Mathematical Statistics

Kenichi ISHIKAWA

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

Oct. 18 Combination and probability
Oct. 25 Random variables and probability distributions
Nov. 1 Representative probability distributions
Nov. 8 (First half) Random walk and gambler's ruin problem
(Latter half) Brownian motion and diffusion
Nov. 22 Noise theory

Only exercises are provided on Nov. 15.

Reference books

[1] Satsuma, J. (2001). *"Probability /Statistics"* -- Beginning course of mathematics of science and technology 7, Iwanami Shoten

[2] Kolmogorov, A.N., Žurbenko, I. G., Prokhorov, A.V. (2003). *"Beginning of Kolmogorov's Probability Theory"*. (Trans. Murayama , T. & Baba, Y.): Morikita Shuppan

[3] Kitahara, K. (1997). *"Nonequilibrium Statistical Mechanics" – Iwanami Fundamental Physics Series 8*, Iwanami Shoten

Mathematical Statistics

Kenichi ISHIKAWA

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

Oct. 25 Random variables and probability distributions

- Random variables
- Expectation values and variance
- Multivariate probability distribution
- Law of large numbers
- Variable transformation

2-1 Random Variables

- Each event is given a proper value, in order to handle probabilities more mathematically.
  - Numbers on a die: Integral values from 1 to 6
  - Courses available at the Department of Systems Innovation : A→1, B→2, C→3, D→4
  - Courses available at the Department of Systems Innovation: ＊ eg, A→−1, B→20, C→33, D→400 (Not impossible, but it merely complicates handling.)
  - It rains →1, It does not rain →0
  - Momentum of gaseous molecule → Momentum values

They are assigned various values, being ruled by coincidence.

Random variables (a.k.a. stochastic variable) → Reference book [1], p. 36

Consider the variable X , to which appropriate values correspond for an elementary event in a sample space. Although the value assigned to this variable is ruled by coincidence, this X is defined as the random variable; provided, the probability that X assigned the specific fixed value x (ie, the probability of occurrence of an elementary event).

To fix a random variable, it requires:
- (i) the list of values to be taken by such random variable.
- (ii) The determination of the probability with which such values are taken.

TBL1

TBL1

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

In the discrete case (taking only discrete values): Probability density

→ Reference book [1], p.37

- Consider the case in which the random variable X is discrete, taking a finite quantity of values. The probability that X turns out to be x_i ($i=1,2,\cdot\cdot,n$) is represented as $P(X=x_i) = p_i$.

  – [Example] In the case of die casting:: $p_1 = p_2 = \cdot\cdot = p_6 = 1/6$

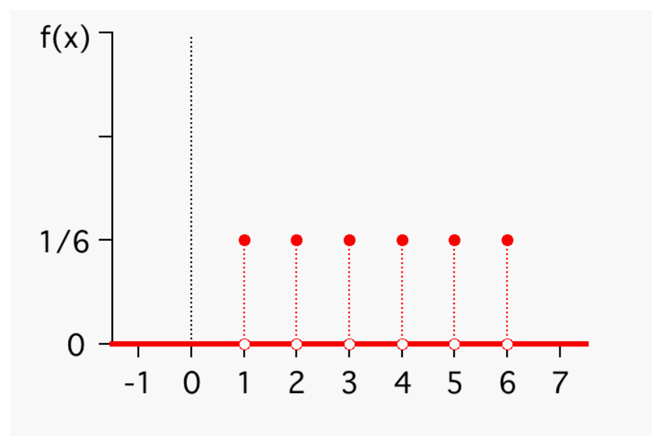
The probability can be written in the form of a function as follows:

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

Probability density

In some discrete cases, it is referred to as the probability function.

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

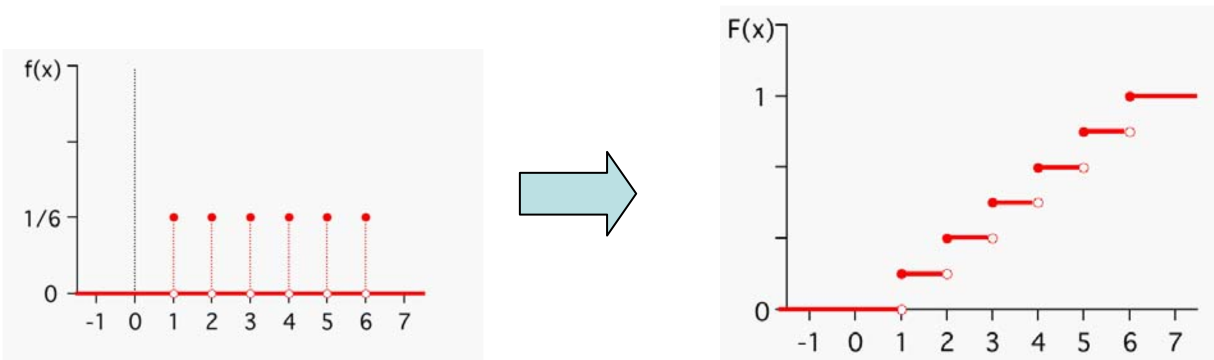


In the discrete case: Distribution Function

→ Reference book [1], p. 37

- First, consider the probability that the values taken by random variable X does not exceed x. For this probability, a function $F(x) = P(X \le x)$ can be considered. This function is referred to as the distribution function.

[Example] In the case of die casting: F(2)=1/3, F(2.5)=1/3



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

Since $F(x)$ is a stair-like function without decreasing x;

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

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

→ Reference book [1], p.38

2-1 Random Variables

In the continuous case: Probability density/distribution function

→ Reference book [1], p.38

- A function f(x) is defined as the probability density, if the probability that the random variable X exists between x and x+Dx assigns:

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

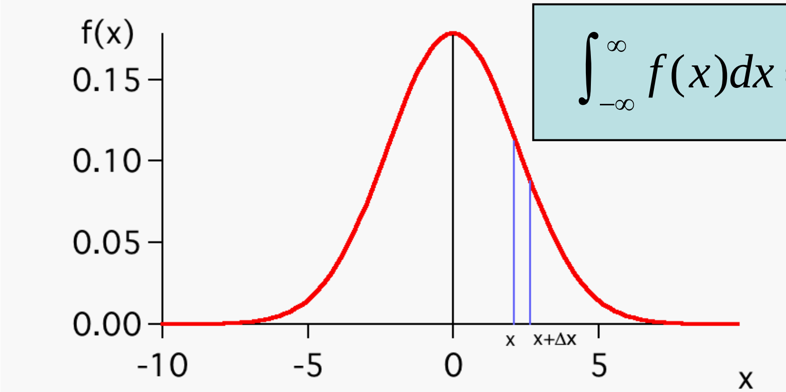
Note: Always try to consider the probability pertaining to the range of the random variable *X*. The probability that such a value will be certain is zero!

Distribution Function $F(x) = \int_{-\infty}^{x} f(y)dy$ ⇒

Monotonically increasing function:

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

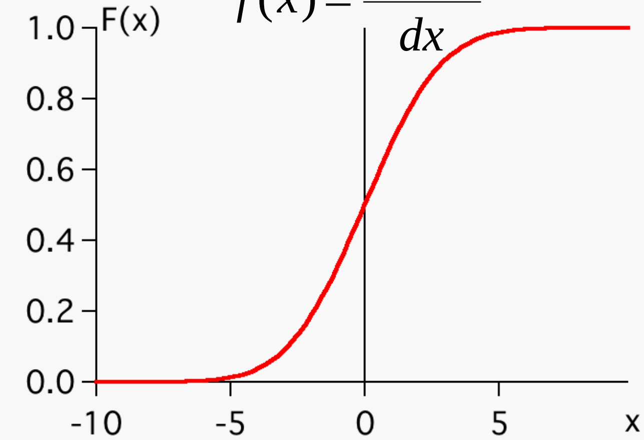
$$P(\alpha < X \le \beta) = F(\beta) - F(\alpha) = \int_{\alpha}^{\beta} f(x)dx$$



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

Probability Density

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



Distribution Function

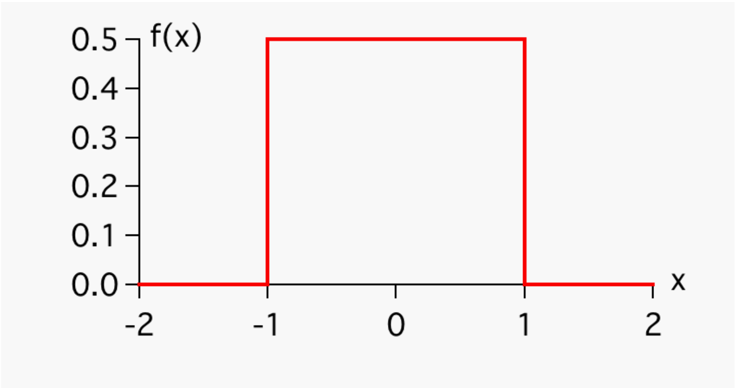
In the continuous case

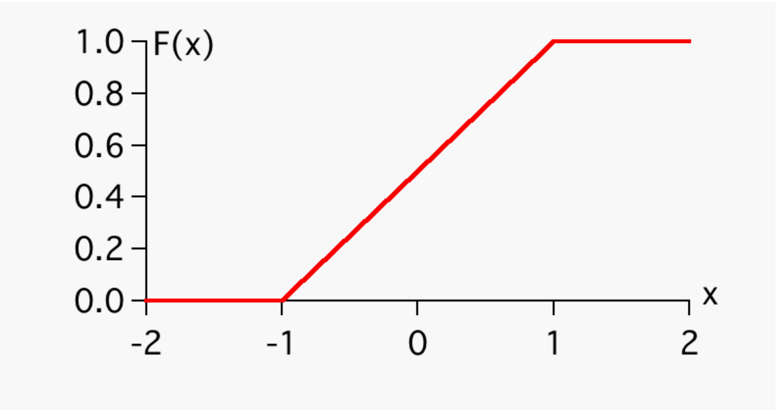
[Example] Find the distribution function $F(x)$ by fixing the value c so that the function $f(x)=\begin{cases} c & (|x|\le 1) \\ 0 & (|x|>1) \end{cases}$ (uniform distribution) can be the probability density.

→ Reference book [1], p.40

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

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





2-2 Expectation values and Variance

→ Reference book [1]

Probability density (probability distributions) → Completely characterizes the random variables

- The probability density includes many unknowns in its detail.
- Much detail of the probability density is inconsequential.

} Some index/indicator will be required.

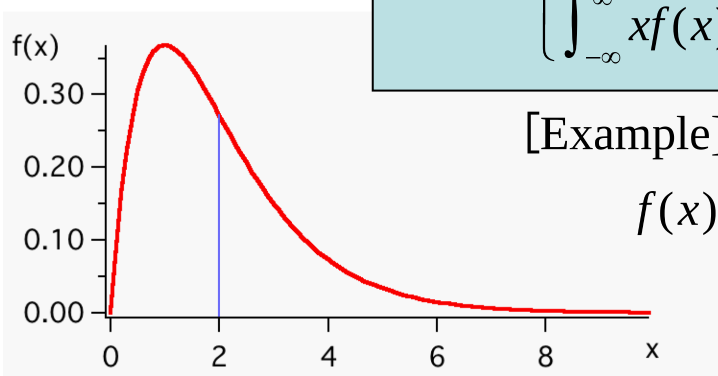
Extremely important

Expectation values or mean/average

$$\mu = \begin{cases} \sum_{i=1}^{n} x_i f(x_i) & \text{(In the discrete case)} \\ \int_{-\infty}^{\infty} x f(x) dx & \text{(In the continuous case)} \end{cases}$$

[Example]

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



→ Reference book [1], p.43

Variance: Indicates where and how widely the random variables X is being distributed.

$$V(X) \equiv \sigma^2 \equiv \begin{cases} \sum_{i=1}^{n} (x_i - \mu)^2 f(x_i) & \text{(In the discrete case)} \\ \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx & \text{(In the continuous case)} \end{cases}$$

→ Reference book [1], p. 44

Standard deviation: Indicates the square root of variance (ie, σ), as well as how the distribution width is being deviated from average.

[Example] Die

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

[Example]

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

→ Reference book [1], p.45

Large standard deviation: Degree of deviation of distribution is high.

Small standard deviation: Distribution is concentrated around average value.

Moment

→ Reference book [1], p. 48

Expectation value of Function $\varphi(X)$ of the random variable X

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

Especially,

k -th moment

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

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

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

$$\sigma^2 = E[(X-\mu)^2]$$ ⇐ Second moment around the average

$$= E[X^2] - 2\mu E[X] + \mu^2$$

$$= E[X^2] - \mu^2$$

Strain Rate $\gamma = E[(X-\mu)^3]$

$\gamma = 0$, if the distribution is symmetric.

2-3 Multivariate probability distribution

→ Reference book [1]

- Simultaneous-probability distributions → Probability distributions shown in simultaneously performing multiple trials.
  - [Example] Cast two dice, letting spots on each die be the random variables X and Y. Then, consider the events of occurrence of $2<X<5$ and $3<Y<5$ and the probability thereof (Two-dimensional probability distribution) .
- In the discrete case
  - Values taken by X: $x_1, x_2, \cdot\cdot, x_m$
  - Values taken by Y: $y_1, y_2, \cdot\cdot, y_n$
  - Let us represent the probability that X takes the values of x_i ($i = 1, 2,\cdot\cdot, m$), and that Y takes the values of y_j ($j = 1, 2,\cdot\cdot, n$) as p_{ij}.

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

Probability density

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

Distribution function

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

In the case of two dice

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

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

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

2-3 Multivariate probability distribution

- Simultaneous probability distributions（In the continuous case） → Reference book [1], p.56

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

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

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

- Marginal probability distributions（in the discrete case） → Reference book [1], p.57
  - To find, for example, how X is being distributed irrespective of the values of Y in the two-dimensional probability distributions, it suffices to add every Y value at each X.

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

Marginal distribution function

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

2-3 Multivariate probability distribution

→ Reference book [1], p.58

- Marginal probability distributions（in the continuous case）
  - To find how *X* is being distributed irrespective of the values of *Y* in the two-dimensional probability distributions, it suffices to perform integration of the *Y* values at each *X*.

Marginal probability density	$f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy$
Marginal distribution function	$F_1(x) = \int_{-\infty}^{x} f_1(x')dx'$

2-3 Multivariate probability distribution

→ Reference book [1] p.59

- Conditional probability distributions (In the continuous case)

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

Marginal probability density

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

Conditional probability density $f(x \mid y) = \dfrac{f(x,y)}{f_2(y)}$

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

- Independence of random variables (Statistical independence)

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

[Example] Spots on each die, X and Y, when casting two dice

2-3 Multivariate probability distribution

- Expectation Values and Variance
  - Definable in the same way as the case of 1-variable

Function of X and Y: $\varphi(X,Y)$

$$E[\varphi(X,Y)]=\begin{cases} \sum_{i=1}^{m}\sum_{j=1}^{n}\varphi(x_i,y_j)f(x_i,y_j) & \text{In the discrete case} \\ \int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dy\,\varphi(x,y)f(x,y) & \text{In the continuous case} \end{cases}$$

Average $\mu_x = E[X],\ \mu_y = E[Y]$ → Reference book [1] p.60

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

Covariance $\sigma_{xy} = E\left[(X-\mu_x)(Y-\mu_y)\right]$ → Reference book [1], p.61

Correlation coefficient $\rho_{xy} = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y}$

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

X and Y are independent. ⇒ $\rho_{xy} = 0$ "X and Y are not correlated with one another"

(the converse, ⇐, is crossed out)

Consider $E[\{\lambda(X-\mu_x)+(Y-\mu_y)\}]^2$.

2-3 Multivariate probability distribution

- Expectation Values and Variance
  - If the expectation value and variance of each individual random variable are known, then, what will the expectation value and variance of the sum of such random variables be?

The expectation value of the sum of the random variables is the sum of each individual average value.

$$E(X+Y)=E(X)+E(Y)$$ → Reference book [2] p.128

The expectation value of the product of independent random variables is the product of each individual expectation value.

$$E(XY)=E(X)E(Y)$$ → Reference book [2], p.130

The variance of the sum of the independent random variables is the sum of each individual variance.

$$V(X+Y)=V(X)+V(Y)$$ → Reference book [2], p.133

2-4 Law of Large Numbers

- Chebyshev's inequality (P.L. Chebyshev, 1821-1894) → Reference book [1], p. 46

This inequality mathematically represents the fact that the variance or the standard deviation shows the degree of distribution expansion.

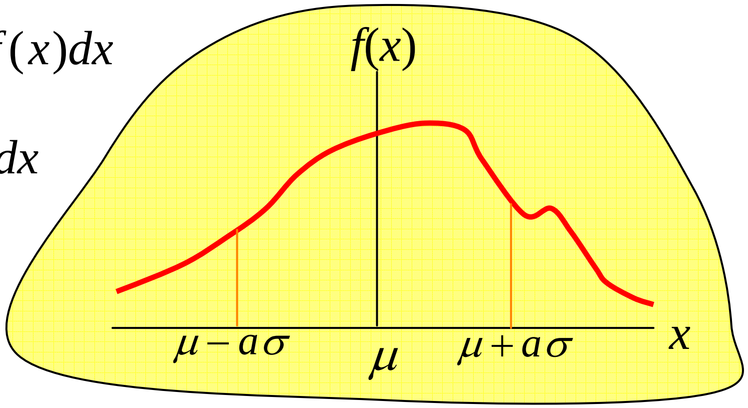
$$\sigma^2(X) = \int_{-\infty}^{\mu-a\sigma}(x-\mu)^2 f(x)dx + \int_{\mu-a\sigma}^{\mu+a\sigma}(x-\mu)^2 f(x)dx + \int_{\mu+a\sigma}^{\infty}(x-\mu)^2 f(x)dx$$

$$\geq \int_{-\infty}^{\mu-a\sigma}(x-\mu)^2 f(x)dx + \int_{\mu+a\sigma}^{\infty}(x-\mu)^2 f(x)dx$$

$$\geq \int_{-\infty}^{\mu-a\sigma}(a\sigma)^2 f(x)dx + \int_{\mu+a\sigma}^{\infty}(a\sigma)^2 f(x)dx$$

$$= a^2\sigma^2\left(\int_{-\infty}^{\mu-a\sigma} f(x)dx + \int_{\mu+a\sigma}^{\infty} f(x)dx\right)$$

$$= a^2\sigma^2 P(|X-\mu| > a\sigma)$$



⇨ $P(|X-\mu| > a\sigma) \leq \dfrac{1}{a^2}$ Chebyshev's inequality

Effected by any probability distribution

-The probability that the values of the random variable X deviate two-times of standard deviation from the average value: Not exceeding 1/4.

-The probability that the values of the random variable X deviate three-times of standard deviation from the average value: Not exceeding 1/9.

- Chebyshev's inequality
  This inequality mathematically represents the fact that the variance or the standard deviation shows the degree of distribution expansion.

$$P(|X-\mu| > a\sigma) \leq \frac{1}{a^2} \quad \text{Chebyshev's inequality}$$

[Example] 200 examinees took an exam. The average mark was 60, and the standard deviation was 6. How many examinees who attained marks ranging from 42 to 78 are there at the minimum?

→ Reference book [1], p. 47

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

⇒ 178 or more examinees

The Chebyshev's inequality can also be written as:

$$P(|X-\mu| > \alpha) \leq \frac{\sigma^2}{\alpha^2} \quad \text{Chebyshev's inequality}$$

2-4 Law of Large Numbers

- Law of large numbers (Chebyshev's Theorem) → Extremely important base for mathematically handling experimental probabilities! Many academic domains involving the regularity of mass phenomenon including physics, economics, and others rely on this Law.

→ Reference book [2], p. 137

$X_1, X_2, \ldots, X_n$: Random variables independent of each other

$\sigma^2(X_k) \le c,\ k = 1, 2, 3, \cdots, n$, where

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

for a given positive α .

[Proof] $Y_n = \dfrac{X_1 + \cdots + X_n}{n} \implies P(|Y_n - E(Y_n)| > \alpha) \le \dfrac{\sigma^2(Y_n)}{\alpha^2}$

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

In particular, if $E(X_k) = a \ (k = 1, 2, 3, \cdots, n)$ for every k, then,

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

→ Reference book [2], p. 138

2-5 Variable Transformation

→ Reference book [1]

- It may be necessary to transform the random variables when handling diverse probability distributions.

If a new random variable *Y* is introduced when the probability density of the random variables *X* is $f(X)$, then, what will the probability density $g(Y)$ thereof be?

When the probability density of the random variable *X* is $f(X)$, we need to introduce the random variables $Y = \Phi(X)$ sufficing the probability density $g(Y)$ which is distinct from the random variable $f(X)$. To this end, what function should $\Phi(X)$ be?

Irrespective of being denoted by *X* or *Y*, the probability does not differ.

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

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

2-5 Variable Transformation

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

[Example] In the case of linear transformation $Y = aX + b$ (a and b are constants.)：

→ Reference book [1], p.52

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

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

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

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

2-5 Variable Transformation

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

[Example]

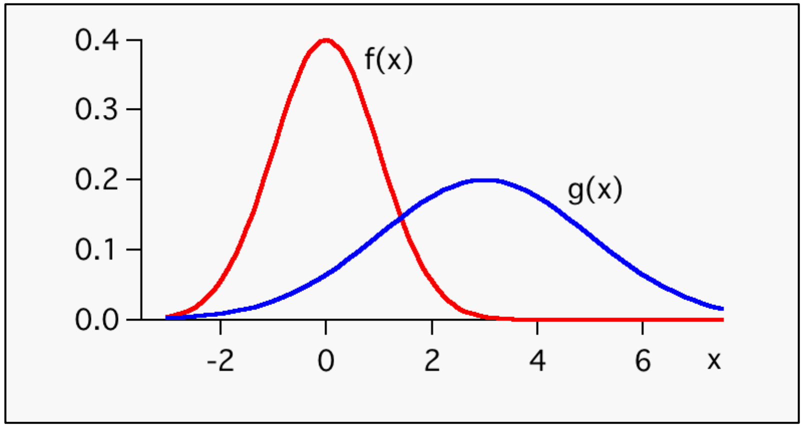
Given that the probability density of X is $f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ (standard normal distribution) $\mu_x = 0,\ \sigma_x^2 = 1$, find the probability density, average, and variance of Y when linearly transforming Y=2X+3.

→ Reference book [1], p. 53

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

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

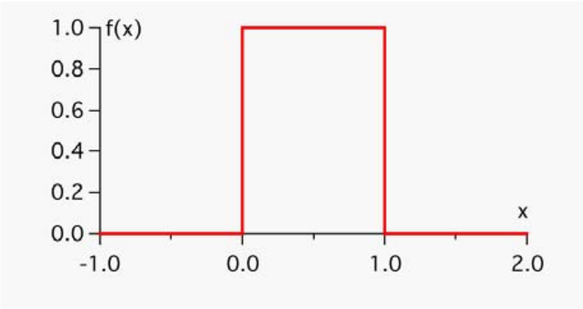
$$\sigma_y^2 = 2^2\sigma_x^2 = 2^2 \cdot 1^2 = 4$$

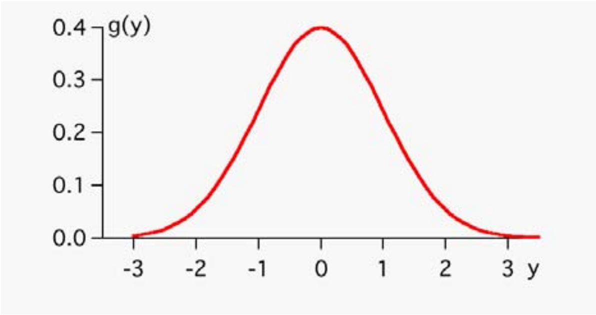


2-5 Variable Transformation

[Example] When the probability density of X is $f(x)=\begin{cases}1 & 0\le x\le 1\\ 0 & x<0, x>1\end{cases}$ (uniform distribution), we need to perform a transform $Y=\Phi(X)$ that enables the probability density of Y to be $g(y)=\frac{1}{\sqrt{2\pi}}e^{-y^2/2}$.

To this end, find $\Phi(X)$.





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

Find y numerically.