

テーマ：期待値と分散、共分散

・期待値

ある確率分布があって、その累積分布関数  $F$  が与えられているとき、その期待値とは、

$$E(x) = \int_{-\infty}^{\infty} x dF$$

で与えられる。

もちろん、関数  $x$  は有界ではないので、期待値はすべての確率に対して定義できるものではない。例として、経済学で有名な「サントペテルブルグの逆説」と呼ばれる賭けを紹介しよう。この賭けは、以下のステップで行われる。

1. 参加者は  $X$  円の参加費を払う。
2. 参加者はコインを投げて、表か裏かを見る。裏が出たらそこで終わり、表が出たら続けて投げることができる。
3. 投げた回数を  $n$  としたとき、賭けの報酬は  $10 \times 2^{n-1}$  円である。

この賭けの分布関数は離散型であり、 $10 \times 2^{n-1}$  に  $2^{-n}$  の確率を割り振る形になっている。期待値を取ると、

$$10 \times \frac{1}{2} + 20 \times \frac{1}{4} + 40 \times \frac{1}{8} + \dots = 5 + 5 + 5 + \dots = +\infty$$

となって、この賭けの期待値は  $+\infty$  に発散している。

では、この賭けはいくら払っても期待値的には得なのだから参加するか？ といえば、 $X$  がたとえば10万円だったら参加する人間はほぼいないだろう。この例は「人はリスクのある出来事に対して、期待値だけを見て参加することはない」ということを表しているが、同時に「我々は簡単に期待値が存在しない分布を作ってしまう」ということも意味している。

しかし、そうはいつでも実用的な多くの分布は有限の期待値を持っている。そして、それらを用いて分析する手法も議論されている。さしあたり、前回紹介した幾何分布について期待値を計算すると、

$$\begin{aligned} E(x) &= p \times 1 + p(1-p) \times 2 + p(1-p)^2 \times 3 + \dots \\ &= p[1 + 2(1-p) + 3(1-p)^2 + \dots] \end{aligned}$$

という値になる。ダランベールの収束判定定理を用いれば、この括弧の中身が収束することが示せるので、この値は有限である。一方で、

$$(1-p)E(x) = p[(1-p) + 2(1-p)^2 + 3(1-p)^3 + \dots]$$

を上から引き算すると、

$$pE(x) = p[1 + (1-p) + (1-p)^2 + (1-p)^3 + \dots]$$

となる。括弧内に出てくるのは等比級数なので、等比級数の和の公式

$$1 + a + a^2 + \dots = \frac{1}{1-a}$$

を  $a = 1 - p$  に適用すれば、

$$pE(x) = p \times \frac{1}{1 - 1 + p} = 1$$

となり、よって

$$E(x) = \frac{1}{p}$$

が得られる。

ポワソン分布の場合はどうであろうか？ やってみると、

$$\begin{aligned} E(x) &= 0 \times e^{-\lambda} + 1 \times \frac{\lambda e^{-\lambda}}{1!} + 2 \times \frac{\lambda^2 e^{-\lambda}}{2!} + 3 \times \frac{\lambda^3 e^{-\lambda}}{3!} + \dots \\ &= \lambda e^{-\lambda} \left[ \frac{1}{0!} + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \dots \right] \end{aligned}$$

を得る。括弧内は有名な  $e^x$  のマクローリン展開

$$e^x = \frac{1}{0!} + \frac{1}{1!}x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \dots$$

の  $x$  に  $\lambda$  を代入した値なので、結局

$$E(x) = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

となる。

このように、多くの分布は期待値を持つ。この期待値を持つ意味については、後ほど解説する。

・正規分布

正規分布と呼ばれる確率分布の累積分布関数は、

$$F(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$$

で与えられる。

これが正しく分布関数であることを示すためには、

$$\lim_{x \rightarrow \infty} F(x) = 1$$

でなければならない。つまり、

$$\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sqrt{2\pi\sigma^2}$$

である必要がある。これを示すのには、数学で有名な次の公式（証明はしない）

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

を利用する。上の積分の式を変数変換  $y = \frac{x-\mu}{\sqrt{2\sigma^2}}$  で計算すると、 $dy = \frac{1}{\sqrt{2\sigma^2}} dx$  だから、

$$\begin{aligned} & \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \int_{-\infty}^{\infty} \sqrt{2\sigma^2} e^{-y^2} dy = \sqrt{2\pi\sigma^2} \end{aligned}$$

となって、確かに正しいことが確かめられる。

正規分布については、 $E(x) = \mu, E(x^2) = \sigma^2 + \mu^2$  となることが知られているが、これの証明は省略する。

#### ・大数の法則と中心極限定理

いま、 $X_1, \dots, X_N$  が同じ分布に従う確率変数だとする。このとき、

$$\frac{1}{N}[X_1 + \dots + X_N]$$

は、統計学で標本平均と呼ばれるものである。

我々は  $X_i$  が従う確率自体を知ることができないので、その性質を知る方法が欲しい。特に、たとえば  $X_i$  の分布の期待値などは、上の標本平均の値と似た値であってくれないと、困る。これは、たとえばポワソン分布などを考えれば、 $\lambda$  は元々上の標本平均から類推された値だったのだから、期待値  $\lambda$  と標本平均が無関係であってくれと困るのであ

る。しかし、標本平均と期待値は、少なくとも数学的には、違う概念である。これらに関係はあるのか？

ここで、用語をいくつか付け足しておこう。ある確率変数  $Y_N$  が  $Y$  に確率収束するとは、どんな小さな数  $\varepsilon > 0$  を取って来ても、

$$\lim_{N \rightarrow \infty} P(|Y_N - Y| > \varepsilon) = 0$$

が成り立つことを言う。一方、 $Y_N$  が  $Y$  に概収束するとは、

$$P\left(\lim_{N \rightarrow \infty} Y_N = Y\right) = 1$$

であることを言う。このふたつは違う概念であるが、一般に概収束の方が、確率収束よりも強い。

なんらかの条件下で標本平均  $Y_N$  が  $X_i$  の期待値  $E(x)$  に確率収束する、という定理を、弱い大数の法則 (weak law of large numbers) と呼ぶ。一方で、やはりなんらかの条件下で標本平均  $Y_N$  が  $E(x)$  に概収束する、という定理を、強い大数の法則 (strong law of large numbers) と呼ぶ。

一般に、弱い大数の法則を示すのには、 $E(X_i X_j) = E(X_i)E(X_j)$  が常に成り立っていればよい。この条件が成り立つことを、 $X_i$  と  $X_j$  は無相関であると言う。一方、強い大数の法則を示すためにはこれだけではダメで、どんな関数  $f, g$  に対しても、 $E(f(X_i)g(X_j)) = E(f(X_i))E(g(X_j))$  が成り立たなければならない。この条件が成り立つことを、 $X_i$  と  $X_j$  は独立であると言う。

ともかく、大数の法則は確率論の基礎のひとつである。これは、標本の数が十分に多ければ、標本平均は真の期待値の推定値として使ってよいことを意味する。

一方、標本平均と真の期待値の差、いわゆる推定誤差はどう評価されるだろうか。これについては、 $X_1, \dots, X_N$  が独立であるとして、

$$\frac{1}{N}[X_1 + \dots + X_N] - E(x)$$

を計算すると、 $N \rightarrow \infty$  とすれば、これは強い大数の法則から 0 に概収束する。より興味深い結果は、これに  $\sqrt{N}$  を掛けた値、つまり

$$Z_N = \frac{[X_1 + \dots + X_N] - NE(x)}{\sqrt{N}}$$

である。十分に  $N$  が大きいとき、 $Z_N$  の分布関数は、 $\mu = 0, \sigma = 1$  の正規分布の分布関数に近づいていくことがわかっている。この結果を中心極限定理 (central limitation theorem) と呼ぶ。

中心極限定理のよいところは、 $X_i$  の分布関数がなんであろうと成り立つことである。これによって、統計学者は本来の分布が知らなくてもだいたいの計算ができることになる。すべての状態を調査することが難しい社会科学において、分布の特性がわからなくても分析できるこれらの定理は極めて重要な意味を持つ。

・分散と共分散

期待値  $E(x)$  が実数で与えられるような、分布関数  $F$  で表される確率の分散 (variance)  $V(x)$  とは、 $E(x^2) - (E(x))^2$  のことである。積分で書くと、

$$\int_{-\infty}^{\infty} x^2 dF - \left( \int_{-\infty}^{\infty} x dF \right)^2$$

である。

期待値が存在しない分布があるように、分散が定義できない分布も存在する。

分散の定義で注意すべきことは、これは

$$E((x - E(x))^2)$$

と書いても同値だということである。実際、スティルチェス積分の定義から

$$\int_a^b c dF = c(F(b) - F(a))$$

となるので、 $\lim_{x \rightarrow -\infty} F(x) = 0$  と  $\lim_{x \rightarrow \infty} F(x) = 1$  からただちに

$$\int_{-\infty}^{\infty} c dF = c$$

であることがわかる。したがって  $c = (E(x))^2$  に適用して整理してみると、

$$\begin{aligned} E((x - E(x))^2) &= E(x^2 - 2E(x)x + (E(x))^2) \\ &= E(x^2) - E(2E(x)x) + E((E(x))^2) \\ &= E(x^2) - 2E(x)E(x) + (E(x))^2 \\ &= E(x^2) - (E(x))^2 = V(x) \end{aligned}$$

となって、たしかに正しいことがわかる。

分散に続いて共分散 (covariance)  $Cov(x, y)$  の説明をしたいのだが、これについては多少補足を要する。いま、 $F(x, y)$  という二変数関数を考え、これが以下の条件を満たすとしよう。

1)  $F$  は単調非減少。

2)  $\lim_{x \rightarrow -\infty, y \rightarrow -\infty} F(x, y) = 0$  で、 $\lim_{x \rightarrow \infty, y \rightarrow \infty} F(x, y) = 1$  である。

確率論的には、 $F$  は  $X$  と  $Y$  というふたつの確率変数の分布を表している。つまり、 $F(x, y)$  は  $X$  が  $x$  以下で、 $Y$  が  $y$  以下である確率を表している。当然ながら、たとえばここから  $X$  だけの分布を得るためには、

$$F_X(x) = \lim_{y \rightarrow \infty} F(x, y)$$

を分布関数と見ればよい。同様に、

$$F_Y(y) = \lim_{x \rightarrow \infty} F(x, y)$$

を分布関数として見れば、これが  $Y$  の分布関数である。

さて、 $F$  についてスティルチェス重積分をする。この定義は、重積分の定義自体をやっていないのでわかりにくい、必要あれば Fubini の定理という定理を自習して欲しい。さしあたり、ふんわりと

$$E(f(x, y)) = E_F(f(x, y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dF$$

という概念が定義できると考えよう（もちろん、リーマン和を正方形で議論して、厳密な議論をすることはそれほど難しくないので、チャレンジ精神のある学生はやってみて欲しい）。このとき、

$$\text{Cov}(x, y) = E(xy) - E(x)E(y)$$

として定義される。

分散の時と同様に、

$$\text{Cov}(x, y) = E((x - E(x))(y - E(y)))$$

という公式が成り立つ。証明は学生各自で行うこと。

ところで、先ほどの重積分について興味深い結果がある。いま、 $L_c^2$  を連続で、 $E((f(x, y))^2)$  が実数になるような関数  $f(x, y)$  をすべて集めてできた集合としよう。この集合上で、

$$\|f\| = \sqrt{E((f(x, y))^2)}$$

としてノルム (norm) を定義し、また

$$f \cdot g = E(f(x, y)g(x, y))$$

として内積 (inner product) を定義する。ノルムは定義できるが内積は定義できるかどうかかわからないじゃないか、と思われるかもしれないが、実は定義できる。実際、

$$\begin{aligned} 0 &\leq |f(x, y)g(x, y)| \\ &\leq \frac{1}{2}(f(x, y) + g(x, y))^2 \end{aligned}$$

であるが、

$$0 \leq (f(x, y) + g(x, y))^2 \leq 4 \max\{(f(x, y))^2, (g(x, y))^2\} \leq 4[(f(x, y))^2 + (g(x, y))^2]$$

なので、

$$0 \leq E((f(x, y) + g(x, y))^2) \leq 4E((f(x, y))^2) + 4E((g(x, y))^2) < \infty$$

となる。したがって

$$0 \leq E(|f(x, y)g(x, y)|) < \infty$$

であるが、一方で

$$h^+(x, y) = \max\{f(x, y)g(x, y), 0\}, \quad h^-(x, y) = \min\{0, f(x, y)g(x, y)\}$$

とすると、

$$f(x, y)g(x, y) = h^+(x, y) + h^-(x, y)$$

であり、

$$|h^\pm(x, y)| \leq |f(x, y)g(x, y)|$$

なので、 $E(h^\pm(x, y))$  は実数値である。よって  $E(f(x, y)g(x, y))$  も実数値であることがわかった。特に、 $f(x, y) = x - E(x)$  とし、 $g(x, y) = y - E(y)$  とすれば、これは  $V(x)$  と  $V(y)$  が共に実数値であれば  $Cov(x, y)$  も実数値であるという意味を有することに注意が必要である。

上の論証の途中で、実は興味深い結果が得られている。いま、集合  $L_c^2$  に

$$(f + g)(x, y) = f(x, y) + g(x, y), \quad (cf)(x, y) = cf(x, y)$$

と定義することで、足し算と定数倍が定義できる。 $f \in L_c^2$  のとき  $cf \in L_c^2$  は当たり前だが、 $f, g \in L_c^2$  であっても  $f + g \in L_c^2$  ではないのでは？ と思うかもしれないが、上で示したように  $f, g \in L_c^2$  ならば  $f + g \in L_c^2$  である。そこで、第8章で議論したように、 $L_c^2$  はベクトル空間の性質 1.~8. をすべて満たし、また内積は第8章の性質 (i)-(iv) をすべて満たす。さらに、

$$\|f\|^2 = f \cdot f$$

が得られている。そこで

$$h(t) = \|f + tg\|^2$$

と定義してみると、すべての  $t$  について  $h(t) \geq 0$  である。一方、

$$h(t) = (f + tg) \cdot (f + tg) = \|f\|^2 + 2(f \cdot g)t + \|g\|^2 t^2$$

とできる。  $g \equiv 0$  でないとするれば  $\|g\|^2 > 0$  なので、二次方程式の判別式条件から

$$(f \cdot g)^2 \leq \|f\|^2 \|g\|^2$$

が得られ、平方根を取って

$$|f \cdot g| \leq \|f\| \|g\|$$

が得られる。もちろん  $g \equiv 0$  ならば上の不等式は当たり前である。こうして我々は **Cauchy-Schwarz** の不等式に到達した。

第8章の Cauchy-Schwarz の不等式と、出し方がまったく同じであることを注意して欲しい。この当たりの帰結として、 $f(x, y) = x - E(x)$ ,  $g(x, y) = y - E(y)$  に上の不等式を適用すると

$$\begin{aligned} Cov(x, y) &= E((x - E(x))(y - E(y))) \\ &\leq \sqrt{E((x - E(x))^2)E((y - E(y))^2)} \\ &= \sqrt{V(x)V(y)} \end{aligned}$$

が得られる。つまり、分散の積は共分散の二乗を常に上回る。この結果は重要なので、覚えておくとよい。

なお、 $X$  と  $Y$  が無相関であることは、 $Cov(x, y) = 0$  と同値である。これは定義から当たり前である。