経済数学（法政用）: 第１３章

細矢祐誉

テーマ：確率とスティルチェス積分

・確率と累積分布関数

確率とはなにか。これを議論し出すと哲学の話になって収拾が付かないので、このテキストではあくまで、確率を数学的にどう表現するかという話だけを扱う。確率を表す手段は複数存在するが、ここで扱うのは、累積分布関数を用いた表現である。累積分布関数は $F:\mathbb{R}\to[0,1]$ という形で表される関数で、次の 3 つの性質を持つものである。

1. 非減少である。つまり、$x<y$ ならば $F(x)\le F(y)$ である。
2. 右側連続である。つまり、$x\downarrow y$ ならば $F(x)\downarrow F(y)$ である。
3. $\lim_{x\to-\infty}F(x)=0$ かつ $\lim_{x\to\infty}F(x)=1$ である。

この $F(x)$ はなにを表すかというと、ある確率に従う数 X が、x 以下になる確率を表現している。たとえば、X がさいころの出目だったとしてみよう。さいころは六面体で、出る目の確率はすべて同じだったとすると、

$$F(x)=\begin{cases}0 & \text{if } x<1,\\ \frac{k}{6} & \text{if } 1\le k\le 5,\ k\le x<k+1,\\ 1 & \text{if } x\ge 6.\end{cases}\tag{1}$$

という関数が、このさいころの出目 X に対応する累積分布関数である。

当然ながら、$x<y$ ならば、$F(y)-F(x)$ は X が $x<X\le y$ という条件を満たす確率を表している。このようにして、我々が確率と呼ぶものは、すべて累積分布関数で表すことができる。逆に累積分布関数をひとつ与えれば、それはすなわち、確率をひとつ与えたということになる。

・スティルチェス積分

リーマン和 $R(f,a,b,\Delta,s)$ を思い出そう。この正確な定義は

$$R(f,a,b,\Delta,s)=f(s_n)(t_n-t_{n-1})+f(s_{n-1})(t_{n-1}-t_{n-2})+...+f(s_1)(t_1-t_0)$$

で与えられた。ただし Δ は $(t_0,...,t_n)$ という有限点列で、ただし $a=t_0<t_1<...<t_n=b$ となるもの、また s_i は t_{i-1} と t_i の間にある数であった。$|\Delta|$ は t_i-t_{i-1} の最大

値を表すとし、

$$\lim_{|\Delta|\to 0} R(f,a,b,\Delta,s) = \alpha$$

となる数 α が存在するとき、f は $[a,b]$ 上でリーマン積分可能と呼び、

$$\alpha = \int_a^b f(x)dx$$

と表すのであった。

なぜこんなことを書くかというと、これを拡張したいからである。いま累積分布関数 F が与えられているとして、

$$\begin{aligned}R(f,a,b,\Delta,s,F) &= f(s_n)(F(t_n)-F(t_{n-1})) + f(s_{n-1})(F(t_{n-1})-F(t_{n-2})) \\ &\quad + ... + f(s_1)(F(t_1)-F(t_0))\end{aligned}$$

と定義する。$[a,b]$ 内で $F(x)=x$ であればこの定義は普通のリーマン和と同じであるが、F があるおかげで、形が変わっている。前と同様に、

$$\lim_{|\Delta|\to 0} R(f,a,b,\Delta,s,F) = \alpha$$

となる数 α が存在するとき、f は $[a,b]$ 上で F について**スティルチェス積分可能**と呼び[*1]、

$$\alpha = \int_a^b f(x)dF$$

と書く。

リーマン積分の通常の定義と同様、

$$\int_a^b (f(x)+g(x))dF = \int_a^b f(x)dF + \int_a^b g(x)dF,$$

$$\int_a^b cf(x)dF = c\int_a^b f(x)dF$$

が当然のように成り立つ。また、$a<c<b$ であれば、

$$\int_a^b f(x)dF = \int_a^c f(x)dF + \int_c^b f(x)dF$$

[*1] 正確にはリーマン=スティルチェス積分可能。これに対してカラテオドリの拡張定理をベースとして定義される、ルベーグ=スティルチェス積分というべつのものが存在する。

も成り立つ。一方で、$b < a$ のときに、

$$\int_a^b f(x)dF = -\int_b^a f(x)dF$$

と**定めてはならない！**　なぜならば、$a < b$ の時には、$f(x) \equiv 1$ であれば、

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

となり、確率変数 X が $a < X \leq b$ を満たす確率である。この不等号を見ればわかるとおり、a と b の取り扱いはもはや対称ではないので、同じように議論しようとすると痛い目にあう。

　一方で、

$$\int_{-\infty}^b f(x)dF,$$

$$\int_a^\infty f(x)dF,$$

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

等の定義は、リーマン積分のときと同様である。ここで重要な定理がある。

定理１：任意の累積分布関数 F に対して、f が $[a, b]$ 上連続関数であれば、それはスティルチェス積分可能である。

　というわけで、だいたいの関数を扱う限り、スティルチェス積分可能かどうかについて我々が思い悩む必要はない。

　特に $f(x)$ が $\mathbb{R}$ 全体で積分可能（つまり、$\int_{-\infty}^\infty f(x)dF$ が実数値として存在する）なとき、

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

と定義し、これを累積分布関数 F で表される確率についての関数 $f(x)$ の**期待値** (expectation) と呼ぶ。文脈上いま考えている F が明確であるときにはその表記を省略して単に

$$E(f(x))$$

と書くが、右辺の定義には F が必要なので、書いてないだけで常に考えていると思っておいたほうがよい。

なお、当然のことながら、定数 c については

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

である。したがって、

$$E(c) = c$$

が成り立つ。

・離散型の確率

累積分布関数 F は、有限個の不連続点を除いてその付近で定数関数であるとき、その表現する確率は**離散型の確率**であると言う。たとえば (1) 式で表したさいころの出目 X の確率を表す累積分布関数 F は、$\{1,2,3,4,5,6\}$ という不連続点を除いてその付近で定数であるから、これは離散型の確率である。

F が X を表現する離散型の確率であり、その不連続点が $\{x_1, ..., x_n\}$ であるとする。ただし $x_1 < x_2 < ... < x_n$ である。このとき、$x_0 < x_1$ を適当に取れば、$F(x_0) = 0, F(x_n) = 1$ である。そこで、

$$p_i = F(x_i) - F(x_{i-1})$$

と置くと、

$$p_1 + p_2 + ... + p_n = F(x_1) - F(x_0) + F(x_2) - F(x_1) + ... + F(x_n) - F(x_{n-1}) = F(x_n) - F(x_0) = 1$$

となることがわかる。この p_i は X が x_i を取る確率を表現している。

当然ながら、期待値を議論するためには以下のことが成り立っていないといけないであろう。

定理２：F が離散型の確率で不連続点 $x_1, ..., x_n$ と確率 $p_1, ..., p_n$ が上で書かれたように定まっているとき、

$$E(f(x)) = p_1 f(x_1) + p_2 f(x_2) + ... + p_n f(x_n)$$

が成り立つ。

証明：まず、$a < x_0 < x_1 < x_n < b$ となる任意の a, b, x_0 を取る。そして、$\Delta = (t_0, ..., t_m)$ とし、ただし $a = t_0 < t_1 < ... < t_m = b$ で、また、$t_i = x_0$ となる i と、$t_j = x_n$ となる

j が存在するとしよう。$\Delta' = (t_i, ..., t_j)$ とし、対応して $s' = (s_{i+1}, ..., s_j)$ を取れば、x_n 以降と x_0 以前では F は定数であるから、

$$\begin{aligned} R(f, a, b, \Delta, s, F) &= f(s_1)(F(t_1) - F(t_0)) + ... + f(s_m)(F(t_m) - F(t_{m-1})) \\ &= f(s_{i+1})(F(t_{i+1}) - F(t_i)) + ... + f(s_j)(F(t_j) - F(t_{j-1})) \\ &= R(f, x_0, x_n, \Delta', s', F) \end{aligned}$$

となる。$|\Delta'| \leq |\Delta|$ なので、$|\Delta| \to 0$ として極限を取れば、

$$\int_a^b f(x)dF = \int_{x_0}^{x_n} f(x)dF$$

となることがわかる。したがって $a \to -\infty, b \to \infty$ とすれば、

$$E(f(x)) = \int_{x_0}^{x_n} f(x)dF$$

である。

次に、$\Delta = (t_0, ..., t_m)$ を取り直し、今度は $x_0 = t_0 < t_1 < ... < t_m = x_n$ とし、またすべての x_i に対して $x_i = t_{j_i}$ となる j_i が存在したとする。このとき、$s_i = t_i$ として s を取ると、

$$\begin{aligned} R(f, x_0, x_n, \Delta, s, F) &= f(t_1)(F(t_1) - F(t_0)) + ... + f(t_m)(F(t_m) - F(t_{m-1})) \\ &= f(t_{j_1})(F(t_{j_1}) - F(t_{j_1 - 1})) + ... + f(t_{j_n})(F(t_{j_n}) - F(t_{j_n - 1})) \\ &= f(x_1)(F(x_1) - F(x_0)) + f(x_2)(F(x_2) - F(x_1)) \\ &\quad + ... + f(x_n)(F(x_n) - F(x_{n-1})) \\ &= p_1 f(x_1) + ... + p_n f(x_n) \end{aligned}$$

が成り立つ。$|\Delta| \to 0$ とすればただちに

$$E(f(x)) = \int_{x_0}^{x_n} f(x)dF = p_1 f(x_1) + ... + p_n f(x_n)$$

となって証明が完成する。

なお、上では有限個の点だけで不連続な F を考えたが、これは無限個でも場合によっては拡張できる。以下に、ひとつだけ例を挙げてみよう。

例 1（幾何分布）: $0 < p < 1$ とし、X は確率 p で成功する独立[*2]な試行を何度も行ったときに、最初に成功したときの行った回数を指しているとする。このとき、たとえば $X = 3$

[*2] この独立という言葉の定義は次回以降に精密に議論する。

である確率 p_3 は「一回目失敗して、二回目失敗して、三回目成功する」確率であるから、$(1-p)^2p$ である。同様にして $p_n = (1-p)^{n-1}p$ とすれば、

$$F(x) = \begin{cases} 0 & x < 1, \\ p_1 & 1 \leq x < 2, \\ p_1 + p_2 & 2 \leq x < 3, \\ ... & \\ p_1 + p_2 + ... + p_n & n \leq x < n-1, \\ ... & \end{cases}$$

というのがこの X が従う確率の累積分布関数であって、この F は厳密な意味では離散型の確率を表していない（可算無限個の不連続点がある）。しかし、これも離散型の確率の亜種であり、あまり区別はされていない。実際、定理 2 と同様に計算すれば、$n \leq b < n+1$ であるときには

$$\int_{-\infty}^{b} f(x)dF = \sum_{i=1}^{n} p_i f(i)$$

であることがわかるので、$b \to \infty$ とすれば、

$$E(f(x)) = \sum_{i=1}^{\infty} p_i f(i) = \lim_{n\to\infty} \sum_{i=1}^{n} p_i f(i)$$

であることが証明できる。

例 2（ポワソン分布）: $\lambda > 0$ とし、この λ はある「現象」が一定時間（この時間を 1 としておく）以内に平均的に起こる回数を指す（ポワソンの元の研究によると、一年間に馬に蹴られて死ぬ兵士の数）とする。X はこの「現象」が時間 1 の間に起こる回数であるとする。$X = k$ である確率を考えるために、次のようなことを考えよう。まず、時間 1 を N 等分して、$T_i = \frac{i}{N}$ とする。T_{i-1} から T_i の間に「現象」が二回以上起こることは、T_{i-1} と T_i の差が極めて小さければ、ほとんど無視できる確率だろうから、$X = k$ である確率は、T_{i-1} と T_i の間に「現象」が「一回」起こるような i がちょうど k 個ある確率と、ほぼ同一視できるであろう。「現象」は時間 1 で平均 λ 発生するので、T_{i-1} と T_i の間に発生する確率は

$$\frac{\lambda}{N}$$

である。したがって $X = k$ である確率は近似的に

$${}_NC_k \left(1 - \frac{\lambda}{N}\right)^{N-k} \left(\frac{\lambda}{N}\right)^k$$

である。（ただし $_NC_k$ は二項係数 $\frac{k!(N-k)!}{N!}$ を表す）

　たいへん面倒だが、上の式は $N \to \infty$ としたときに収束することが知られていて、その極限は

$$p_k = \frac{\lambda^k e^{-\lambda}}{k!}$$

で表される。そこで、

$$F(x) = \begin{cases} 0 & x < 1, \\ p_1 & 1 \leq x < 2, \\ p_1 + p_2 & 2 \leq x < 3, \\ ... & \\ p_1 + p_2 + ... + p_n & n \leq x < n+1, \\ ... & \end{cases}$$

とすれば、これが X が従う確率の累積分布関数である。これも上と同様に、

$$E(f(x)) = \sum_{i=1}^{\infty} p_i f(i)$$

であることが示せる。

・連続型の確率

　今度は、F が連続微分可能である場合を考える。この場合は、この F で表される確率は連続型の確率と呼ばれる。いま、$\Delta = (t_0, ..., t_n)$ を取ってきて、ただし $a = t_0 < t_1 < ... < t_n = b$ とする。平均値の定理から、$t_{i-1} < s_i < t_i$ を満たすある s_i について、

$$\frac{F(t_i) - F(t_{i-1})}{t_i - t_{i-1}} = F'(s_i)$$

が成り立つ。これを変形して

$$F(t_i) - F(t_{i-1}) = F'(s_i)(t_i - t_{i-1})$$

と書き直すと、

$$\begin{aligned} R(f, a, b, \Delta, s, F) &= f(s_1)(F(t_1) - F(t_0)) + f(s_2)(F(t_2) - F(t_1)) + ... + f(s_n)(F(t_n) - F(t_{n-1})) \\ &= f(s_1)F'(s_1)(t_1 - t_0) + f(s_2)F'(s_2)(t_2 - t_1) + ... + f(s_n)F'(s_n)(t_n - t_{n-1}) \\ &= R(f \times F', a, b, \Delta, s) \end{aligned}$$

である。$|\Delta| \to 0$ として極限を取れば、

$$\int_a^b f(x)dF = \int_a^b f(x)F'(x)dx$$

という形で、通常のリーマン積分で議論が完結してしまう。まとめると次の定理を得る。

定理３：F が連続微分可能であれば、

$$\int_a^b f(x)dF = \int_a^b f(x)F'(x)dx$$

が成り立つ。

これについて、興味深い議論をしてみよう。いま、$g(x)$ は実数 $\mathbb{R}$ 上で定義された 0 以上の連続関数で、

$$\int_{-\infty}^{\infty} g(x)dx = 1$$

が成り立っていたとする。

$$F(x) = \int_{-\infty}^{x} g(x)dx$$

と定義すれば、微分積分学の基本定理から

$$F'(x) = g(x)$$

である。言うまでもなく、$f(x) \equiv 1$ であれば、

$$\int_{-\infty}^{x} f(x)g(y)dy = F(x) = \int_{-\infty}^{x} f(x)dF$$

である。定理 3 はこの式を、はるかに一般の関数 f の場合に証明したものになっている。

・一般の確率

以上で、離散型の確率と連続型の確率について話した。しかしこれらはどのくらい一般的だろうか？　実を言うと、確率論にはこれをはるかに一般化した議論がある。それによると、確率の分類で特異確率と呼ばれるものがあって、それは離散型の確率を少し発展させた形をしている。一方、絶対連続確率と呼ばれるものがあって、これは連続型の確率における F の連続微分可能性を少しだけ弱めたものになっている。ルベーグの分解定理と

いう定理によれば、任意の累積分布関数 F は、特異確率の累積分布関数 F_1 と、絶対連続確率の累積分布関数 F_2 について、

$$F = aF_1 + (1-a)F_2$$

という形で書ける。しかもこの記述において、a と F_1 と F_2 は F からたったひとつに定まる[*3]。

というわけで、離散型の確率と連続型の確率という我々の分類は、現代的な積分論の枠組みにおいて、それほど的外れではないことがわかった。この分け方は、ある意味でとても、本質的である。

[*3] なお、このとき $F_2'(x)$ に当たる関数について定理 3 と同じ主張が成り立つが、これを**ラドン＝ニコディムの定理**と呼ぶ。