確率分布

高校 3 年 4 組 3 番　石関　海也

1　はじめに

本日は灘校文化祭に、及び、数学研究部にお越しいただきありがとうございます。この記事では数学ではマイナー（？）な分野である確率・統計の確率分布について書きたいと思います。

2　確率空間と分布

2.1　確率と確率空間の定義

Ω を全事象を表す集合とし、F を Ω の σ-加法的集合族[*1] とする。このとき、F 上で定義され、下の 3 条件を満たす実数値関数 P を**確率**、$P(A)$ を**事象 A の確率**という。

1. $0 \leq P(A) \leq 1 \quad (A \in F)$
2. $P(\Omega) = 1$
3. $A_1, A_2, \cdots, A_n, \cdots \in F$ が互いに共通部分を持たない（つまり、任意の $i \leq j\ (i, j \in \mathbb{N})$ で、$A_i \cap A_j = \phi$）とき、

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$$

そして、Ω, F, P を 3 つ組み合わせた (Ω, F, P) を確率空間という。

[*1] F が Ω の σ-加法的集合族であるとき、下の 3 条件を満たす。

1. $\Omega \in F$
2. $A \in F \Rightarrow A^c \in F$
3. $A_1, A_2, \cdots, A_n, \cdots \in F \Rightarrow \displaystyle\bigcup_{i=1}^{\infty} A_i \in F$

2.2　確率分布

例 1. （確率空間の例）

$\Omega : \mathbb{Z}$

$F : \Omega$の部分集合の全体

として、$p(n)$ を $\mathbb{Z}$ 上で定義された、正の値をとる関数で

$$\sum_{n=-\infty}^{\infty} p(n) = 1$$

となるとする。

このとき $P(A) = \sum_{n \in A} p(n)$ とすると (Ω, F, P) は確率空間である。

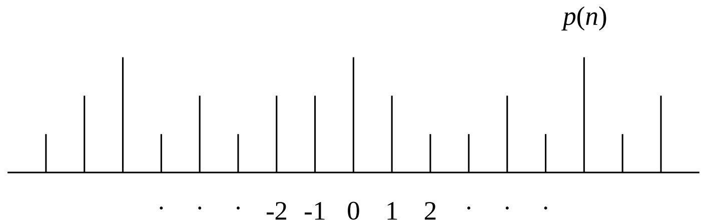

図 1

例 2. 確率空間の例

$\Omega : \mathbb{R}$（実数直線）

$F :$ 一次元ボレル集合族[*2]

として、$\rho(x)$ を積分可能な正の値をとる関数で

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

となるとする。このとき $P(A) = \int_A \rho(x) dx$ とおくと、(Ω, F, P) は *1* つの確率空間である。

例 1,2 で Ω, F, P が本当に確率空間であることを示すのは簡単なので、省略する。

このとき、例 1 における図 1 のような分布を**離散分布**、例 2 における図 2 のような分布を**連続分布**という。確率分布が離散分布なのは、Ω が高々加算な濃度をもつ集合、例えば、$\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ あるいはそれらの部分集合であるときである。

[*2] n 次元ボレル集合族とは $\mathbb{R}^n$ 上の全ての開集合（閉集合としても同じ、[*1]の 2. を参照）を含む σ-加法族で最小のものである。

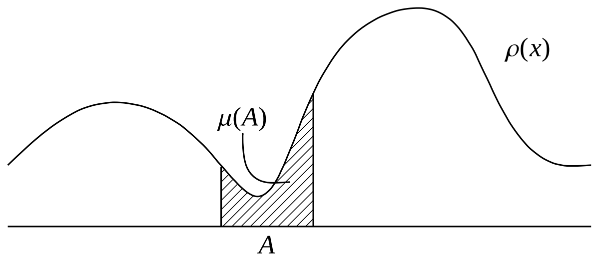

図 2

　また、これらは n 次元拡張して同じことがいえるが、1 変数分布について扱うので、ここでは詳しく述べない。

3　1 変数分布での基本事項

3.1　平均の定義

$$\mu = \begin{cases} \displaystyle\int_{\Omega} x\rho(x)dx & （連続） \\ \displaystyle\sum_{x \in \Omega} xp(x) & （離散） \end{cases}$$

となる μ をこの分布の平均と呼ぶ。

3.2　分散と標準偏差の定義

$$\sigma^2 = \begin{cases} \displaystyle\int_{\Omega} (x-\mu)^2 \rho(x)dx \\ \displaystyle\sum_{x \in \Omega} (x-\mu)^2 p(x) \end{cases}$$

となる σ^2 を分散と呼び、σ を標準偏差と呼ぶ。（ただし $\sigma \geq 0$）

　分散が大きいとき、分布のばらつきが大きく、分散が小さいとき、分布のばらつきが小さいといえる。

3.3　平均と分散

(1)

$$\sigma_a^2 = \begin{cases} \displaystyle\int_\Omega (x-a)^2 \rho(x) dx \\ \displaystyle\sum_{x\in\Omega} (x-a)^2 p(x) \end{cases}$$

が存在するとき、任意の a で $\sigma^2 \leq \sigma_a^2$ である。

（証明）

（連続のとき）

$$\begin{aligned} \sigma_a^2 - \sigma^2 &= \int_\Omega (x-a)^2 \rho(x) dx - \int_\Omega (x-\mu)^2 \rho(x) dx \\ &= \int_\Omega \left((x-a)^2 - (x-\mu)^2\right) \rho(x) dx \\ &= \int_\Omega \left(2(\mu-a)x + a^2 - \mu^2\right) \rho(x) dx \\ &= 2(\mu-a) \int_\Omega x\rho(x) dx + (a^2-\mu^2) \int_\Omega \rho(x) dx \\ &= 2(\mu-a)\mu + (a^2-\mu^2)\cdot 1 \\ &= a^2 - 2a\mu + \mu^2 \\ &= (a-\mu)^2 \geq 0 \end{aligned}$$

（離散のとき）

同様にして示せるので省略。 以下も、連続のときのみを示す。 □

この σ_a^2 は a のまわりの分散といったりもするが、平均 μ は、そのまわりの分散が最小であるという特徴を持っているといえる。

(2) 平均が μ で、分散が σ^2 となる分布で

$$\sigma^2 = \begin{cases} \displaystyle\int_\Omega x^2 \rho(x) dx - \mu^2 \\ \displaystyle\sum_{x\in\Omega} x^2 p(x) - \mu^2 \end{cases}$$

（証明）

$$
\begin{aligned}
\sigma^2 &= \int_\Omega (x-\mu)^2 \rho(x) dx - \mu^2 \\
&= \int_\Omega x^2 \rho(x) dx - 2\mu \int_\Omega x\rho(x) dx + \mu^2 \int_\Omega \rho(x) dx \\
&= \int_\Omega x^2 \rho(x) dx - 2\mu^2 + \mu^2 \\
&= \int_\Omega x^2 \rho(x) dx - \mu^2
\end{aligned}
$$

□

3.4　チェビシェフの不等式

標準偏差 σ が存在する任意の分布で

$$
P\left(\left|\frac{x-\mu}{\sigma}\right| \geq k\right) \leq \frac{1}{k^2}
$$

（証明）

$$
\begin{aligned}
\sigma^2 &= \int_\Omega (x-\mu)^2 \rho(x) dx \\
&= \int_{|x-\mu|\geq k\sigma} (x-\mu)^2 \rho(x) dx + \int_{|x-\mu|<k\sigma} (x-\mu)^2 \rho(x) dx \\
&\geq \int_{|x-\mu|\geq k\sigma} k^2\sigma^2 \rho(x) dx + \int_{|x-\mu|<k\sigma} (x-\mu)^2 \rho(x) dx \\
&= k^2\sigma^2 \rho\left(\left|\frac{x-\mu}{\sigma}\right| \geq k\right) + \int_{|x-\mu|<k\sigma} (x-\mu)^2 \rho(x) dx \\
&\geq k^2\sigma^2 P\left(\left|\frac{x-\mu}{\sigma}\right| \geq k\right)
\end{aligned}
$$

$$
\therefore\ P\left(\left|\frac{x-\mu}{\sigma}\right| \geq k\right) \leq \frac{1}{k^2}
$$

□

この不等式は一見便利そうだが、

$$
\int_{|x-\mu|<k\sigma} (x-\mu)^2 \rho(x) dx \geq 0
$$

などという大雑把な範囲の絞込みをしていて、極限における法則（例えば大数の法則）について考えるとき以外にはあまり使われない。

4　様々な分布

4.1　二項分布

一回の試行である事象が起こる確率を p とする。（p は $0 < p < 1$ を満たす定数）このとき、この試行を n 回繰り返して、その事象が r 回起こる確率

$$ {}_nC_r p^r q^{n-r} \quad (q = 1 - p) $$

を二項分布という。

二項分布の平均 μ は np、分散 σ^2 は npq である。

（証明）

$$ (x + y)^n = \sum_{r=0}^{n} {}_nC_r x^r y^{n-r} $$

で両辺を x について偏微分してから x をかけると

$$ nx(x + y)^{n-1} = \sum_{r=0}^{n} r\,{}_nC_r x^r y^{n-r} \tag{1} $$

ここで

$$ \begin{aligned} \mu &= \sum_{r=0}^{n} r\,{}_nC_r p^r q^{n-r} \\ &= np(p + q)^{n-1} \quad (\because (1) \text{ に } x = p, y = q \text{ を代入}) \\ &= np \quad (\because p + q = 1) \end{aligned} $$

$\square$

(1) をさらに x で偏微分して、両辺に x をかけると、

$$ n(n-1)x^2(x + y)^{n-2} + nx(x + y)^{n-1} = \sum_{r=0}^{n} r^2\,{}_nC_r x^r y^{n-r} \tag{2} $$

$$ \begin{aligned} \therefore\ \sigma^2 &= \sum_{r=0}^{n} r^2\,{}_nC_r p^r q^{n-r} - \mu^2 \\ &= n(n-1)p^2 + np - n^2p^2 \quad (\because (2), p + q = 1) \\ &= np(1 - p) \\ &= npq \end{aligned} $$

□

二項分布では他にも、
$n \to \infty$ のとき、$t = \frac{r}{n}$ の分布 $p_n(t)$ は

$$\delta(x) = \begin{cases} 0 & (t \neq p) \\ \infty & (t = p) \end{cases}$$

$$\int_{-\infty}^{\infty} \delta(x)dx = 1\,,\ \int_{-\infty}^{\infty} \delta(x)f(x)dx = f(p)$$

となるような $\delta(x)$ のグラフに近づくことが知られていたり、$\lambda = np$ とおいたとき、λ を一定として $n \to \infty$ にすると、ポアソン分布（後述）に近づくことなどが知られている。

4.2　ポアソン分布

$$P_\lambda(r) = \frac{e^{-\lambda}\lambda^r}{r!}$$

で定義される O 及び自然数上の分布を**ポアソン分布**という。

まず、二項分布で $\lambda = np$ とおいて λ を一定としたとき、$n \to \infty$ とすればポアソン分布 $P_\lambda(r)$ に近づくことを示す。

（証明）

$$\begin{aligned} {}_nC_r p^r q^{n-r} &= \frac{n!}{r!(n-r)!}\left(\frac{\lambda}{n}\right)^r \left(1 - \frac{\lambda}{n}\right)^{n-r} \\ &= \frac{\lambda^r}{r!} \cdot \frac{n}{n} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdot \ \cdots \ \cdot \frac{n-r+1}{n} \cdot \left(1 - \frac{\lambda}{n}\right)^{-r} \cdot \left(1 - \frac{\lambda}{n}\right)^{\left(-\frac{n}{\lambda}\right)\cdot(-\lambda)} \\ &\to \frac{\lambda^r}{r!} \cdot 1 \cdot 1 \cdot \ \cdots \ \cdot 1 \cdot 1 \cdot e^{-\lambda} \\ &= \frac{\lambda^r e^{-\lambda}}{r!} = P_\lambda(r) \end{aligned}$$

□

また、ポアソン分布 $P_\lambda(r)$ の平均 μ は λ、分散 σ^2 は λ である。

（証明）

$$e^x = \sum_{r=0}^{\infty} \frac{x^r}{r!} \quad \text{（マクローリン展開）}$$

を x について微分してから x をかけると、

$$xe^x = \sum_{r=0}^{\infty} \frac{rx^r}{r!} \tag{3}$$

$$\begin{aligned}\therefore\ \mu &= \sum_{r=0}^{\infty} rP_\lambda(r) \\ &= \sum_{r=0}^{\infty} \frac{r\lambda^r e^{-r}}{r!} \\ &= \lambda e^\lambda \cdot e^{-\lambda} \quad (\because (3)) \\ &= \lambda\end{aligned}$$

(3) を再び x について微分してから x をかけると、

$$x(x+1)e^x = \sum_{r=0}^{\infty} \frac{\lambda^2 r^x}{r!} \tag{4}$$

$$\begin{aligned}\therefore\ \sigma^2 &= \sum_{r=0}^{\infty} r^2 P_\lambda(r) - \mu^2 \\ &= \sum_{r=0}^{\infty} \frac{r^2\lambda^r e^{-r}}{r!} - \lambda^2 \\ &= \lambda\cdot(\lambda+1)e^\lambda \cdot e^{-\lambda} - \lambda^2 \quad (\because (4)) \\ &= \lambda\end{aligned}$$

□

4.3 幾何分布

1 回の試行である事象が起こる確率を p（p は $0 < p < 1$ を満たす定数）として r 回目に初めてその事象が起こる確率

$$q^{r-1}p \quad (q = 1 - p)$$

が表す分布を幾何分布と呼ぶ。

幾何分布の平均 μ は $\dfrac{1}{p}$、分散 σ^2 は $\dfrac{q}{p^2}$ である。

（証明） $0 < x < 1$ で

$$\sum_{r=0}^{\infty} x^r = \frac{1}{1-x}$$

x について両辺微分して

$$\sum_{r=0}^{\infty} r x^{r-1} = \frac{1}{(1-x)^2} \tag{5}$$

$$\begin{aligned}
\therefore\ \mu &= \sum_{r=0}^{\infty} r q^{r-1} p \\
&= \frac{p}{(1-q)^2} \quad (\because (5)) \\
&= \frac{p}{p^2} \\
&= \frac{1}{p}
\end{aligned}$$

(5) の両辺に x をかけたものを両辺 x について微分して

$$\sum_{r=0}^{\infty} r^2 x^{r-1} = \frac{1+x}{(1-x)^3}$$

$$\begin{aligned}
\therefore\ \mu &= \sum_{r=0}^{\infty} r^2 q^{r-1} p - \mu^2 \\
&= \frac{1+q}{(1-q)^3} \cdot p - \frac{1}{p^2} \\
&= \frac{1+q}{p^2} - \frac{1}{p^2} \\
&= \frac{q}{p^2}
\end{aligned}$$

□

4.4 正規分布

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

で定義される分布を平均 μ、分散 σ^2 の正規分布といい、$N(\mu, \sigma^2)$ で表す。特に $N(0,1)$ となるものを**標準正規分布**という。この分布の平均は μ、分布は σ^2 である。

（証明）

まず、上記の $f(x)$ で定義される分布が確率分布であることを示す。

$$I = \int_0^{\infty} e^{-\frac{x^2}{2}} dx$$

とおく。$D = \{(x,y) \mid x \geq 0, y \geq 0\}$ で

$$\begin{aligned}
&\iint_D e^{-\frac{x^2+y^2}{2}} dxdy \\
&= \int_0^\infty e^{-\frac{x^2}{2}} dx \int_0^\infty e^{-\frac{y^2}{2}} dy \\
&= I^2
\end{aligned}$$

ここで $x = r\cos\theta, y = r\sin\theta$ とおく（極座標変換）と、D に対応する (r,θ) の領域 Ω は $\Omega = \left\{(r.\theta) \mid 0 \leq r, 0 \leq \theta \leq \dfrac{\pi}{2}\right\}$

　また、

$$\begin{aligned}
dxdy &= \begin{vmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} \\ \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} \end{vmatrix} drd\theta \\
&= \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} d\theta \\
&= rdrd\theta
\end{aligned}$$

$$\begin{aligned}
\therefore \iint_D e^{-\frac{x^2+y^2}{2}} dxdy &= \iint_\Omega re^{-\frac{r^2}{2}} drd\theta \\
&= \int_0^{\frac{\pi}{2}} d\theta \int_0^\infty re^{-\frac{r^2}{2}} dr \\
&= \int_0^{\frac{\pi}{2}} d\theta \left[-e^{-\frac{r^2}{2}}\right]_0^\infty \\
&= \int_0^{\frac{\pi}{2}} d\theta = \frac{\pi}{2} \\
\therefore I &= \sqrt{\frac{\pi}{2}} \\
\therefore \int_{-\infty}^\infty f(x)dx &= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-\frac{t^2}{2}} dt \quad \left(t = \frac{x-\mu}{\sigma}\right) \\
&= \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^0 e^{-\frac{t^2}{2}} dt + \int_0^\infty e^{-\frac{t^2}{2}} dt\right) \\
&= \frac{2I}{\sqrt{2\pi}} = 1
\end{aligned}$$

□

以下、平均と分散をそれぞれ求める。

$$
\begin{aligned}
\text{（平均）} &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \left(\int_{-\infty}^{\infty} \sigma t e^{-\frac{t^2}{2}} dt + \int_{-\infty}^{\infty} \mu e^{-\frac{t^2}{2}} dt \right) \\
&= \frac{1}{\sqrt{2\pi}} \left(\sigma \left[-e^{-\frac{t^2}{2}} \right]_{-\infty}^{\infty} + \mu \cdot 2I \right) \\
&= \frac{1}{\sqrt{2\pi}} (0 + \sqrt{2\pi} \cdot \mu) \\
&= \mu \\
\text{（分散）} &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx \\
&= \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 t^2 e^{-\frac{t^2}{2}} dt \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \left(\left[-t e^{-\frac{t^2}{2}} \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} dt \right) \\
&= \frac{\sigma^2}{\sqrt{2\pi}} (0 + \sqrt{2\pi}) \\
&= \sigma^2
\end{aligned}
$$

□

この分布は、確率分布の中でも最も重要な分布だが、それは次の中心極限定理によるものである。

定理 4.1 (中心極限定理)**.** ある母集団が平均 μ、標準偏差 σ のある分布に従うならば、大きさ n の標本を無作為にとりだしたときの標本の平均の分布は、$n \to \infty$ のとき、平均 μ、標準偏差 $\frac{\sigma}{\sqrt{n}}$ の正規分布に近づく。

正規分布は本来連続分布であるが、離散分布においても正規分布が近似的に用いられることがある。また、統計的推測や仮説検定、区間推定など、様々な場面で利用される。

5　あとがき

僕はもともと、セルオートマトン、パーコレーションや、コンタクト・プロセスにおける臨界確率について書こうと思っていましたが、準備期間の短さがたたって、自分の仮説を示すことができなかったりして、諦めることにしました。結局、もともとある程度知識はかじっていた確率統計について書くことにしましたが、それも思ったようにいきませんでした。こんなことならあきらめずに書けばよかったなぁと思いまして、一応名前だけ紹介しました。非常に面白い分野であることは保証します。

また、この記事では他の分布やランダム・ウォークなども扱いたかったんですが、時間の制約などもあり、今回は書くことができませんでした。最後の部誌だったので、非常に残念な思いです。

というわけで、ここに書いてある内容は、本当に確率統計の序の序です。しかし、これだけでも、この分野に興味を抱いた人がいれば幸いです。

ここまで、拙い文章を読んでいただきありがとうございました。

参考文献

[1]「確率・統計入門」森　真・藤田岳彦著