

連続量  $x$  の分布がカテゴリ化された度数分布  
 $(x_i, p_i)$  として与えられた場合に、  
 $x$  の単調関数  $y = f(x)$  の  
 任意の  $y$  を中点とする区間の確率  $\pi_y$  を求める方法。

980907,08,071022 前川眞一

1.  $(x_i, p_i)$ ,  $i = 1, 2, \dots, n$ , をデータとして与えられた度数分布とする。
2. 累積確率  $c_i = \sum_{k=1}^i p_k / \sum_{k=1}^n p_k$  を求める。
3.  $\mathbf{x} = \{x_1 < x_2 < \dots < x_n\}$  をカテゴリ化された  $x$  の中点とみなし、各カテゴリの上端点  $b_i$  を求める。

$$b_i = (x_i + x_{i+1})/2, \quad i = 1, 2, \dots, n-1$$

そして、連続量としての  $x$  の分布関数を  $(b_i, c_i)$ ,  $i = 1, 2, \dots, n-1$  で近似する。すなわち、

$$\Pr(X \leq b_i) = F_x(b_i) \approx c_i$$

4.  $\mathbf{y} = \{y_1 < y_2 < \dots < y_m\}$  をカテゴリ化された  $y$  の中点とみなし、各カテゴリの端点  $\beta_i$  を求める。

$$\beta_i = (y_i + y_{i+1})/2, \quad i = 1, 2, \dots, m-1$$

5.  $x$  の上端点  $b_i$  で関数  $f$  を評価し、 $y$  の累積分布関数表

$$(f(b_i), c_i), \quad i = 1, 2, \dots, n-1$$

を作る。

6.  $(f(b_i), c_i)$ ,  $i = 1, 2, \dots, n-1$ , を補間して、 $y$  の上端点までの累積分布関数表

$$(\beta_i, \gamma_i), \quad i = 1, 2, \dots, m-1$$

を求める。ただし、

$$f(b_1) \leq \beta_1, \quad \beta_{m-1} \leq f(b_{n-1})$$

であるとする。

これが、連続量  $y$  の分布関数の近似

$$\Pr(Y \leq \beta_i) = F_y(\beta_i) \approx \gamma_i, \quad i = 1, 2, \dots, m-1$$

を与える。

7.  $y$  のカテゴリの中点における確率  $\pi_i$  を

$$\pi_1 = \gamma_1, \quad \pi_i = \gamma_i - \gamma_{i-1}, \quad i = 2, 3, \dots, m$$

として求める。これは、上下の終端カテゴリに関して、

$$\Pr(Y \leq y_1) = F_y(y_1) = 0, \quad \Pr(Y \leq y_m) = F_y(y_m) = 1$$

を仮定している。

$n = 5$  の場合の  $x$  の中点と上端点の関係

	MIN				MAX
	1	2	3	4	5 = n
$x$	----- ----- ----- -----				
$b$					
	1	2	3	4 = n - 1	