統計数理

石川顕一

http://ishiken.free.fr/lecture.html

10/18　組み合わせと確率
10/25　確率変数と確率分布
11/1　代表的な確率分布
11/8（前半）　ランダムウォークと破産問題
11/8（後半）　ブラウン運動と拡散
11/22　ノイズの理論

11/15は演習のみ

参考書

[1] 理工系の数学入門コース7・薩摩順吉著
「確率・統計」岩波書店 2001年

[2] コルモゴロフ、ジュルベンコ、プロホロフ 共著
丸山 哲郎、馬場 良和 共訳
「コルモゴロフの確率論入門」森北出版 2003年

[3] 岩波基礎物理シリーズ8・北原和夫著
「非平衡系の統計力学」岩波書店 1997年

統計数理

石川顕一

10/25　確率変数と確率分布

- 確率変数
- 期待値と分散
- 多変数の確率分布
- 大数の法則
- 変数変換

2—1 確率変数

- 確率をより数学的に取り扱うために、各事象に適当な数値を与える。
  - サイコロの目：1〜6の整数
  - システム創成学科のコース：A→1, B→2, C→3, D→4
  - システム創成学科のコース：A→−1, B→20, C→33, D→400 (可能だが扱いを面倒にするだけ)
  - 雨が降る→1、雨が降らない→0
  - 気体分子の運動量→運動量の値

偶然に支配され、様々な値を取る。

確率変数(random variable or stochastic variable) → 参考書[1] p.36

標本空間の中の根元事象に対して適当な数値を対応させた変数 X を考えて、その変数がどの数値をとるかは偶然に支配されるけれども、X がある特定の数値 x をとる確率、すなわちある根元事象の起こる確率が定まっているとき、X を確率変数という。

確率変数を決定するには
- それがとる値のリスト
- その値をどういう確率でとるか

離散的な場合（とびとびの値しかとらない）：確率密度

→ 参考書[1] p.37

- 確率変数Xが離散的で有限個の値をとる場合を考える。Xがx_i (i=1,2,・・,n)となる確率を

  $P(X=x_i) = p_i$

  と表す。
  - ［例］サイコロ投げの場合、$p_1=p_2=\cdots=p_6=1/6$

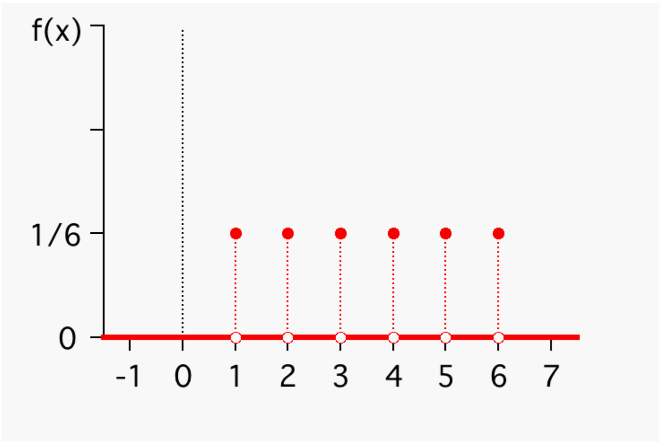
確率は関数の形で、

$$f(x)=\begin{cases} p_i & (\text{if} \quad x=x_i) \\ 0 & (\text{otherwise}) \end{cases}$$

確率密度(probability density)

と書くことができる。離散的な場合には、確率関数ということもある。

$$\sum_{i=1}^{n} f(x_i) = f(x_1) + f(x_2) + \cdots + f(x_n) = 1$$



離散的な場合：分布関数

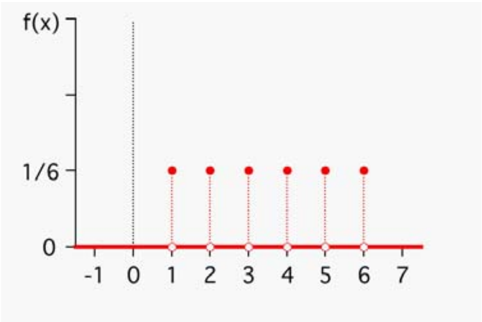
→ 参考書[1] p.37

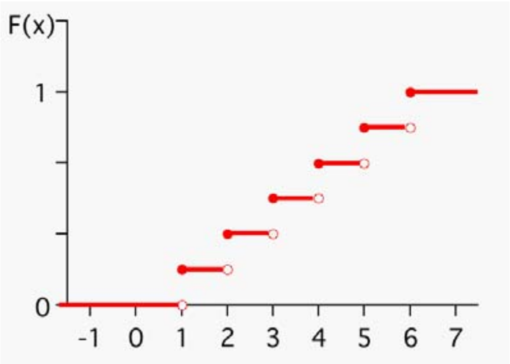
- 確率変数Xのとる値がx以下である確率というものを考え、それに対して、

$$F(x) = P(X \le x)$$

という関数を考えることができる。これを分布関数と呼ぶ。

  – [例] サイコロ投げの場合、F(2)=1/3, F(2.5)=1/3





$$F(x) = \sum_{x_i \le x} f(x)$$

$F(x)$はxについて減少することのない階段状の関数であり、

$$F(\infty) = 1,\ F(-\infty) = 0$$

$$P(\alpha < X \le \beta) = F(\beta) - F(\alpha) = \sum_{\alpha < x_i \le \beta} f(x_i)$$

→ 参考書[1] p.38

連続的な場合:確率密度、分布関数

→ 参考書[1] p.38

- 確率変数Xがxとx+Δxの間にある確率が

$$P(x < X \le x+\Delta x) = \int_x^{x+\Delta x} f(y)dy \cong f(x)\Delta x$$

となるような関数$f(x)$が**確率密度**である。

注意!
常に、確率変数Xの範囲に対する確率を考える。ある特定の値になる確率はゼロ!

分布関数 $F(x) = \int_{-\infty}^{x} f(y)dy$

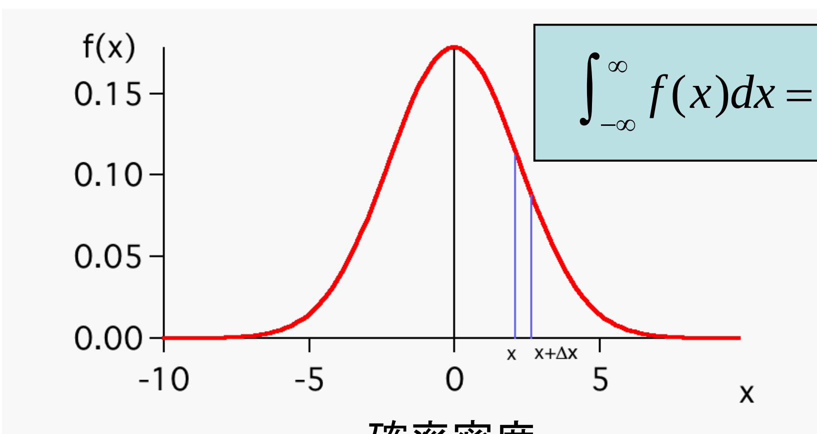
⇨

単調増加関数 $F(\infty)=1,\ F(-\infty)=0$

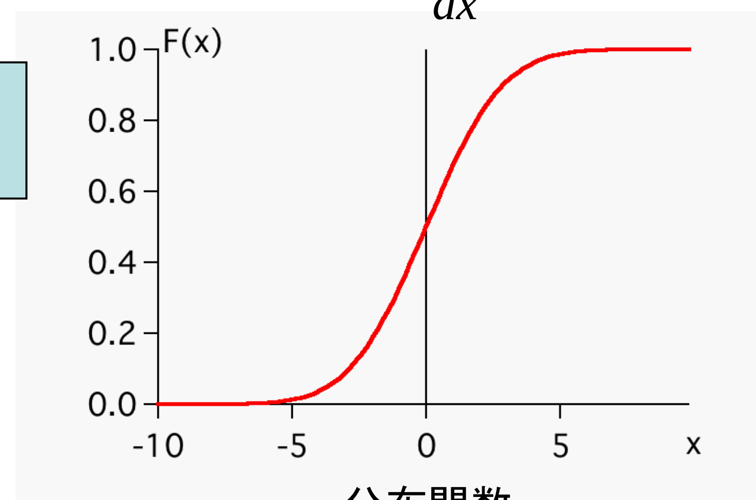
$$P(\alpha < X \le \beta) = F(\beta) - F(\alpha) = \int_\alpha^\beta f(x)dx$$

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

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



確率密度



分布関数

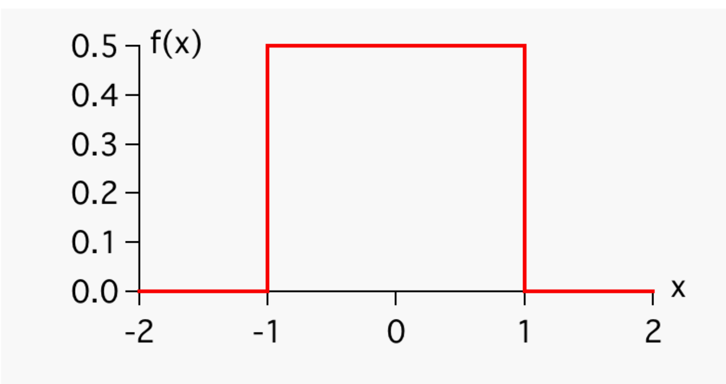
連続的な場合

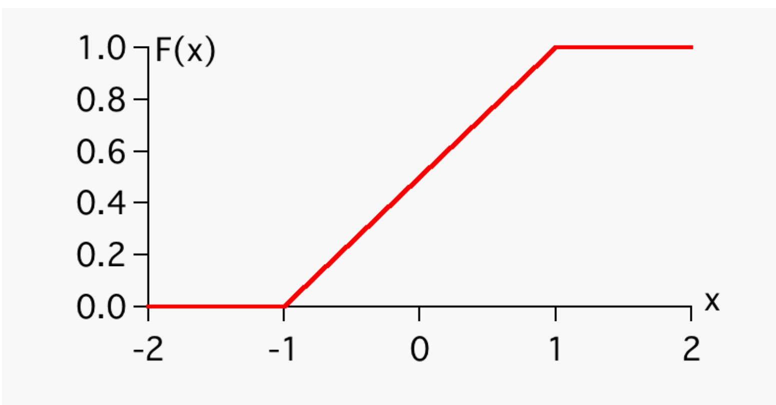
[例] 関数 $f(x)=\begin{cases} c & (|x|\le 1) \\ 0 & (|x|>1) \end{cases}$ （一様分布）　　→ 参考書[1] p.40

が確率密度となるようにcの値を定め、分布関数$F(x)$を求めよ。

$$\int_{-\infty}^{\infty} f(x)dx = 2c = 1 \Rightarrow c = \frac{1}{2}$$

$$F(x) = \int_{-\infty}^{x} f(y)dy = \begin{cases} 0 & (x<-1) \\ (x+1)/2 & (-1\le x\le 1) \\ 1 & (x>1) \end{cases}$$





2−2　期待値と分散

→ 参考書[1]

確率変数を完全に特徴づけるもの　→　確率密度（確率分布）

- 確率密度の詳細は分からないことも多い。
- 確率密度の詳細は重要でないことも多い。

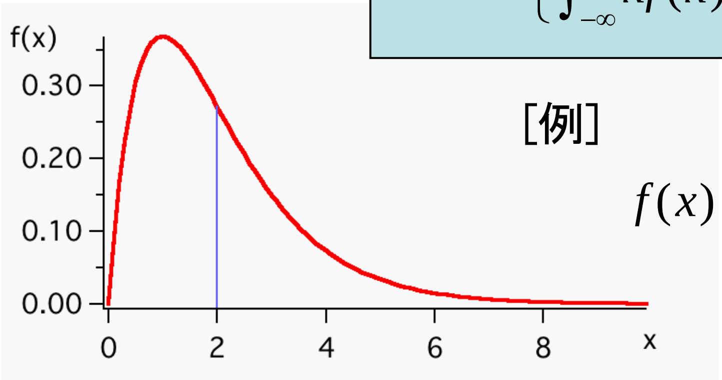
指標が必要

最重要

期待値(expectation value)または平均(mean, average)

$$\mu = \begin{cases} \sum_{i=1}^{n} x_i f(x_i) & \text{（離散的な場合）} \\ \int_{-\infty}^{\infty} x f(x) dx & \text{（連続的な場合）} \end{cases}$$

[例]



$$f(x) = \begin{cases} xe^{-x} & (x \geq 0) \\ 0 & (x < 0) \end{cases} \quad \Rightarrow \quad \mu = \int_0^{\infty} x^2 e^{-x} dx = 2$$

→ 参考書[1] p.43

2−2 期待値と分散

分散(variance) 確率変数Xがだいたいどれぐらいの範囲に分布しているのか

$$V(X) \equiv \sigma^2 \equiv \begin{cases} \displaystyle\sum_{i=1}^{n} (x_i - \mu)^2 f(x_i) & \text{（離散的な場合）} \\ \displaystyle\int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx & \text{（連続的な場合）} \end{cases}$$

→ 参考書[1] p.44

標準偏差(standard deviation)・・分散の平方根、すなわちσ
分布が平均からどれぐらいの幅にあるのか

[例] サイコロ

$$\mu = 3.5 \qquad \sigma^2 = \frac{8.75}{3} = 2.92 \qquad \sigma = \sqrt{2.92} = 1.71$$

[例]

$$f(x) = \begin{cases} xe^{-x} & (x \geq 0) \\ 0 & (x < 0) \end{cases} \Rightarrow \sigma^2 = \int_0^{\infty} (x-2)^2 xe^{-x} dx = 2 \qquad \sigma = \sqrt{2}$$

→ 参考書[1] p.45

標準偏差大・・分布のばらつきの程度が大きい

標準偏差小・・分布はほとんど平均値のまわりに集中

東京大学工学部 システム創成学科 C コース Program for Biological & Information Systems

モーメント

→ 参考書[1] p.48

確率変数Xの関数 $\varphi(X)$ の期待値

$$E[\varphi(X)]=\begin{cases}\displaystyle\sum_{i=1}^{n}\varphi(x_i)f(x_i)\\ \displaystyle\int_{-\infty}^{\infty}\varphi(x)f(x)dx\end{cases}$$

特に

k 次のモーメント

$$E[X^k]=\begin{cases}\displaystyle\sum_{i=1}^{n}(x_i)^k f(x_i)\\ \displaystyle\int_{-\infty}^{\infty}x^k f(x)dx\end{cases}$$

$$E[X^0]=1$$

$$E[X^1]=\mu$$

$$\begin{aligned}\sigma^2 &= E[(X-\mu)^2] \\ &= E[X^2]-2\mu E[X]+\mu^2 \\ &= E[X^2]-\mu^2\end{aligned}$$

⇦ 平均のまわりの2次のモーメント

ひずみ度　$\gamma=E[(X-\mu)^3]$

対称な分布なら $\gamma=0$

2−3　多変数の確率分布

→ 参考書[1]

- 同時確率分布→同時にいくつかの試行を行ったときの確率分布
  - [例] 2個のサイコロを振って、それぞれのサイコロの目を確率変数X, Yとし、$2<X<5$かつ$3<Y<5$となる事象やその確率を考える。（2次元確率分布）
- 離散的な場合
  - Xのとる値：$x_1, x_2, \cdot\cdot, x_m$
  - Yのとる値：$y_1, y_2, \cdot\cdot, y_n$
  - Xがx_i $(i = 1, 2,\cdot\cdot, m)$の値をとり、かつYがy_j $(j = 1, 2,\cdot\cdot, n)$の値をとる確率をp_{ij}と書くことにする。

$$P(X=x_i, Y=y_j) = p_{ij}$$

確率密度　$$f(x,y)=\begin{cases} p_{ij} & (x=x_i, y=y_j) \\ 0 & \text{otherwise} \end{cases}$$

分布関数　$$F(x,y)=P(X\le x, Y\le y)=\sum_{x_i\le x}\sum_{y_j\le y} f(x_i,y_j)$$

2つのサイコロの場合

$$f(i,j)=\frac{1}{36}\quad(1\le i,j\le 6)$$

$$F(2,3)=\frac{1}{6}$$

$$\Rightarrow\quad \sum_{i=1}^{m}\sum_{j=1}^{n} f(x_i,y_j)=1$$

2−3　多変数の確率分布

- 同時確率分布（連続的な場合）　　→ 参考書[1] p.56

確率密度　$P(x < X \le x+\Delta x, y < Y \le y+\Delta y) = \int_x^{x+\Delta x} dx' \int_y^{y+\Delta y} dy' f(x',y')$

分布関数　$F(x,y) = P(X \le x, Y \le y) = \int_{-\infty}^{x} dx' \int_{-\infty}^{y} dy' f(x',y')$

$$F(\infty,\infty) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy f(x,y) = 1$$

- 周辺確率分布（離散的な場合）　　→ 参考書[1] p.57
  - ２次元確率分布で、例えば、Y の値にかかわらずXがどういう分布をしているか知りたいときは、それぞれのXにおけるYの値をすべて足しあわせればよい。

$$f_1(x) = \begin{cases} p_{i1} + p_{i2} + \mathcal{L} + p_{in} = \sum_{i=1}^{m} p_{ij} & (\text{if } x = x_i) \\ 0 & \text{otherwise} \end{cases}$$

周辺分布関数

$$F_1(x) = P(X \le x) = \sum_{x_1 \le x} f_1(x_i)$$

2−3　多変数の確率分布

- 周辺確率分布（連続的な場合）　　→ 参考書[1] p.58
  - ２次元確率分布で、例えば、Y の値にかかわらずXがどういう分布をしているか知りたいときは、それぞれのXにおけるYの値を積分すればよい。

周辺(marginal)確率密度	$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy$
周辺分布関数	$F_1(x) = \int_{-\infty}^{x} f_1(x')dx'$

2−3 多変数の確率分布

- 条件付き確率分布(連続的な場合)

→ 参考書[1] p.59

$$\frac{P(x<X\le x+\Delta x, y<Y\le y+\Delta y)}{P(y<Y\le y+\Delta y)} = \frac{\int_x^{x+\Delta x} dx' \int_y^{y+\Delta y} dy' f(x',y')}{\int_y^{y+\Delta y} dy' f_2(y')}$$

周辺確率密度

$$\frac{f(x,y)dxdy}{f_2(y)dy} = \frac{f(x,y)}{f_2(y)}dx$$

条件付き確率密度 $f(x\,|\,y) = \dfrac{f(x,y)}{f_2(y)}$

$$\int_{-\infty}^{\infty} f(x\,|\,y)dx = \frac{\int_{-\infty}^{\infty} f(x,y)dx}{f_2(y)} = 1$$

- 確率変数の独立性(統計的独立)

$$f(x,y) = f_1(x) f_2(y) \quad \Longleftrightarrow \quad f(x\,|\,y) = f_1(x)$$

[例] 2個のサイコロを振るときの、それぞれのサイコロの目 X, Y

2−3　多変数の確率分布

- 期待値と分散
  - 1変数の場合と同様に定義できる。

X, Yの関数　$\varphi(X,Y)$

$$E[\varphi(X,Y)] = \begin{cases} \displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n}\varphi(x_i,y_j)f(x_i,y_j) & \text{離散的な場合} \\ \displaystyle\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dy\,\varphi(x,y)f(x,y) & \text{連続的な場合} \end{cases}$$

→ 参考書[1] p.60

平均　$\mu_x = E[X],\ \mu_y = E[Y]$

分散　$\sigma_x^2 = E\left[(X-\mu_x)^2\right],\ \sigma_y^2 = E\left[(Y-\mu_y)^2\right]$

共分散　$\sigma_{xy} = E\left[(X-\mu_x)(Y-\mu_y)\right]$

→ 参考書[1] p.61

相関係数　$\rho_{xy} = \dfrac{\sigma_{xy}}{\sigma_x\sigma_y}$

$-1 \le \rho_{xy} \le 1$

XとYが独立　$\Rightarrow$　$\rho_{xy} = 0$　「XとYは相関がない」

（逆は成り立たない：$\not\Leftarrow$）

$E[\{\lambda(X-\mu_x)+(Y-\mu_y)\}]^2$を考える。

2−3　多変数の確率分布

- 期待値と分散
  - ひとつひとつの確率変数の期待値や分散が分かっている時、それらの確率変数の和の期待値や分散はどうなるであろうか？

確率変数の和の期待値は、それぞれの平均値の和である。

$$E(X+Y)=E(X)+E(Y)$$

→ 参考書[2] p.128

独立な確率変数の積の期待値は、それぞれの期待値の積である。

$$E(XY)=E(X)E(Y)$$

→ 参考書[2] p.130

独立な確率変数の和の分散は、それぞれの分散の和である。

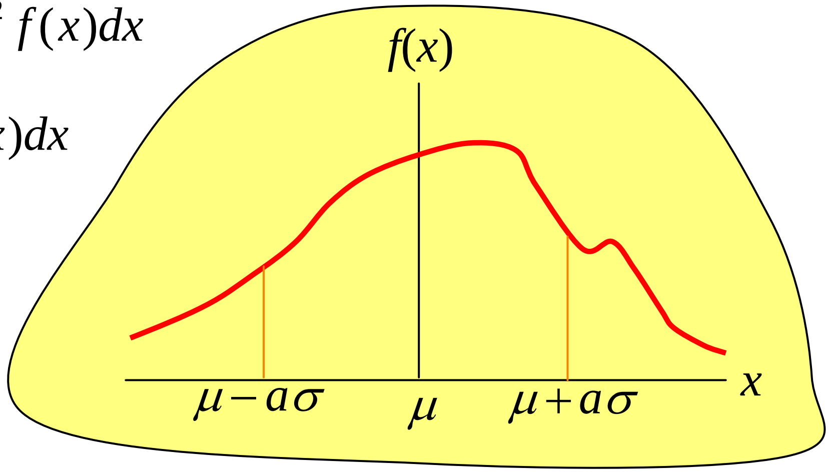
$$V(X+Y)=V(X)+V(Y)$$

→ 参考書[2] p.133

2−4　大数の法則

- チェビシェフ(1821−1894)の不等式　　　→ 参考書[1] p.46
  分散または標準偏差が分布の広がりの度合いを示すという事実を数学的に述べたもの。

$$
\begin{aligned}
\sigma^2(X) &= \int_{-\infty}^{\mu-a\sigma}(x-\mu)^2 f(x)dx + \int_{\mu-a\sigma}^{\mu+a\sigma}(x-\mu)^2 f(x)dx + \int_{\mu+a\sigma}^{\infty}(x-\mu)^2 f(x)dx \\
&\geq \int_{-\infty}^{\mu-a\sigma}(x-\mu)^2 f(x)dx + \int_{\mu+a\sigma}^{\infty}(x-\mu)^2 f(x)dx \\
&\geq \int_{-\infty}^{\mu-a\sigma}(a\sigma)^2 f(x)dx + \int_{\mu+a\sigma}^{\infty}(a\sigma)^2 f(x)dx \\
&= a^2\sigma^2\left(\int_{-\infty}^{\mu-a\sigma} f(x)dx + \int_{\mu+a\sigma}^{\infty} f(x)dx\right) \\
&= a^2\sigma^2 P(|X-\mu| > a\sigma)
\end{aligned}
$$



$$\Rightarrow \quad \boxed{P(|X-\mu| > a\sigma) \leq \frac{1}{a^2} \quad \text{チェビシェフの不等式}}$$

どんな確率分布に対しても成り立つ。

平均値から標準偏差の2倍以上離れている確率は1/4以下。
平均値から標準偏差の3倍以上離れている確率は1/9以下。

2–4 大数の法則

- チェビシェフの不等式
  分散または標準偏差が分布の広がりの度合いを示すという事実を数学的に述べたもの。

$$P(|X-\mu|>a\sigma)\le \frac{1}{a^2}$$ チェビシェフの不等式

[例] 200人の試験で、平均点が60点、標準偏差が6点であった。得点が42点から78点の間にある受験生は何人以上か。 → 参考書[1] p.47

$$200\times\left(1-\frac{1}{3^2}\right)=177.8 \Rightarrow$$ 178人以上

チェビシェフの不等式は次のような形にも書ける。

$$P(|X-\mu|>\alpha)\le \frac{\sigma^2}{\alpha^2}$$ チェビシェフの不等式

2−4 大数の法則

経験的確率を数学的に扱う大切な根拠!大量現象の法則性にかかわる物理学、経済学その他多くの学問分野が、これに基礎をおく。

- 大数の法則(チェビシェフの定理)

→ 参考書[2] p.137

$X_1, X_2, \ldots, X_n$: 互いに独立な確率変数

$\sigma^2(X_k) \le c,\ k = 1,2,3,\mathcal{L}\ ,n$

このとき、任意の正数 α に対して

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \mathcal{L}\ + X_n}{n} - \frac{E(X_1) + \mathcal{L}\ + E(X_n)}{n}\right| < \alpha\right) = 1$$

[証明]

$$Y_n = \frac{X_1 + \mathcal{L}\ + X_n}{n} \quad \Longrightarrow \quad P(|Y_n - E(Y_n)| > \alpha) \le \frac{\sigma^2(Y_n)}{\alpha^2}$$

$$\sigma^2(Y_n) = \frac{\sigma^2(X_1) + \mathcal{L}\ + \sigma^2(X_n)}{n^2} \le \frac{c}{n} \quad \Longrightarrow \quad P(|Y_n - E(Y_n)| < \alpha) \ge 1 - \frac{c}{n\alpha^2}$$

特にすべてのkについて $E(X_k) = a \ \ (k = 1,2,3,\mathcal{L}\ ,n)$ であれば

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}(X_1 + \mathcal{L}\ + X_n) - a\right| < \alpha\right) = 1$$

→ 参考書[2] p.138

2–5 変数変換

→ 参考書[1]

- いろいろな確率分布を扱う際、確率変数を変換したいことが起こる。

確率変数Xの確率密度が$f(X)$の場合に、$Y=\Phi(X)$で新しい確率変数Yを導入したとき、その確率密度$g(Y)$はどうなるか？

確率変数Xの確率密度が$f(X)$の場合に、それと異なる確率密度$g(Y)$を満たす確率変数$Y=\Phi(X)$を導入したい。$\Phi(X)$ はどんな関数？

X で表しても、Y で表しても確率は変わらない。

$$P(x < X < x+\Delta x) = P(y < Y < y+\Delta y) \Rightarrow \int_x^{x+dx} f(x')dx' = \int_y^{y+dy} g(y')dy'$$

$$\Rightarrow f(x)dx = g(y)dy \Rightarrow \boxed{g(y) = f(x)\frac{dx}{dy} = f(x)\left(\frac{dy}{dx}\right)^{-1} = f(x)\left[\Phi'(x)\right]^{-1}}$$

2−5 変数変換

$$g(y) = f(x)\frac{dx}{dy} = f(x)\left(\frac{dy}{dx}\right)^{-1} = f(x)[\Phi'(x)]^{-1}$$

[例] 1次変換 $Y = aX + b$ の場合(a, bは定数)　　→ 参考書[1] p.52

$$x = \frac{y-b}{a},\ \frac{dx}{dy} = \frac{1}{a}$$ だから、 $$g(y) = \frac{1}{a} f\left(\frac{y-b}{a}\right)$$

$$\mu_y = \int_{-\infty}^{\infty} y g(y) dy = \int_{-\infty}^{\infty} (ax+b) f(x) dx = a\mu_x + b$$

$$\sigma_y^2 = \int_{-\infty}^{\infty} (y-\mu_y)^2 g(y) dy = \int_{-\infty}^{\infty} (ax+b-a\mu_x-b)^2 f(x) dx$$

$$= a^2 \int_{-\infty}^{\infty} (x-\mu_x)^2 f(x) dx = a^2 \sigma_x^2$$

2−5　変数変換

$$g(y) = f(x)\frac{dx}{dy} = f(x)\left(\frac{dy}{dx}\right)^{-1} = f(x)[\Phi'(x)]^{-1}$$

[例] Xの確率密度が　$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$　（標準正規分布）　$\mu_x = 0,\ \sigma_x^2 = 1$

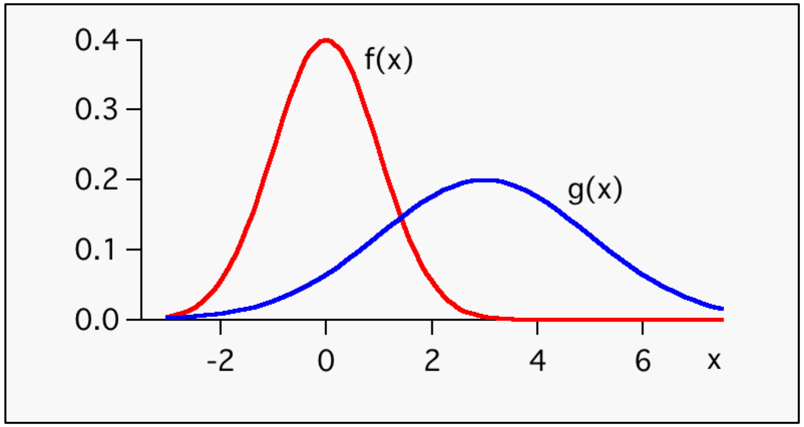
である場合に、Y=2X+3の一次変換をしたときのYの確率密度、平均、分散

→ 参考書[1] p.53

$$g(y) = \frac{1}{2}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{y-3}{2}\right)^2\right]$$

$$\mu_y = 2\mu_x + 3 = 2\cdot 0 + 3 = 3$$

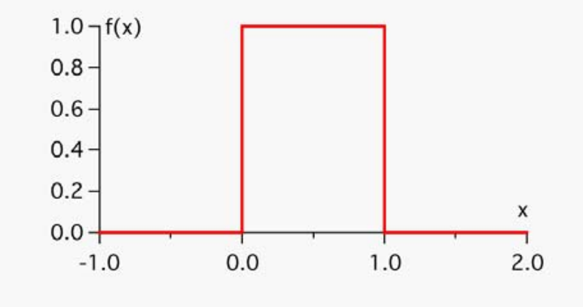
$$\sigma_y^2 = 2^2\sigma_x^2 = 2^2\cdot 1^2 = 4$$

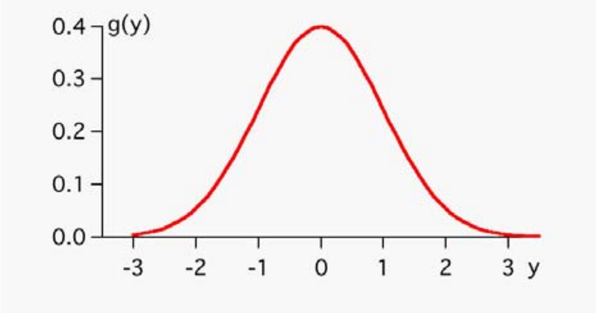


2–5 変数変換

[例] Xの確率密度が $f(x)=\begin{cases}1 & 0\le x\le 1\\ 0 & x<0, x>1\end{cases}$ （一様分布）

である場合に、変換$Y=\Phi(X)$をしてYの確率密度が $g(y)=\dfrac{1}{\sqrt{2\pi}}e^{-y^2/2}$
となるようにしたい。$\Phi(X)$を求めよ。





$$x=\int_{-\infty}^{y}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz \quad\Rightarrow\quad x-\frac{1}{2}=\int_{0}^{y}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz$$

数値的にyを求める。