鹿児島大学　渡邉研究室
wata104@eee

範囲の期待値：係数 d_2 の計算式

1. 確率密度関数と累積分布関数

区間 $-\infty < x < \infty$ で定義される確率変数が $x \sim x + dx$ の間の値をとる確率を確率密度関数 $f(x)$ で表す。累積分布関数 $F(x)$ は、確率変数が x 以下の値をとる確率であり、

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

で表される。ここで、$F(-\infty) = 0,\ F(\infty) = 1$ である。また、

$$f(x) = \frac{dF(x)}{dx} = F'(x)$$

が成り立つ。

2. 最大値の期待値

大きさ n のサンプル x_i の最大値 x_L の期待値の計算式を求める。ある x_i に対して、それ以外の $n-1$ 個の x_j（$j \neq i$）が x_i 以下になる確率を考えると

$$\begin{aligned}
\overline{x_L} &= \sum_{i=1}^{n} \int_{-\infty}^{\infty} x_i f(x_i) F(x_i)^{n-1} dx_i \\
&= n \int_{-\infty}^{\infty} x f(x) F(x)^{n-1} dx \\
&= \int_{-\infty}^{\infty} x \cdot n F'(x) F(x)^{n-1} dx \\
&= \int_{-\infty}^{\infty} x \cdot (F(x)^n)' dx
\end{aligned}$$

となる。

3. 最小値の期待値

大きさ n のサンプル x_i の最小値 x_S の期待値の計算式も、同様に求めることができる。ある x_i に対して、それ以外の $n-1$ 個の x_j（$j \neq i$）が x_i 以上になる確率を考えると

$$\begin{aligned}
\overline{x_S} &= \sum_{i=1}^{n} \int_{-\infty}^{\infty} x_i f(x_i) (1 - F(x_i))^{n-1} dx_i \\
&= n \int_{-\infty}^{\infty} x f(x) (1 - F(x))^{n-1} dx \\
&= \int_{-\infty}^{\infty} x \cdot n F'(x) (1 - F(x))^{n-1} dx \\
&= -\int_{-\infty}^{\infty} x \cdot ((1 - F(x))^n)' dx
\end{aligned}$$

となる。

4. 範囲の期待値

最大値 x_L と最小値 x_S の差を範囲（レンジ）という。大きさ n のサンプル x_i の範囲 r の期待値は

$$\bar{r} = \overline{x_L} - \overline{x_S} = \int_{-\infty}^{\infty} x \cdot (F(x)^n)' dx + \int_{-\infty}^{\infty} x \cdot ((1 - F(x))^n)' dx$$

$$= \int_{-\infty}^{\infty} x \cdot (F(x)^n + (1 - F(x))^n)' dx$$

$$= \int_{-\infty}^{\infty} x \cdot (-1 + F(x)^n + (1 - F(x))^n)' dx$$

$$= \left[x \cdot \left(-1 + F(x)^n + \left(1 - F(x)\right)^n\right)\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} (-1 + F(x)^n + (1 - F(x))^n) dx$$

$$= \infty \cdot (-1 + 1^n + 0^n) + \infty \cdot (-1 + 0^n + 1^n)$$

$$+ \int_{-\infty}^{\infty} (1 - F(x)^n - (1 - F(x))^n) dx$$

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

で求められる。

（別解）

$$\bar{r} = \overline{x_L} - \overline{x_S} = \int_{-\infty}^{\infty} x \cdot (F(x)^n)' dx + \int_{-\infty}^{\infty} x \cdot ((1 - F(x))^n)' dx$$

$$= \int_{-\infty}^{\infty} x \cdot (F(x)^n + (1 - F(x))^n)' dx$$

$$= [x \cdot (F(x)^n + (1 - F(x))^n)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} (F(x)^n + (1 - F(x))^n) dx$$

$$= [x \quad]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} (F(x)^n + (1 - F(x))^n) dx$$

$$= \int_{-\infty}^{\infty} dx - \int_{-\infty}^{\infty} (F(x)^n + (1 - F(x))^n) dx$$

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

と計算することもできる。

5. 範囲と標準偏差

確率変数 x が元の変数 X の平均 μ と標準偏差 σ によって $x = (X-\mu)/\sigma$ と規格化された値であるとき、X の範囲の期待値 $\bar{R}$ と標準偏差 σ の関係は

$$\bar{R} = \int_{-\infty}^{\infty} (1 - F(X)^n - (1 - F(X))^n) dX$$

$$= \sigma \int_{-\infty}^{\infty} (1 - F(x)^n - (1 - F(x))^n) dx$$

となる。したがって、$\bar{R}$ と σ の比をサンプルの大きさ n によって決まる係数 d_2 で表して $\bar{R} = d_2 \sigma$ とすると

$$d_2 = \frac{\bar{R}}{\sigma} = \int_{-\infty}^{\infty} (1 - F(x)^n - (1 - F(x))^n) dx$$

となる。

累積分布関数 $F(x)$ が標準正規分布

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

にしたがうとき、上式から d_2 の値を計算すると、下の表のようになる。

表　d_2 の値

n	d_2
2	1.12838
3	1.69257
4	2.05875
5	2.32593
6	2.53441
7	2.70436
8	2.84720
9	2.97003
10	3.07750

この値は、品質管理で用いられる $\bar{X}-R$ 管理図で、$\bar{X}$ の管理限界を

$$E(\bar{X}) \pm 3D(\bar{X}) = \bar{\bar{X}} \pm 3\frac{\sigma}{\sqrt{n}} = \bar{\bar{X}} \pm \frac{3}{d_2\sqrt{n}}\bar{R} = \bar{\bar{X}} \pm A_2\bar{R}$$

として定める際に用いられる。

参考文献

・森口　繁一「初等数理統計学」改訂版、培風館、1957
・中村　達男「管理図の作り方と活用」（新版　QC 入門講座　7）、日本規格協会、1999