$$\textit{(i)}\ \mu = np \quad \textit{(ii)}\ \sigma^2 = np(1 - p)$$

[*2] つまり、$P(X \leq Y) = 1$ あるいは, $X(\omega) \leq Y(\omega) a.e.\omega$, ほとんどすべての ω で大小関係をもつとき。

この計算には、

方法 (1) n 個のベルヌーイ分布 $Binom(1,p)$ の和を用いると、$e_i = 1; p, 0; 1-p$ から $X = e_1 + e_2 + \cdots + e_n$ をつくる。期待値の線形性や独立の場合の結果をつかい、もちいて $\mu = EX = p + p + \cdots + p = np$ あるいは $\sigma^2 = E(X^2) - \mu^2 = \sum_i E(e_i^2) + 2\sum_{i<j} E(e_i e_j) - \mu^2 = np + 2\binom{n}{2} p^2 - (np)^2 = np(1-p)$。

方法 (2) 階乗モーメントを利用して $k\binom{n}{k} p^k (1-p)^{n-k} = np\binom{n-1}{k-1} p^{k-1}(1-p)^{n-1-(k-1)}, k = 1, 2, \cdots, n$ $k(k-1)\binom{n}{k} p^k (1-p)^{n-k} = n(n-1)p^2 \binom{n-2}{k-2} p^{k-2}(1-p)^{n-2-(k-2)}, k = 2, \cdots, n$ であるから $E[X] = np, E[X(X-1)] = n(n-1)p^2$ を得る。したがって $\mu = np, \sigma^2 = E(X^2) - \mu^2 = E[X(X-1)] + EX - \mu^2 = n(n-1)p^2 + np - (np)^2 = np(1-p)$ となる。ここでは和の計算に独立性が満たされていない場合にも期待値の和の計算が適用可能である。たとえば非復元抽出（標本の取り出しで、取り出したものをもとに戻さないから、独立ではない）のときにも計算できるから重要である。

定理 3.1 非負整数値の確率変数に対して、その期待値はつぎの形でも求められる。

$$E[X] = \sum_k P(X \geq k)$$

なぜなら

$$\begin{array}{rclllllllll}
E[X] & = & \sum_k kP(X=k) \\
 & = & 1 \times P(X=1) & + & 2 \times P(X=2) & + & 3 \times P(X=3) & + & 4 \times P(X=4) & + \cdots \\
 & = & P(X=1) & + & P(X=2) & + & P(X=3) & + & P(X=4) & + \cdots \\
 & & & + & P(X=2) & + & P(X=3) & + & P(X=4) & + \cdots \\
 & & & & & + & P(X=3) & + & P(X=4) & + \cdots \\
 & & & & & & & + & P(X=4) & + \cdots \\
 & = & P(X \geq 1) & + & P(X \geq 2) & + & P(X \geq 3) & + & P(X \geq 4) & + \cdots \\
 & = & \sum_k P(X \geq k)
\end{array}$$

例題 3.2 幾何分布の期待値：

幾何分布の期待値計算を上の方法で求めてみる。幾何分布とは、コイン投げにおいて表の出る確率 p のとき、投げ続けてはじめて表が出るまでに必要であった回数を表す。一般にパラメータ p の幾何分布とは

$$P(X=k) = (1-p)^{k-1}p, \qquad k = 1, 2, \cdots$$

で定めると、

$$\begin{array}{ll}
P(X \geq k) & = p(1-p)^{k-1}\{1 + (1-p) + (1-p)^2 + \cdots\} \\
 & = (1-p)^{k-1} \\
\sum_k P(X \geq k) & = \sum_k (1-p)^{k-1} = \dfrac{1}{p}
\end{array}$$

この幾何分布の平均は $E(X) = \sum_k k(1-p)^{k-1}p = 1/p$ となる。

このような方法は、連続型分布についても考えられる。すなわち

定理 3.2 区間 $[0, \infty]$ における連続型確率変数 X、密度関数 $f_X(x), 0 < x\infty$ その分布関数を $F_X(x) = P(X \leq x)$, すなわち $P(X \geq x) = 1 - F_X(x)$ に対して、その期待値はつぎの形でも求められる。

$$E[X] = \int_0^\infty \{1 - F_X(x)\} dx$$

証明は部分積分の用いれば得られる。

注意として、期待値が有限な値 $E[X] < \infty$ であることと $\lim_{t\to\infty} \int_t^{\infty} \{1 - F_X(x)\}dx = 0$ が成り立つこととは同値である。

例題 3.3 指数分布例題としては指数分布（しすうぶんぷ、英: *exponential distribution*）が典型的である。

$$f_X(x) = \lambda e^{-\lambda x}, x > 0 \qquad F_X(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}, \quad x > 0$$

指数分布も重要な分布である。ポアソン過程とも密接な関連をもつ。単位時間ごとに生起する事象の数がパラメータ λ のポアソン分布に従うとき、そのような事象の生起間隔はパラメータ λ の指数分布に従う。離散時間パラメータでの、２項過程と幾何分布の関係に対応する。指数分布は、幾何分布と同様に無記憶性 (メモリーレス) とよばれる性質をもつ。これは、確率変数 X が

$$P(X > s + t | X > s) = P(X > t), \quad s, t > 0$$

なる等式を満たすことをいう。すなわち、時刻 s までに事象が生起しなかったという情報が与えられたとき、その事象がさらに t 時間の間生起しない条件付き確率は、（時刻 s まで事象が生起しなかったという情報が完全に忘れ去られ、改めてその時点から観測を始めて）t 時間の間事象が生起しない確率に一致するという意味である。信頼性理論の分野ではハザード関数（故障率関数）の定義では指数分布は一定の値であり、故障が起こりやすい、起こりにくいという性質に関して変化が一定という性質に対応する。上述した確率分布関数の定義より、指数分布に従う確率変数がこの性質を満たすことは容易に示される。逆に、この性質を満たす連続確率分布が指数分布のみであることも証明されている。

平均、分散の計算は

$$E[X] = \int_0^{\infty} (1 - F_X(t))dt = \int_0^{\infty} e^{-\lambda x} dt = 1/\lambda$$

さらに分散[*3] の計算は $E[X^2] = \displaystyle\int_0^{\infty} 2x(1 - F_X(t))dt = 2\int_0^{\infty} xe^{-\lambda x} dt$ より

$$var(X) = E(X^2) - (EX)^2 = 2/\lambda^2 - (1/\lambda)^2 = 1/\lambda^2$$

となる。

定義 3.1 ガンマ分布

パラメータ (r, λ) のガンマ分布の密度関数は

$$f_X(x) = \lambda^k x^{k-1} \frac{e^{-\lambda x}}{\Gamma(k)}, \quad x > 0$$

とおく。

これはパラメータ λ の指数分布を r 個加えた和であるから、平均、分散は

$$E[X] = \sum_{i=1}^{r} E[X_i] = r/\lambda, \qquad var[X] = \sum_{i=1}^{r} var[X_i] = r/\lambda^2$$

ここで密度関数の係数には、積分値が 1 となるよう、つぎの関数をもちいている。

[*3] 連続型のモーメントは $E(X^r) = \int_0^{\infty} x^r f_X(x)dx = r\int_0^{\infty} x^{r-1}(1 - F_X(x))dx$ となる。

定義 3.2 ガンマ関数

$\Gamma(r) = \int_0^\infty x^{r-1}e^{-x}dx$, $r > 0$ でガンマ関数とよばれ、

$$\Gamma(r) = (r-1)\Gamma(r-1), \quad \Gamma(1) = 1, \quad \Gamma(1/2) = \sqrt{\pi}$$

が成り立つ。

また待ち行列理論ではよく用いられ、再生性（ガンマ分布の和もガンマ分布）をもつ。

例題 3.4 超幾何分布の期待値：有限母集団からの抽出

超幾何分布とは、つぼに入っている 2 種類のボールからの取り出しで、「非復元抽出」した場合の結果である。一度取り出したボールはもとに戻さないから、つぎの取り出しは、以前の結果によって、つぼ内のボール構成が異なってくる。このとき取り出したボールの数についての分布である。もし「復元抽出」であれば、ボールの構成は同じとなるから、これは繰り返しが行われる 2 項分布の場合に他ならない。

いまつぼには合計 $M = R + B$ 個のボールが入っているが、赤ボールが R 個、黒ボールが B 個あるものとする。これからランダムに n 個のボールを取り出すが、一度取り出したボールはもとに戻さないものと仮定する。確率変数 X を取り出した n 個のボールのうち、赤ボールの個数とする。$M = R + B$ であるから

$$(1) \quad P(X = x) = \frac{\binom{R}{x}\binom{B}{n-x}}{\binom{M}{n}}, \qquad x = 0, 1, 2, \cdots n$$

$$(2) \quad EX = n\frac{R}{M} \qquad (3) \quad var(X) = n\left(\frac{R}{M}\right)\left(\frac{B}{M}\right)\frac{M-n}{M-1}$$

このとき、次の注意を述べておく。

(1) 二項分布でのパラメータ p は, 超幾何分布の比率 $\frac{R}{M}$ に対応している。また $M = R + B$ は母集団の総数を表すから、$M(R, B) \to \infty$ とすれば、復元と非復元の違いはなくなる。平均はどちらでも同じであるが、分散は $\frac{M-n}{M-1} \to 1 (M \to \infty)$ で同じ値に近づく。これを母集団が有限であること生じているから、有限補正項とよばれる。

(2) 平均の計算は分布の期待値から直接計算できたが、条件つき期待を使い、確率変数の和の期待値からも計算できる。n 個中の赤ボールの総数は $X = X_1 + X_2 + \cdots + X_n$、ただし $X_i = 1$, i 番目のボールが赤のとき、$X_i = 0$, i 番目のボールが黒のときとおく。赤を成功、黒を失敗とみなしてもよい。ただしこの 2 種類の結果はそれ以前の取り出しに依存しているから、「独立ではない」。しかし条件つき確率は簡単に求められる。なぜならまず最初は

$$P(X_1 = 1) = R/(B + R), \qquad P(X_1 = 0) = B/(B + R),$$

であり、2 回目は $P(X_2 = 1|X_1 = 1) = (R-1)/(M-1)$, $P(X_2 = 0|X_1 = 1) = B/(M-1)$, $P(X_2 = 1|X_1 = 0) = R/(M-1)$, $P(X_2 = 0|X_1 = 0) = (B-1)/(M-1)$ より

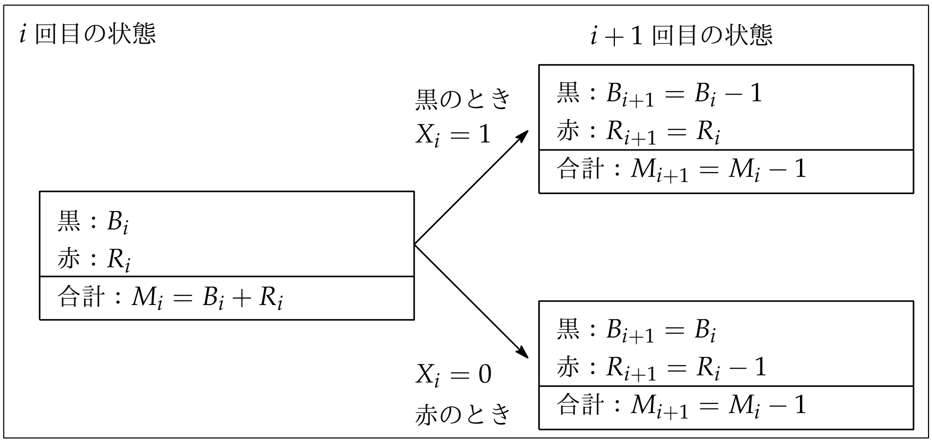
$$\begin{aligned} P(X_2 = 1) &= P(X_1 = 1)P(X_2 = 1|X_1 = 1) + P(X_1 = 0)P(X_2 = 1|X_1 = 0) \\ &= R/M * (R-1)/(M-1) + B/M * R/(M-1) \\ &= R/M \\ P(X_2 = 0) &= B/M \end{aligned}$$

以下同様に $i = 2, 3, \cdots, n$ で

$$P(X_{i+1} = 1) = R/(R + B), \qquad P(X_{i+1} = 0) = B/(R + B),$$

であるから、$EX = EX_1 + EX_2 + \cdots + EX_n = n\dfrac{R}{R+B} = n\dfrac{R}{M}$ を得る。もし分散を計算する場合、独立でないときには、2 次モーメントの計算に相関係数を求める必要がでてくるのでここでは省略する。

状態推移の図



イェンセン (Jensen) の不等式: 凸関数 $\varphi(x) : 0 \leq \forall t \leq 1, \varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y)$ にたいして

$$\varphi(E[X]) \leq E[\varphi(X)] \tag{3.11}$$

凸関数の例として 2 次関数 $\varphi(x) = x^2$ を考える。$(E[X])^2 \leq E[X^2]$ となるから、$E[X^2] < \infty$ ならば、$E[X] < \infty$ である。つまり 2 次モーメントが有限すれば、1 次モーメントが有限な値となる。分散が存在するならば、平均は必ず存在する。高次のモーメントが存在すれば、低次のモーメントが存在する。しかし逆は必ずしも成り立たない。また密度関数となっているが、平均が存在しないこともある。その例として、コーシー分布 $f_X(x) = \dfrac{1}{1+x^2}, -\infty < x < \infty$ が知られている。

チェビシェフ (Chebyshev) の不等式: 確率変数 X が有限な平均 μ と分散 σ^2 をもつとき、任意の正の数 c にたいして

$$P(|X - \mu| > c) \leq \frac{\sigma^2}{c^2} \tag{3.12}$$

例題 3.5 2 項分布 $Binom(n, p)$ の確率評価；

$n = 30, p = 0.5$ とおき、計算する。

$\mu = np = 15, \sigma^2 = np(1-p) = 7.5$ であり、$P(|X - \mu| > c)$ と $\dfrac{\sigma^2}{c^2}$ の比較をおこなう。

k	11	12	13	14	15	16	17	18	19
$P(X = k)$	.0509	.0806	.1115	.1354	.1445	...	...	...	...
$np - k$	-4	-3	-2	-1	0	1	2	3	4

(i) $c = 2$: $\dfrac{\sigma^2}{c^2} = \dfrac{npq}{2^2} = 1.875$

$$P(|X - 15| > 2) = P(X \neq 13, 14, 15, 16, 17) = 1 - (.1115 + \cdots + .1115) = 1 - .6384 = .3616$$

(ii) $c = 3$: $\dfrac{\sigma^2}{c^2} = \dfrac{npq}{3^2} = .8333$

$$P(|X - 15| > 3) = P(X \neq 12, 13, \cdots, 17, 18) = 1 - (.0806 + \cdots + .0806) = 1 - .7995 = .2004$$

(iii) $c = 4$: $\dfrac{\sigma^2}{c^2} = \dfrac{npq}{4^2} = .4688$

$$P(|X-15| > 4) = P(X \neq 11, 13, \cdots, 17, 19) = 1 - (.0509 + \cdots + .0509) = 1 - .9013 = .1987$$

この結果を見ると、確率とその評価値とは差が大きいと感じるかも知れないが、チェビシェフの不等式の長所は確率変数に制限が少ないことである。

独立な確率変数の積の期待値: 二つの可積分な確率変数 X と Y が独立の場合は

$$E(XY) = E(X)E(Y) \tag{3.13}$$

が成立する。

ただし逆は不成立であることに注意。簡単な例として $f_{X,Y}(x,y) = 1/4, (x,y) \in \{(-1,0),(0,1),(0,-1),(1,0)\}$ では $E(XY) \neq E(X)E(Y)$。

定理 3.3 一般に n 個の独立な確率変数の積の期待値では

$$E(X_1X_2\cdots X_n) = E(X_1)E(X_2)\cdots E(X_n) \tag{3.14}$$

さらに独立に加えてすべて同一の分布をもつばあい、それを X とすれば、べき乗となる。

$$E(X_1X_2\cdots X_n) = (EX)^n \tag{3.15}$$

独立で同一分布する確率変数は *iid r.v.* (independently identically distributed random variable) と略記されることが多い。

4 くじ引きの期待値

例題 4.1 次のようなゲームを考える。*100* 円支払えば、*6* 面サイコロ *1* 個を *1* 回振ることができる。サイコロの目に応じて、次の金額を貰える。

さいころの目 $X = x$:	*1*	*2*	*3*	*4*	*5*	*6*
確率 $P(X = x)$:	*1/6*	*1/6*	*1/6*	*1/6*	*1/6*	*1/6*
金額 $h(x)$:	*20* 円	*50* 円	*100* 円	*100* 円	*150* 円	*150* 円

このとき、貰える金額の期待値を求めると、$E[h(X)] = \sum_x h(x)P(X = x) = 20 \times (1/6) + 50 \times (1/6) + \cdots + 150 \times (1/6) = 95$ である。得られる金額の期待値 *95* 円が参加費 *100* 円を下回ることから、このゲームは参加者が得をする可能性もあるものの平均的には損をするということが分かる。特に回数を多く行なうほど、大数の法則から行なった回数ごとに *5* 円の損をした状態に限りなく近づく。このようなゲームは多数回繰り返すと、胴元が得をする仕掛けで損をしていく。賭けをしようとするときには、１回、２回だけでサイコロの目が５，６が出ることを祈ることが賢明である。極言にいえば、不利なゲームに対しては、一か八かの大勝負 *(boldly)* にでるべきで、不利なことをチミチミ *(timid)* 繰り返され続けていては、ますます苦境に落ちていくだけである。

例題 4.2 サンクト・ペテルブルクのパラドックス（逆理）

1738 年、サンクトペテルブルクに住んでいたダニエル・ベルヌーイ[*4]が、学術雑誌『ペテルブルグ帝国アカデミー論集』の論文「リスクの測定に関する新しい理論」で発表した。その目的は、期待値による古典的な「公平さ」が現実には必ずしも適用できないことを示し、「効用」（ラテン語: *emolumentum*）についての新しい理論を展開することであった。

偏りのないコインを表が出るまで投げ続け、表がでたときに、賞金をもらえるゲームがあるとする。もらえる賞金は *1* 回目に表がでたら *1* 円（通貨単位は本質ではなく、どんなものでもいい。ベルヌーイの原論文ではダカット（英語）金貨で、現在の約 *500* 日本円相当だった)、*1* 回目は裏で *2* 回目に表がでれば倍の *2* 円、*2* 回目まで裏で *3* 回目が表ならまたその倍の *4* 円というふうに倍々でふえる。つまり表が出るまでに投げた回数を *n* とすると、*2(n-1)* 円もらえる賭けであるとしよう。はたして期待値を効用として考えると、このような賭けに参加することは有利なこととみなせるであろうか。具体的数字では *10* 回目に表が出れば *512* 円、*20* 回目に出れば *52* 万 *4288* 円、*30* 回目なら *5* 億 *3687* 万 *0912* 円である。ここで、このゲームには賭け金が必要であるとしたら、賭け金の金額が何円までなら払っても損ではないと言えるだろうか。期待値の考えから、参加することの有利不利の判断基準が定められるはずであった。多くの場合、この種の問題では賞金の期待値を算出して、賭け金がそれ以下であれば良いとする。ところが、この問題で実際に賞金の期待値を計算してみると、その数値は無限大に発散してしまうのである。すなわち期待値を W とすると、

$$W = \sum_i (2^{k-1}\frac{1}{2^k}) = \frac{1}{2} + \frac{1}{2} + \cdots = \infty$$

となる。したがって、期待値によって判断するならば、賭け金がいくら大金であっても参加すべきであるということになる。ところが実際には、このゲームでは *1/2* の確率で *1* 円、*1/4* の確率で *2* 円、*1/1024* の確率で *512* 円の賞金が得られるに過ぎない。したがって、そんなに得であるはずがないことは直観的に分かる。ゆえにこれはパラドックスとされる。

期待値は多くの分野でも利用されるが、このようにすべての評価基準として通用する概念ではない。これに注意しながら、さまざまな応用に対処する必要がある。

5　条件付き期待値

定義と応用を述べよう

(1) $E[X|Y] = E[X|Y = y] \quad \text{on} \quad \{Y = y\}$
(2) $E[X] = E\big(E[X|Y]\big)$

つまり、離散型であれば、$E\big(E[X|Y]\big) = \sum_y E[X|Y = y]P(Y = y) = E[X]$ を意味する。

$$E[X|Y = y] = \begin{cases} \displaystyle\int_{\mathcal{R}} x f_{X|Y}(x|y)dx = \int_{\mathcal{R}} x\frac{f_{X,Y}(x,y)}{f_Y(y)}dx & \text{(連続型)} \\ \displaystyle\sum_i x_i p_{X|Y}(x_i|y) = \sum_i x_i\frac{p_{X,Y}(x_i,y)}{p_Y(y)} & \text{(離散型)} \end{cases}$$

[*4] Daniel Bernoulli, 1700 年 2 月 9 日 - 1782 年 3 月 17 日, オランダ生まれ、父、伯父に数学、自然科学者一家。スイス、ロシアの大学で教鞭をとった数学者、物理学者。流体力学（流れに沿って成り立つエネルギー保存の法則、ベルヌーイの定理)、振動理論などの自然科学の功績が大きいが、このような経済理論でのリスク評価でも有名

定理 5.1 (ワルド (Wald) の等式) 確率変数のランダムな個数の和 : $\sum_{i=1}^{N} X_i = X_1 + X_2 + \cdots + X_N$ を考える。確率変数 $X_i, i = 1, 2, \cdots$ は独立同一分布 (*i.i.d.*) で、その分布は X と同じとする。また N は正の整数値をとる確率変数で、$X_i, i = 1, 2, \cdots$ とは独立であるとする。すなわち和の個数が変動するような場合である。ただし重要な仮定として、$EN < \infty$ のもとで考える。このとき期待値について次が成り立つ。

$$E(X_1 + X_2 + \cdots + X_N) = E(N)E(X) \tag{5.16}$$

これをワルド[*5]の等式とよぶ。$EN < \infty$ が満たされないと反例が知られている。

N と X_i の独立性から、任意の n で $E[X_1 + X_2 + \cdots + X_N | N = n] = E[X_1 + X_2 + \cdots + X_n | N = n] = E[X_1 + X_2 + \cdots + X_n] = nE[X]$ よって条件つき期待値に対して期待値をとれば $E(E[X_1 + X_2 + \cdots + X_N | N = n]) = E(X_1 + X_2 + \cdots + X_N)$ より、結論を得る。

応用として、幾何分布の平均を求めてみる。幾何分布とは、コイン投げにおいて表の出る確率 p のとき、投げ続けてはじめて表が出るまでに必要であった回数を表す。たとえば、試行が「裏裏裏表」となれば、4 回目ではじめて表が出たから、「X＝4」とする。この確率は $(1-p)^3 p$ である。一般にパラメータ p の幾何分布とは

$$P(X = k) = (1-p)^{k-1} p, \qquad k = 1, 2, \cdots$$

で定める。この分布の平均（期待値）は $E(X) = \sum_k k(1-p)^{k-1} p = 1/p$ となる。直感的には、もし表の出る確率が $1/5$ であれば、平均して 5 回繰り返せば表が出ると考えられる。この式を別の方法で求めてみよう。いま確率変数 Y をはじめに表が出たら、１、そう出なければ０とおく。

$$Y = \begin{cases} 1 & \text{はじめに表} \quad \text{確率 } p \\ 0 & \text{はじめに裏} \quad \text{確率 } 1-p \end{cases}$$

条件つき期待値をもちいて

$$E[X] = E[X|Y=1]P(Y=1) + E[X|Y=0]P(Y=0) = pE[X|Y=1] + (1-p)E[X|Y=0]$$

ここで $E[X|Y=1] = 1, E[X|Y=0] = 1 + E[X]$ を代入して、$E[X]$ について解けば、結果 $E[X] = 1/p$ が示される。

6 正規分布の平均、分散、モーメント

正規分布のモーメントを計算するには微積分の知識をより多く用いる必要に迫られる。

正規分布が統計学上特別な地位をもつ理由は、中心極限定理があるためである。中心極限定理は、「独立な同一の分布に従う確率変数の算術平均（確率変数の合計を変数の数で割ったもの）の分布は、もとの確率変数に分散が存在するならば、もとの分布の形状に関係なく、離散型であっても連続型であってもモーメントの条件さえあれば、極限分布として正規分布に収束する。」という。大標本の平均値の統計には、正規分布が仮定されることが非常に多い。”自然界”の事象には、正規分布に従うものがあることが知られている。しかしそれは必ずしも”多数派”というわけではない。19 世紀ではさながら「正規分布万能主義」といったものがまかり通っていたが、20 世紀以降そういった考え方に修正が見られた。今日においては社会現象、生物集団の現象等々、種別から言えば、正規分布に従うものはむしろ少数派であることが確認されている。例えば、フラクタ

[*5] エイブラハム・ワルド（Abraham Wald, 1902 年 10 月 31 日 - 1950 年 12 月 13 日インド出身の数学者。飛行機事故で、若くして死亡）Sequential Analysis（1947 年）Statistical Decision Functions（1950 年）が有名。

ルな性質をもつ場合は正規分布よりも、パレート分布になることが多い。また人間は自然界の事象とはちがって自分の意思をもっているため、たとえば、子供たちの成績などは決して正規分布にはならないといわれる。何らかの事象について法則性を捜したり理論を構築しようとしたりする際、その確率分布がまだ分かっていない場合にはそれが正規分布であると仮定して推論することは珍しくないが、誤った結論にたどりついてしまう可能性がある。本当にその事象が正規分布であるかどうかは実際のデータから確認するしかない。十分というわけではないが、最低限、データの尖度と歪度を調べるべきである。正規分布 は、その平均を μ, 分散を σ^2 とするとき、次の形の確率密度関数 $X \sim N(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big), \quad -\infty < x < \infty$$

をもつ。この正規分布を $N(\mu, \sigma^2)$ と表す（N は「正規分布」を表す英語 ”Normal Distribution” の頭文字に由来する）。特に $\mu = 0, \sigma^2 = 1$ のとき、この分布は標準正規分布（または基準正規分布）とよばれる。つまり標準正規分布 $Z \sim N(0, 1)$ は（複素数積分では複素数を z で用いるが、ここでは実数であることに注意）

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{z^2}{2}\Big), \quad -\infty < z < \infty$$

なる確率密度関数をもつ確率分布として与えられる。ガンマ関数の $\Gamma(1/2) = \sqrt{2}$ が実数の積分計算で $\int_{-\infty}^{\infty} f_Z(z)dz = 1$ と同値。正規分布の確率密度関数をグラフ化した正規分布曲線は左右対称なつりがね状の曲線であり、鐘の形に似ているからベル・カーブともよばれる。両端が急激にゼロに近づき、漸近する。直線 $x = \mu$ を軸に左右対称であり、x-軸が漸近線である。なお、曲線は σ の値が大きいほど扁平になる。正規分布は、いろいろな分布でのパラメータの極限を考えると、その極限分布となる。たとえば統計の歴史的進展から得られた正規分布の発見（中心極限定理の特別な場合）では、大きな n に対する二項分布ともみなすことができる。

２項分布の平均と分散から

$$E(Z) = 0, \quad var(Z) = E(Z^2) - (EZ)^2 = 1$$

定理 6.1 一般の正規分布 $X \sim N(\mu, \sigma^2)$ については

$$\begin{array}{llll} (1) & E(X) = \mu & (1') & E(X-\mu) = 0 \\ (2) & E(X^2) = \mu^2 + \sigma^2 & (2') & var(X) = E(X-\mu)^2 = \sigma^2 \\ (3) & E(X^3) = 3\mu\sigma^2 + \mu^3 & (3') & E(X-\mu)^3 = 0 \\ (4) & E(X^4) = 3\sigma^4 + 6\mu^2\sigma^2 + \mu^4 & (4') & E(X-\mu)^4 = 3\sigma^4 \end{array} \tag{6.17}$$

が成り立つ。

これらの結果を得る方法は、すなわち確率論の歴史といっても過言ではないかも知れない。いままでにも表れてきた確率分布を調べ、解析するためには、分布の特徴をいくつかの量で表す、モーメント（積率）が用いられる。たとえば、1 次の積率は分布の中心がどこにあるかを示し、2 次のそれは分布の中心付近において分布がどれほど集中しているかを示している。モーメントは理論的にその期待値を与える式より求められるが、一般には必ずしも容易ではない。そこで、モーメントを生成する関数として、積率母関数 (moment generating function, mgf) とよばれる次式を定義する。

定義 6.1 モーメント母関数: 確率変数 X のｍｇｆ $\phi_X(t)$ とは

$$\phi_X(t) = E[e^{tX}]$$

形式的に直接計算すると、指数関数のテイラー展開から, r 次のモーメント $\mu_r = E[X^r], r = 1,2,\cdots$ をもちいて、

$$\begin{aligned}\phi_X(t) = E[e^{tX}] \quad &= E\left\{1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!}]+\right\}\\ &= 1 + t(EX) + \frac{t^2}{2!}E(X^2) + \frac{t^3}{3!}E(X^3)+\\ &= 1 + t\mu_1 + \frac{t^2}{2!}\mu_2 + \frac{t^3}{3!}\mu_3+\end{aligned}$$

となる。この係数を取り出すために、$\phi_X(t)$ を微分して $t = 0$ とおくと、

$$\phi_X(0) = 1, \quad \left.\frac{d\phi_X}{dt}\right|_{t=0} = \mu_1, \quad \left.\frac{d^2\phi_X}{dt^2}\right|_{t=0} = \mu_2, \quad \left.\frac{d^3\phi_X}{dt^3}\right|_{t=0} = \mu_3, \cdots \tag{6.18}$$

したがってモーメント $\mu_r, r = 1,2,\cdots$ が計算される。しかし一般にはモーメント、期待値の計算のため、$\phi_X(t)$ が簡単ではないか、もしこのｍｇｆが求まるならば極めて微分で計算できることになる。

正規分布の場合はどうであろうか？　正規分布に対する積率母関数を求めてみよう。

$X \sim N(\mu, \sigma^2)$ の密度関数は $f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $-\infty < x < \infty$ より、モーメント母関数の定義式に代入し、変換 $u = \frac{x-\mu}{\sigma}$ とすることで、$x = \mu + \sigma u$, $dx = \sigma du$ から

$$\begin{aligned}\phi_X(t) \quad &= \int_{-\infty}^{\infty} e^{tx}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\\ &= e^{\mu t + \frac{\sigma^2}{2}t^2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{(u-\sigma t)^2}{2}}\\ &= \exp\left\{\mu t + \frac{\sigma^2}{2}t^2\right\}\end{aligned}$$

最後の等号は、積分式については正規分布型の密度関数であるので、積分すれば、1 となるからである。X のかわりに $X - \mu$ の mgf は

$$\phi_{X-\mu}(t) = E[e^{t(X-\mu)}] = e^{-\mu t}E[e^{tX}] = e^{-\mu t}\phi_X(t)$$

より

$$\begin{aligned}e^{\frac{\sigma^2t^2}{2}} \quad &= 1 + \frac{\sigma^2t^2}{2} + \frac{\sigma^4t^4}{2!\cdot 2^2} + \frac{\sigma^6t^6}{3!\cdot 2^3} + \cdots\\ &= 1 + \frac{\sigma^2}{2!}t^2 + \frac{3\sigma^4}{4!}t^4 + \frac{3\cdot 5\sigma^6}{6!}t^6 + \cdots\end{aligned}$$

さらに係数比較することで $X - \mu$ のモーメントは

$$\mu_1 = 0, \mu_2 = \sigma^2, \mu_3 = 0, \mu_4 = 3\sigma^4, \cdots \tag{6.19}$$

が得られるから、(6.17) の結果となる。(1′) $E(X-\mu) = 0$ 　(2′) $var(X) = E(X-\mu)^2 = \sigma^2$ 　(3′) $E(X-\mu)^3 = 0$ 　(4′) $E(X-\mu)^4 = 3\sigma^4$

したがって、これらを用いると、同値な関係式も導かれる。(1) $E(X) = \mu$ 　(2) $E(X^2) = \mu^2 + \sigma^2$ 　(3) $E(X^3) = 3\mu\sigma^2 + \mu^3$ 　(4) $E(X^4) = 3\sigma^4 + 6\mu^2\sigma^2 + \mu^4$ となる。

ここで定義したモーメント母関数 (6.1) は逆にモーメントから分布を定めることに用いられる。より一般には複素数関数の $E(e^{itX}), i = \sqrt{-1}$ が収束するから、数学の厳密性という理論の観点からはよく解析には導入される。

7　２変数に対する期待値

2 変数の確率変数 X, Y について和 $E(X+Y)$, 積 $E(XY)$ を計算する。確率変数が独立あれば、比較的簡単である。独立でないときには和に対して $E(X+Y) = E(X) + E(Y)$ となるから、同時（結合）分布から、それぞれの周辺分布を求めることで、期待値が求められる。一方、積に対しては $E(X \cdot Y) \neq E(X) \cdot E(Y)$ であるから、それぞれの周辺分布では不足してしまう。

共分散と分散の関係

数学におけるコーシー＝シュワルツの不等式（コーシーシュワルツのふとうしき、英語: Cauchy-Schwarz inequality）、シュワルツの不等式、シュヴァルツの不等式あるいはコーシー＝ブニャコフスキー＝シュワルツの不等式 (Cauchy-Bunyakovski-Schwarz inequality) などとよばれ、簡単な不等式関係では

$$(ac+bd)^2 \leqq (a^2+b^2)(c^2+d^2)$$

（右辺）−（左辺）＝ $(a^2+b^2)(c^2+d^2) - (ac+bd)^2 = (ad-bc)^2$ 等号は $ad = bc$ のとき。少し一般化して n 個の数列では

$$(a_1b_1 + a_2b_2 + \cdots + a_nb_n)^2 \leqq (a_1^2 + a_2^2 + \cdots + a_n^2)(b_1^2 + b_2^2 + \cdots + b_n^2)$$

等号が成り立つのは, ある定数 k があって, $a_i = kb_i, i = 1, 2, \cdots, n$ のとき。もし積分形であれば、

$$\left(\int_0^1 f(x)g(x)dx\right)^2 \leqq \left(\int_0^1 f(x)^2dx\right) \cdot \left(\int_0^1 g(x)^2dx\right)$$

数列に対する不等式はオーギュスタン＝ルイ・コーシーによって 1821 年に、積分系での不等式はまずヴィクトール・ブニャコフスキーによって 1859 年に発見された後ヘルマン・アマンドゥス・シュワルツによって 1888 年に再発見されたといわれている（Wikipedia)。ベクトル空間の内積とノルムの評価関係にも表れる。

２つの確率変数 X, Y では

$$\Big(E(XY)\Big)^2 \leqq E(X^2)E(Y^2) \tag{7.20}$$

平行移動（線形変換）した変数に適用すれば $\Big(E(X-\mu_X)(Y-\mu_Y)\Big)^2 \leqq E[(X-\mu_X)^2]E[(Y-\mu_Y)^2]$。したがって平均と分散、共分散で表現すれば

$$cov(X,Y)^2 \leqq \sigma_X^2 \cdot \sigma_Y^2 \tag{7.21}$$

定義 7.1　相関係数:

相関係数（そうかんけいすう、*correlation coefficient*）ρ とは、

$$\rho = \frac{E(X-\mu_X)(Y-\mu_Y)}{\sqrt{E[(X-\mu_X)^2]E[(Y-\mu_Y)^2]}} = \frac{cov(X,Y)}{\sigma_X \cdot \sigma_Y} = cov\Big(\frac{X-\mu_X}{\sigma_X}, \frac{Y-\mu_Y}{\sigma_Y}\Big) \tag{7.22}$$

で定める。いいかえると、２つの標準化された変数の共分散に他ならない。

ここで ρ に添え数として、確率変数を明示していない理由は、標準化により定めるから、各変量を線形変換しても不変であることによる。

定理 7.1 確率変数 X, Y の線形変換 $V = aX + b, W = cY + d, (a > 0, c > 0)$ とするとき

$$\rho = \frac{E(X-\mu_X)(Y-\mu_Y)}{\sqrt{E[(X-\mu_X)^2]E[(Y-\mu_Y)^2]}} = \frac{cov(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{E(V-\mu_V)(W-\mu_W)}{\sqrt{E[(V-\mu_V)^2]E[(W-\mu_W)^2]}} = \frac{cov(V,W)}{\sigma_V \cdot \sigma_W}$$

条件の中に $a > 0, c > 0$ としているが、これは各々の変量に増減を変えないようにしてしているだけである。つまり、X が増加すれば、V も増加するとした。２変量の変化傾向を変えない、という線形変換である。

2 つの確率変数の間の相関（類似性の度合い）を示す統計学的指標として用いられる。単位は標準化によって、無単位であり、不等式（）により、その数値は -1 から 1 の間の実数値をとることがわかる。この値が 1 に近いときは 2 つの確率変数には「正の相関がある」といい、-1 に近ければ「負の相関がある」という。0 に近いときはもとの確率変数の「相関は弱い」という。等号が成り立つ場合, 1 もしくは -1 となる場合は 2 つの確率変数は線形従属の関係にある。確率変数に対して相関係数が必要になる場合はそう頻度が多くはないが、この 2 変量正規分布では山型の尾根部分の方向を示すこととなる。

記述統計学では、組のデータ値から求めた相関係数、すなわちピアソンの積率相関係数（Pearson product-moment correlation coefficient）をさす。よく用いられるが、確率変数（確率分布）の場合と記述統計でのデータ値とは区別して考えなければならない。２変量正規分布として分布を仮定する（パラメトリック）方法であるが、他にこのような仮定を置かないノンパラメトリックな方法として、スピアマンの順位相関係数、ケンドールの順位相関係数なども記述統計では一般に用いられる。相関係数は、あくまでも確率変数の線形関係を計測しているに過ぎない。また、確率変数間の因果関係を説明するものでもない。相関係数は順序尺度であり間隔尺度ではないので、例えば「相関係数が 0.2 と 0.4 であることから、後者は前者より 2 倍の相関がある」などと言うことはできない。しばしば、相関があるという表現が、あたかも因果関係を示しているかのように誤解あるいは誤用される。

2 変量正規分布 (bivariate normal distribution) の密度関数を式で表現すると

$$\begin{aligned} f(x,y) = f_{X,Y}(x,y) &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \\ &\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\frac{x-\mu_X}{\sigma_X}\cdot\frac{y-\mu_Y}{\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\} \end{aligned} \tag{7.23}$$

2 つの確率変数を対象とした 2 次元同時（結合）分布 (joint distribution) です．ある 1 つの変量の変動だけをみた 1 次元の分布は，周辺分布 (marginal distribution) といい、座標軸による断面図であるが、これも 1 次正規分布である。中心となる頂上は、各周辺分布の期待値（平均）μ_X, μ_Y で決まり、分散 σ_X^2, σ_Y^2 はそれぞれの軸への広がり具合でさらに 2 変量間の相関係数 ρ がパラメータとなって，両軸への広がり、つまり直線 $y = x$ への 2 変量の集中度など分布の形状が定まります．２変量正規分布のイメージを表す図を示しておく。