第2章　連続な確率変数とその分布

さしあたり，根元事象については深く考えないで，ある全事象 Ω が十分大きく与えられているとする．この上で定義された確率変数がよい分布を持つ場合を先に考えよう．難しいことを後回しにするとこれは，$a < b$ を任意に与えたとき，

$$p_X([a,b]) = P(X \in [a,b]) = P(a \leq X \leq b)$$

を考えることに相当する．

2.1　密度関数を持つ分布とその期待値，分散

定義 2.1 X の分布 p_X が密度関数 $f(x)$ を持つとは

$$p_X([a,b]) = \int_a^b f(x)dx$$

が任意の $a \leq b$ に対して成り立つときに言う．

p_X が密度関数 $f(x)$ を持つとき，X の期待値 EX，分散 $V(X)$ はそれぞれ

$$EX = \int_{-\infty}^{\infty} xf(x)dx \tag{2.1}$$

および

$$V(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - (EX)^2 \tag{2.2}$$

によって定義される．形式的にはシグマを積分に書き換えたものである．いくつかこの様な密度関数を持つ分布を紹介する．

2.1.1 $[a,b]$ 上の一様分布 $Unif([a,b])$

これが一番簡単な例．密度関数は

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, \text{その他} & \end{cases}$$

のとき．

練習問題 2.1 X が $[0,5]$ 上の一様分布の時密度関数 $f(x)$ と期待値 EX，分散 $V(X)$ を計算せよ．

2.1.2 指数分布 $Exp(\lambda)$

指数分布もパラメータ $\lambda > 0$ を持っている．その密度関数は

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$$

で与えられる． X がパラメータ λ の指数分布に従うとき，

$$EX = \int_0^\infty x\lambda e^{-\lambda x} dx = \frac{1}{\lambda}\int_0^\infty y e^{-y} dy = \frac{1}{\lambda}$$

同じように計算して

$$EX^2 = \frac{1}{\lambda^2}\int_0^\infty y^2 e^{-y} dy = \frac{2}{\lambda^2}$$

なので，

$$V(X) = EX^2 - (EX)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

2.1.3 正規分布 $N(m,\sigma^2)$

この密度関数は二つのパラメータを持つ．平均 m と標準偏差 $\sigma > 0$ である．普通この分布を $N(m,\sigma^2)$ と書くが，X が正規分布 $N(m,\sigma^2)$ に従うとき，

$$p_X([a,b]) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/(2\sigma^2)} dx$$

となる．一見むちゃくちゃ難しそうに見えるが，大事な密度関数なので覚えてほしい．期待値を計算すると

$$
\begin{aligned}
EX &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{-\infty} xe^{-(x-m)^2/(2\sigma^2)}dx \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{-\infty} (y+m)e^{-y^2//(2\sigma^2)}dy \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\infty} (z\sigma+m)e^{z^2/2}dz \\
&= m
\end{aligned}
$$

これは多変数の微積で習った

$$
\begin{aligned}
\left(\int_0^\infty e^{-ax^2}dx\right)^2 &= \int_0^\infty\int_0^\infty e^{-a(x^2+y^2)}dxdy \\
&= \int_0^\infty\int_0^{\pi/2} e^{-ar^2}rdrd\theta \\
&= \frac{\pi}{2}\int_0^\infty e^{-u}\frac{1}{2a}du \\
&= \frac{\pi}{4a}
\end{aligned}
$$

という計算で，$a=\frac{1}{2}$ をつかうと

$$
\int_0^\infty e^{-z^2/2}dz = \sqrt{\frac{\pi}{2}}
$$

となり，積分の中身は偶関数だから 2 倍して

$$
\int_{-\infty}^\infty e^{-z^2/2}dz = \sqrt{2\pi}
$$

となることから出てくる．

分散の計算には $\int_{-\infty}^\infty z^2e^{z^2/2}dz$ の計算が必要だがこれは部分積分で上の計算に帰着する．$(e^{-x^2/2})' = -xe^{-x^2/2}$ だから，

$$
\int_{-\infty}^\infty x^2e^{-x^2/2}dx = \left[-xe^{-x^2/2}\right]_{-\infty}^\infty + \int_{-\infty}^\infty e^{-x^2/2}dx = 0+\sqrt{2\pi}
$$

したがって

$$
V(X) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^\infty x^2e^{-(x-m)^2/2\sigma^2}dx - m^2
$$

で $y = \frac{x-m}{\sigma}$ とおくと右辺第 1 項は

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma y + m)^2 e^{-y^2/2} dy = \sigma^2 + 0 + m^2$$

となるので，$V(X) = \sigma^2$ となる．(パラメータ m と σ は平均と標準偏差 $\sqrt{V(X)}$ を表している．

練習問題 2.2 X が正規分布 $N(m, \sigma^2)$ に従うとき $Z = \dfrac{x-m}{\sigma}$ は $N(0,1)$ に従うことを確かめよ．実際には任意の実数 t に対して

$$P(Z \leq t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-z^2/2} dz$$

と言う式を確かめよ．