応用数理 E 第 5 回目

永幡幸生
新潟大学工学部

4 月 22 日

多次元分布

前回は確率変数が 1 つだけのケースを考えたが、2 つ以上でも同じように考える。

離散型の場合

確率分布として

$P(X = x_k)$, $P(Y = y_k)$ だけでは不十分で

$P(X = x_k, Y = y_k)$ の表を作る必要がある。

例

さいころを 2 つ投げた場合等確率のモデルを考えて

X : 一つ目のさいころの目、Y : 二つ目のさいころの目

とすると一枚目の表ができ、これを (X, Y) の同時分布、

$Z = \max\{X, Y\}$、$W = |X - Y|$

とすると二枚目の表ができ、これを (Z, W) の同時分布と呼ぶ。

多次元分布

さいころを2つ投げた時：X:1つ目のさいころ、Y:2つ目のさいころ							
X Y	1	2	3	4	5	6	計
1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
3	1/36	1/36	1/36	1/36	1/36	1/36	1/6
4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
5	1/36	1/36	1/36	1/36	1/36	1/36	1/6
6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
計	1/6	1/6	1/6	1/6	1/6	1/6	1

多次元分布

さいころを2つ投げた時：Z = max{X,Y}, W = |X-Y|							
Z W	0	1	2	3	4	5	計
1	1/36	0	0	0	0	0	1/36
2	1/36	1/18	0	0	0	0	1/12
3	1/36	1/18	1/18	0	0	0	5/36
4	1/36	1/18	1/18	1/18	0	0	7/36
5	1/36	1/18	1/18	1/18	1/18	0	1/4
6	1/36	1/18	1/18	1/18	1/18	1/18	11/36
計	1/6	5/18	2/9	1/6	1/9	1/18	1

多次元分布

(X, Y) の同時分布が与えられれば、X (Y) の分布も分かる。次のページで見れば黄色に色づけされたところに、X の分布 (青色のところに、Y の分布）が表れるが、これを X の周辺分布と呼ぶ。一枚目の表では対称性があって、どちらが、X でどちらが Y に対応しているのか分かりにくいが、二枚目の表では Z, W どちらに対応しているかよく見えている。

多次元分布

さいころを2つ投げた時：X:1つ目のさいころ、Y:2つ目のさいころ							
X Y	1	2	3	4	5	6	計
1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
3	1/36	1/36	1/36	1/36	1/36	1/36	1/6
4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
5	1/36	1/36	1/36	1/36	1/36	1/36	1/6
6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
計	1/6	1/6	1/6	1/6	1/6	1/6	1

多次元分布

さいころを2つ投げた時：Z = max{X,Y}, W = |X-Y|							
Z W	0	1	2	3	4	5	計
1	1/36	0	0	0	0	0	1/36
2	1/36	1/18	0	0	0	0	1/12
3	1/36	1/18	1/18	0	0	0	5/36
4	1/36	1/18	1/18	1/18	0	0	7/36
5	1/36	1/18	1/18	1/18	1/18	0	1/4
6	1/36	1/18	1/18	1/18	1/18	1/18	11/36
計	1/6	5/18	2/9	1/6	1/9	1/18	1

多次元分布

連続型の場合

確率変数が 1 つだけのケースでは積分で与えられたが、2 つ（以上）のケースでは重積分になる。

$$P(a_1 \leq X \leq b_1, a_2 \leq Y \leq b_2) = \iint_{[a_1,b_1]\times[a_2,b_2]} f(x,y)dxdy$$

このとき $f(x,y)$ を同時確率密度関数と呼ぶ。

多次元分布

X の周辺分布は

$$P(a_1 \leq X \leq b_1) = P(a_1 \leq X \leq b_1, -\infty < Y < \infty)$$

$$= \iint_{[a_1,b_1]\times\mathbb{R}} f(x,y)dxdy = \int_{[a_1,b_1]} \left(\int_{-\infty}^{\infty} f(x,y)dy \right) dx$$

となり

$$\int_{-\infty}^{\infty} f(x,y)dy$$

を X の周辺確率密度関数と呼ぶ。

同様に Y の周辺分布、周辺確率密度関数はそれぞれ

$$P(a_2 \leq Y \leq b_2) = \int_{[a_2,b_2]} \left(\int_{-\infty}^{\infty} f(x,y)dx \right) dy$$

$$\int_{-\infty}^{\infty} f(x,y)dx$$

になる。

多次元分布

確率変数の独立

定義

全ての $a \leq b,\ c \leq d$ に対して、
$P(a \leq X \leq b, c \leq Y \leq d) = P(a \leq X \leq b)P(c \leq Y \leq d)$
であるとき独立である。

補題

X, Y が独立と以下は同値
離散型の場合 $P(X = x_k, Y = y_j) = P(X = x_k)P(Y = y_j)$, が全ての x_k, y_j に対して成り立つ
連続型の場合 $f(x, y) = g(x)h(y)$ と書ける。

多次元分布

統計で必要になるいくつかの重要な分布とその性質を挙げる。
基本的には独立な正規分布から何らかの形で導き出されているものである。

正規分布の再生性

X_1, X_2 をそれぞれ $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ に従う独立な正規分布とすると $a_1X_1 + a_2X_2 + a_3$ は $N(a_1\mu_1 + a_2\mu_2 + a_3, a_1^2\sigma_1^2 + a_2\sigma_2^2)$ に従う正規分布となる。

正規分布の性質

$X_1, X_2, \ldots, X_n$ を互いに独立で同じ分布（独立同分布と呼ぶ）に従い、さらにその分布が $N(\mu, \sigma^2)$ ならば

$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$ は $N(\mu, \frac{\sigma^2}{n})$ に従う。

多次元分布

問題

正規分布の性質を正規分布の再生性を用いて証明せよ。

多次元分布

χ^2 分布

$X_1, X_2, \ldots, X_n$ を独立同分布で $N(0,1)$ に従うものとする。このとき

$$\chi^2 = \sum_{k=1}^{n} X_k^2$$

の確率密度関数は、

$$f(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} x^{(n/2)-1} e^{-(x/2)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

統計的に重要な上側 α 点 $\chi(n;\alpha)$ は

$P(\chi^2 \geq \chi(n;\alpha)) = \alpha$

となるように定義され、教科書では付表４としてまとめられている。

t-分布

X は標準正規分布に従い、Y は自由度 n の χ^2 分布に従うとき

$$T = \frac{X}{\sqrt{Y/n}}$$

の確率密度関数は、

$$f(t) = \frac{1}{\sqrt{n}B(\frac{n}{2}, \frac{1}{2})}\left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}$$

統計的に重要な上側 α 点 $t(n;\alpha)$ は

$$P(T \geq t(n;\alpha)) = \alpha$$

となるように定義され、教科書では付表 3 としてまとめられている。

多次元分布

問題

X, Y は独立同分布で標準正規分布 $N(0,1)$ に従うものとする。このとき
$P(-0.01 \le X \le 1.35, -2.51 \le Y \le 0.xx)$
を求めよ。ただし数式中最後の xx には学籍番号下 2 桁を代入するように。
(電卓等使用可、付表を使うように)