19 juillet 2012
PROB #14

確率論 (*Probability Theory*) 第 14 週

w.yamamoto

1 一変量確率分布

本講義でこれまで紹介した、標本空間が一次元の確率分布は、離散分布であれ連続分布であれ、平均と分散が決まれば、パラメータが決まることが多かった。一部を表に抜粋しておく。

表 確率分布と確率関数・密度関数と 2 次までのモーメントの例

確率分布	密度関数・確率関数	パラメータとモーメントの関係式
二項分布	${}_nC_k p^k (1-p)^k$	$\mu_1 = np, \mu_2 = np(1-p)$
ポアソン分布	$\frac{\lambda^k}{k!} e^{-\lambda}$	$\mu_1 = 1/\lambda, \mu_2 = 1/\lambda$
指数分布	$\lambda e^{-\lambda x}$	$\mu_1 = 1/\lambda, \mu_2 = 1/\lambda^2$
正規分布	$\exp\left\{-(x-\mu)^2 / \left(2\sigma^2\right)\right\} / \sqrt{2\pi\sigma^2}$	$\mu_1 = \mu, \mu_2 = \sigma^2$
ガンマ分布	$x^{\alpha-1} e^{-x/\beta} / \beta^\alpha / \Gamma(\alpha)$	$\mu_1 = \alpha\beta, \mu_2 = \alpha\beta^2$

これらから多くの確率分布に対して、「確率分布族 と 母平均(1 次のモーメント) と 母分散(期待値まわりの 2 次のモーメント) を知っていること」と「確率分布族 と パラメータの値 を知っていること」は、同等の情報量であることが分かる。

確率分布は次のような 3 種類のパラメータを持つ。

位置パラメータ 確率分布の位置を表すパラメータ μ で、標準的な確率分布 $F_0(x)$ に対して、μ だけ右に移動した確率分布を $F(y-\mu)$ および $f(y-\mu)$ などと得る。[1]

尺度パラメータ 確率分布の標本空間の尺度変更を表すパラメータ σ で、標準的な確率分布 $F_0(x)$ に対して、時間尺度を σ 倍した確率分布を $F(y/\sigma)$ および $f(y/\sigma)/\sigma$ などと得る。尺度パラメータは、分散など散らばりにも関係がある。

形状パラメータ 変えると確率関数あるいは密度関数の形状が大きく変わるパラメータ。ガンマ分布の *alpha* が一例。

これら 3 種類をすべて持つ確率分布もあるし、パラメータの次元が 3 以上の確率分布もあるが、この講義にはパラメータの次元が 2 までの確率分布しか登場しない。

2 多変量確率分布

多次元の確率変数 $\boldsymbol{X} = \left(X_1, X_2, \ldots, X_p\right)$ を、多次元確率変数と呼ぶことが多いが、多変量[2]と呼ぶこともある。特に多次元の確率変数の確率分布のことは、多変量確率分布と呼ぶ。

連続分布であれば、密度関数は p 変数関数。

$$f(\boldsymbol{x}) \quad = \quad f\left(x_1, \ldots, x_p\right) \tag{1}$$

[1] 位置が変わると標本空間も一緒に移動するとき、これは閾値パラメータと呼ばれることがある。

[2] 多変量解析と多変数解析は、全く異なる学問である。英語では前者は multivariate analysis で後者は multivariable analysis と微妙に異なるが、前者は多次元の確率変数の実現値からなるデータの統計解析の一分野で、後者は変数が複数の関数の解析学であり、意味はかなり異なる。

連続多変量確率分布の例は、p 次元ユークリッド空間 $\mathcal{R}^p$ の上の p 変量正規分布。

$$f(\boldsymbol{x}) \quad = \quad \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right\} \tag{2}$$

離散分布でも、確率関数は p 変数関数。

$$p(\boldsymbol{k}) \quad = \quad p\left(k_1, \ldots, k_p\right) = Pr\left[X_1 = k_1 \quad \text{and} \quad \cdots \quad \text{and} \quad X_p = k_p\right] \tag{3}$$

離散多変量確率分布の例は、1 回の試行で p 種類の互いに疎な事象 $A_1, \ldots, A_p$ のうちのどれかひとつのみが起こる試行を、独立に n 回繰り返したときの、各事象が起こる回数の確率分布である多項分布。

$$p(\boldsymbol{k}) \quad = \quad \frac{n!}{k_1!k_2!\cdots k_p!} {p_1}^{k_1} {p_2}^{k_2} \cdots {p_p}^{k_p} \tag{4}$$

2 次元の正規分布を例に話を進める。密度関数は、参考書には

$$f(x_1, x_2) \quad = \quad \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{{\sigma_1}^2} + \frac{(x_1-\mu_1)^2}{{\sigma_1}^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]\right\} \tag{5}$$

とある。これは

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \ \Sigma = \begin{pmatrix} {\sigma_1}^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & {\sigma_2}^2 \end{pmatrix} \tag{6}$$

と置くと、次のように表現できる。

$$f(x_1, x_2) \quad = \quad \frac{1}{\sqrt{|2pi\Sigma|}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right\} \tag{7}$$

更に両辺の対数をとれば

$$\log f(x_1, x_2) \quad = \quad -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) - \frac{1}{2}\log|2pi\Sigma| \tag{8}$$

となる。式 (8) の第 1 項は $(\boldsymbol{x}-\boldsymbol{\mu})$ についての二次形式、第 2 項は $\boldsymbol{x}$ にも $\boldsymbol{\mu}$ にも依らないことを記しておく。

さて、多変数関数 $f(\boldsymbol{x})$ について、定数 c を与えた時の方程式

$$f(\boldsymbol{x}) = c \tag{9}$$

を満たす $\boldsymbol{x}$ の集合が曲線や曲面を描くとき、それらは等高線あるいは登高面と呼ばれる。山の地図を開くと、高さが一定の地点を結んだ等高線が描かれているのも、同じ定義による。

上の 2 変量正規分布 $f(x_1, x_2)$ の等高線は

$$\frac{1}{\sqrt{|2pi\Sigma|}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right\} = c \tag{10}$$

という方程式を c を変えながら求める。しかし、上に示したようにこの方程式は両辺の対数をとれば

$$-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) - \frac{1}{2}\log|2pi\Sigma| = \log c \tag{11}$$

であり、第 2 項が定数で $\boldsymbol{x}$ を含まないことから、左辺に移項すれば

$$(\boldsymbol{x}-\boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) = -2\log c - \frac{1}{2}\log|2pi\Sigma| \tag{12}$$

との表現を得る。あとはこれが、$\boldsymbol{x}-\boldsymbol{\mu}$ に関する二次形式であることに気づけば、この密度関数の等高線が楕円であることを導くのにあと一歩まで近づく。

線形代数学に、ベクトル $\boldsymbol{x}$ と対称行列 A で用いた $\boldsymbol{x}'A\boldsymbol{x}$ という量が登場する。これは $\boldsymbol{X}$ の成分の二次関数であることから、二次形式と呼ばれる。二次形式には、正の定数 c に対して

$$\boldsymbol{x}'A\boldsymbol{x} = c \tag{13}$$

を満たす $\boldsymbol{x}$ の集合が、A の固有値の符号に応じて

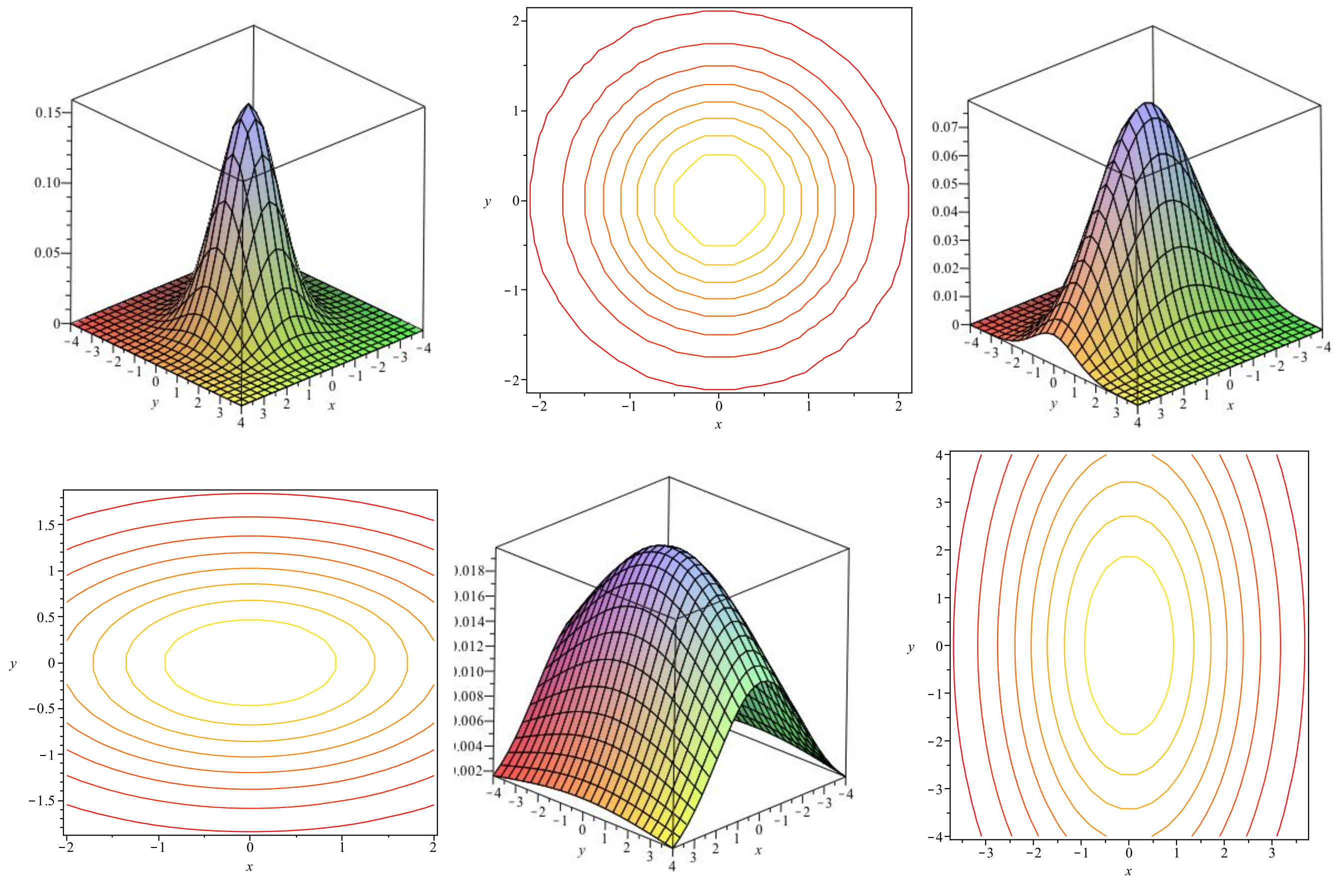
1. A の固有値がすべて正であれば楕円型
2. A の固有値がすべて非負であれば放物型
3. A の固有値に正と負のものがあれば双曲型

の曲線、曲面あるいは超曲面となる、ことが知られている。[3]

この知識と、Σ^{-1} の固有値は、$\rho \neq 1$ あるいは $\sigma_1 \neq 0$ かつ $\sigma_2 \neq 0$ のときに必ず正となる、という事実を用いると

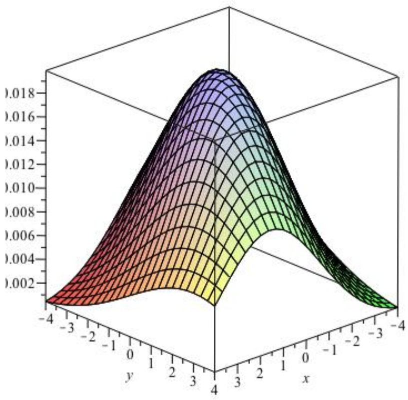
$$(\boldsymbol{x}-\boldsymbol{\mu})'\,\Sigma^{-1}\,(\boldsymbol{x}-\boldsymbol{\mu}) = -2\log c - \frac{1}{2}\log|2pi\Sigma| \tag{14}$$

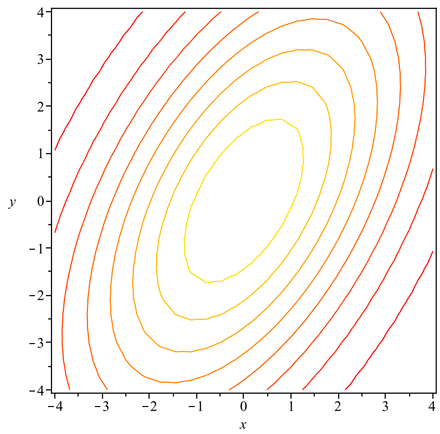
を満たす $\boldsymbol{x}$ は、$\boldsymbol{\mu}$ を中心とする楕円になることが分かる。平面の中心を $\boldsymbol{\mu}$ としたときに、第 1 象限と第 3 象限に多くの確率が含まれるか、第 2 象限と第 4 象限に多くの確率が含まれるかは、ρ の符号で決まる。



A

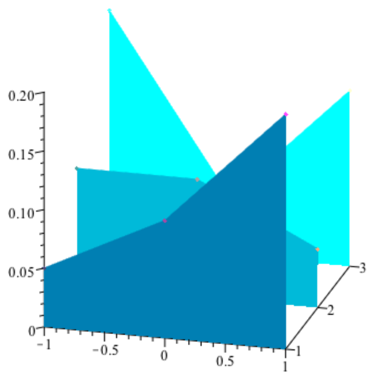
[3] これ、既習得ですよね …?





2.1　条件付き分布と周辺分布

離散分布では、次のようなグラフを思い浮かべると簡単である。



ここでは連続分布の場合に、特に 2 変量正規分布で、片方の確率変数の値が予め分かっているときの、もう一方の確率変数の値の条件付き確率分布を考える。

累積分布関数で表した条件付き確率の公式

$$F(X_1 \leq x_1 \quad \text{and} \quad X_2 \leq x_2) = F(X_1 \leq x_1 | X_2 \leq x_2) F(X_2 \leq x_2) \tag{15}$$

から、確率関数あるいは確率密度関数についても

$$p(x_1, x_2) = p(x_1|x_2)\, p(x_2) \tag{16}$$

および

$$f(x_1, x_2) = f(x_1|x_2)\, f(x_2) \tag{17}$$

が成り立つ。これは、離散分布の場合には累積分布関数の差分が p であること、連続分布の場合には累積分布関数の微分が f であることから、それぞれの条件付き確率や条件付き密度関数も同様に差分や微分で定める。

このことから離散分布における条件付き確率関数は

$$p(x_1|x_2) = \frac{p(x_1, x_2)}{p(x_2)} \tag{18}$$

と得ることができ、連続分布における条件付き確率密度関数も

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f(x_2)} \tag{19}$$

と得られる。

では、周辺分布はどう得るか。2 変量正規分布に話を戻す。2 変量正規分布

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1{}^2} + \frac{(x_1-\mu_1)^2}{\sigma_1{}^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]\right\} \tag{20}$$

では、x_1 について積分

$$\begin{aligned} f(x_2) &= \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1 \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1{}^2} + \frac{(x_1-\mu_1)^2}{\sigma_1{}^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]\right\} dx_1 \end{aligned} \tag{21}$$

を行えば、x_2 の周辺確率分布の確率密度関数が得られる。これを周辺密度と略すことがある。

さて、上の積分を計算すると、

$$f(x_2) = \frac{1}{\sqrt{2\pi\sigma_2{}^2}} \exp\left\{-\frac{(x_2-\mu_2)^2}{2\sigma_2{}^2}\right\} \tag{22}$$

となる。つまり、多変量正規分布の 1 次元の周辺分布は、平均も分散がその成分の平均と分散を持つ、正規分布なのである。この事実を「正規分布はどこから見ても正規分布」ということがある。実際には x_1 の周辺分布、x_2 の周辺分布だけでなく、$a_1x_1 + a_2x_2$ の形のあらゆる 1 次元の周辺分布は正規分布になる。

X_2 の周辺確率分布の確率密度関数が分かれば、$X_1|X_2 = x_2$ の条件付き確率分布の確率密度関数は、割り算をするだけである。

$$\begin{aligned} f(x_1|x_2) &= \frac{f(x_1, x_2)}{f(x_2)} \\ &= \frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1{}^2} + \frac{(x_1-\mu_1)^2}{\sigma_1{}^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]\right\}}{\frac{1}{\sqrt{2\pi\sigma_2{}^2}} \exp\left\{-\frac{(x_2-\mu_2)^2}{2\sigma_2{}^2}\right\}} \\ &= \frac{1}{\sqrt{2\sigma_1{}^2(1-\rho^2)}} \exp\left\{-\frac{\left(x_1-\mu_1-\frac{\sigma_1}{\sigma_2}\rho(x_2-\mu_2)\right)^2}{2\sigma_1{}^2(1-\rho^2)}\right\} \end{aligned} \tag{23}$$

となる。

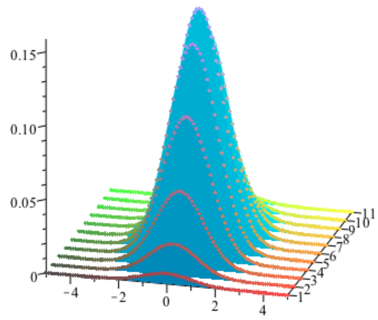
このことから条件付き確率分布も正規分布であり、の期待値は

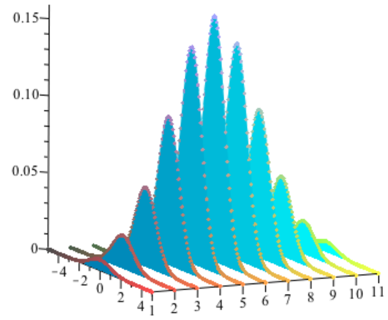
$$E[X_1|X_2 = x_2] = \mu_1 + \frac{\sigma_1}{\sigma_2}\rho(x_2-\mu_2) \tag{24}$$

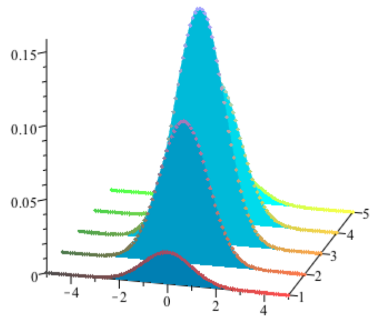
で、分散は

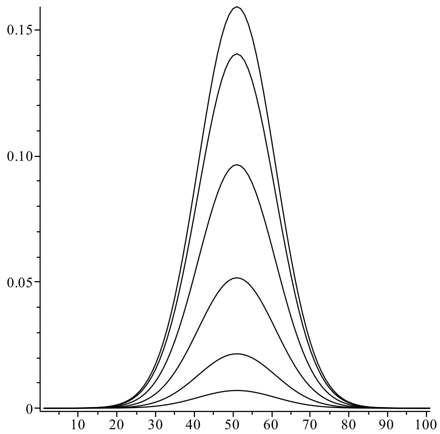
$$V[X_1|X_2 = x_2] = \sigma_1{}^2\left(1-\rho^2\right) \tag{25}$$

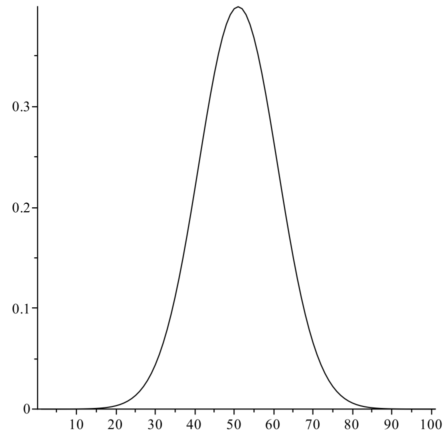
となることが分かる。











2.2 累積分布関数

連続分布の場合：

$$\begin{aligned} F(\boldsymbol{x}) &= Pr\left[X_1 \leq x_1 \quad \text{and} \quad \cdots \quad \text{and} \quad X_p \leq x_p\right] \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} f\left(x_1, \ldots, x_p\right) dx_1 \cdots dx_p \end{aligned} \tag{26}$$

離散分布の場合：

$$\begin{aligned} F(\boldsymbol{k}) &= Pr\left[X_1 \leq x_1 \quad \text{and} \quad \cdots \quad \text{and} \quad X_p \leq x_p\right] \\ &= \sum_{x_1=0}^{k_1} \cdots \sum_{x_p=0}^{k_p} p\left(x_1, \ldots, x_p\right) \end{aligned} \tag{27}$$

2.3 モーメント

多次元の確率変数の場合、モーメントは

$$X_1{}^{k_1} X_2{}^{k_2} \cdots X_p{}^{k_p} \tag{28}$$

の期待値、

$$m_{k_1 k_2 \cdots k_p} \qquad E\left[X_1{}^{k_1} X_2{}^{k_2} \cdots X_p{}^{k_p}\right] = \int \cdots \int_X x_1{}^{k_1} x_2{}^{k_2} \cdots x_p{}^{k_p} f\left(x_1, x_2, \ldots, x_p\right) dx_1 dx_2 \cdots dx_p \tag{29}$$

で定義する。

1 次のモーメントは上の定義で、$k_1 + k_2 + \cdots + k_p = 1$ となる全ての期待値であり、全部で p 個ある。

$$\mu_i = \int \cdots \int_X x_i f\left(x_1, x_2, \ldots, x_p\right) dx_1 dx_2 \cdots dx_p, \quad i = 1, \ldots, p. \tag{30}$$

これは

$$\mu_i = \int x_i dx_i \int \cdots \int_X f\left(x_1, x_2, \ldots, x_p\right) dx_1 dx_2 \cdots dx_p, \quad i = 1, \ldots, p. \tag{31}$$

でもあるので、周辺分布の期待値に他ならない。

2 次のモーメントは上の定義で、$k_1 + k_2 + \cdots + k_p = 2$ となる全ての期待値であり、p^2 個ある。

$$\mu_{ij} = \int \cdots \int_X x_i x_j f\left(x_1, x_2, \ldots, x_p\right) dx_1 dx_2 \cdots dx_p, \quad i = 1, \ldots, p, j = 1, \ldots, p. \tag{32}$$

2 次の期待値まわりのモーメントも、同様に p^2 個ある。

$$\sigma_{ij} = \int \cdots \int_X (x_i - \mu_i)\left(x_j - \mu_j\right) f\left(x_1, x_2, \ldots, x_p\right) dx_1 dx_2 \cdots dx_p, \quad i = 1, \ldots, p, j = 1, \ldots, p. \tag{33}$$

が、対称性から $i \neq j$ ならば $\mu_{ij} = \mu_{ji}$ および $\sigma_{ij} = \sigma_{ji}$ となるので、実際には $p(p-1)/2$ 個で済む。

分散も期待値と同様に、

$$\sigma_{ii} = \int (x_i - \mu_i)^2 dx_i dx_i \int \cdots \int_X f\left(x_1, x_2, \ldots, x_p\right) dx_1 dx_2 \cdots dx_p, \quad i = 1, \ldots, p. \tag{34}$$

より、周辺分布の分散と等しい。

2.4 期待値ベクトル

一次元の確率分布について、期待値 $\mu = E[X]$ は確率分布の 位置 を表す量のひとつである、と説いた。多次元の確率分布についても同様で、期待値ベクトル $\boldsymbol{\mu} = E[\boldsymbol{X}]$ が確率分布の 位置 を表す。

期待値ベクトルは、確率変数の各成分の周辺期待値を並べたベクトルになる。

$$\boldsymbol{\mu} = \left(\mu_1, \mu_2, \ldots, \mu_p\right) \tag{35}$$

2.5 分散共分散行列

一次元の確率分布について、分散 $\sigma^2 u = V[X] = E\left[(X - \mu)^2\right]$ は確率分布の 散らばり具合(広がり具合、ばらつきとも) を表す量のひとつである、と説いた。多次元の確率分 p 布についても同様だが、各成分の分散だけでなく、成分間の関係を表す共分散も用いた、分散共分散行列 $\Sigma = V[\boldsymbol{X}] = E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})']$ が確率分布の 散らばり具合 を表す。

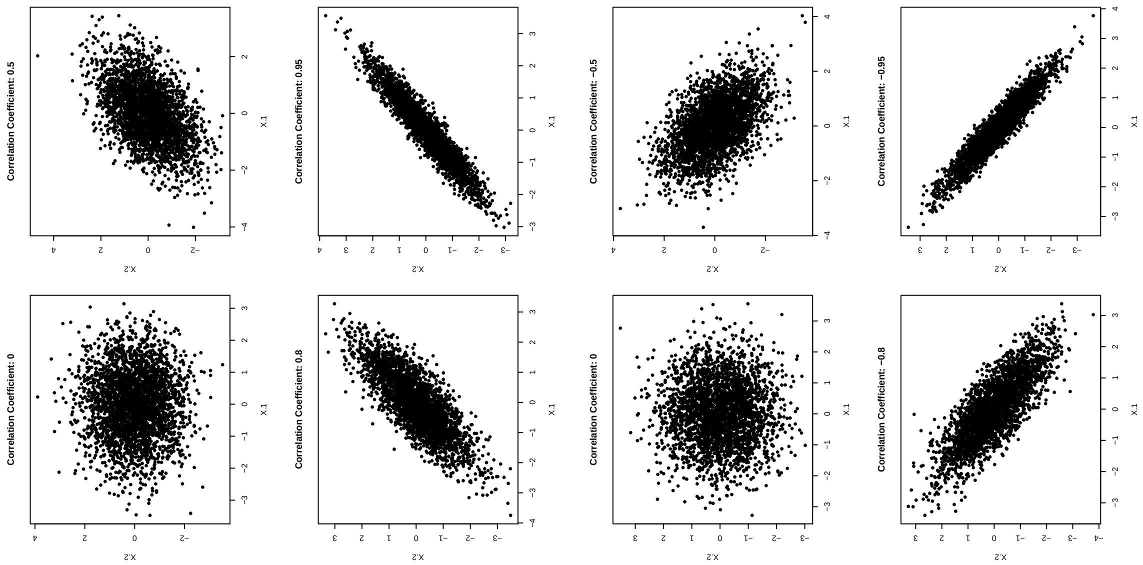
$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & & \sigma_{2p} \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & & \sigma_{pp} \end{pmatrix}$$

$$
= \begin{pmatrix}
V[X_1] & Cov[X_1, X_2] & \dots & Cov\left[X_1, X_p\right] \\
Cov[X_2, X_1] & V[X_2] & \dots & Cov\left[X_2, X_p\right] \\
\vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
Cov\left[X_p, X_1\right] & Cov\left[X_p, X_2\right] & \dots & V\left[X_p\right]
\end{pmatrix} \tag{36}
$$

ここで共分散 $Cov\left[X_i, X_j\right]$ とは

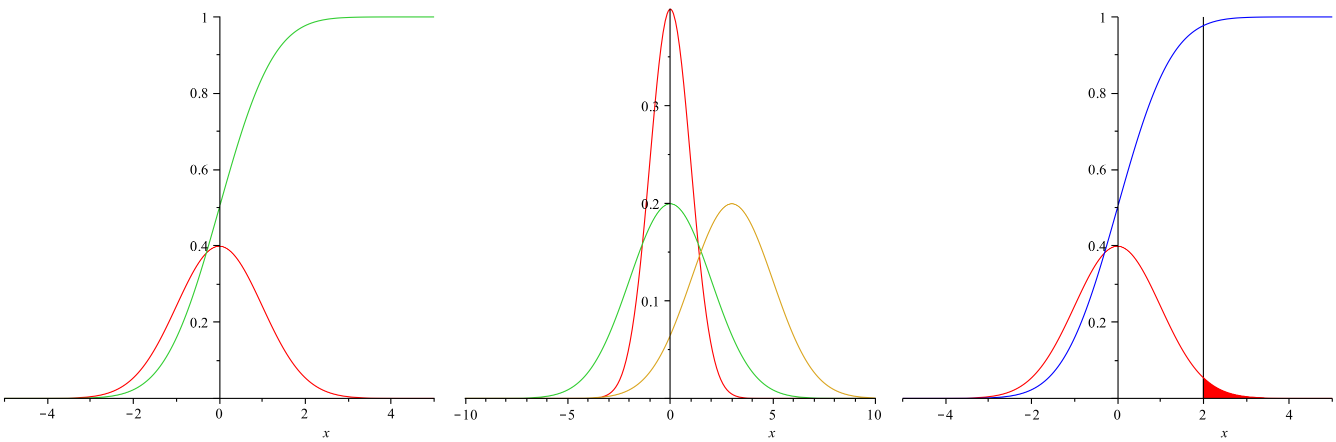
$$
Cov\left[X_i, X_j\right] = E\left[(X_i - \mu_i)\left(X_j - \mu_j\right)\right] \tag{37}
$$

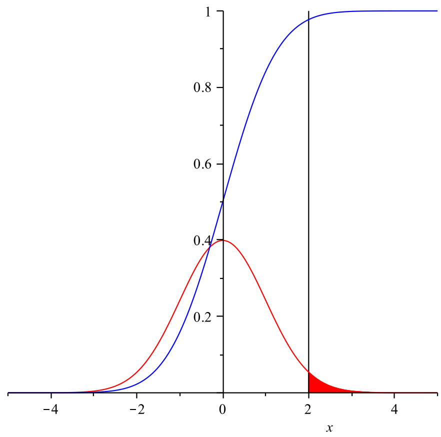
で定義される。$i = j$ であれば X_i の分散 $V[X_i] = E\left[(X_i - \mu_i)^2\right]$ と等しい。



2.6　モーメント母関数

$$
E\left[\exp\left(X_1 t_1 + X_2 t_2 + \cdots X_p t_p\right)\right] = E\left[e^{X_1 t_1} e^{X_2 t_2} \cdots e^{X_p t_p}\right] \tag{38}
$$




1
0.8
0.6
0.4
0.2
−4
−2
0
2
4
x

ギリシャ文字一覧 (英語表記、括弧は同一文字の別書体)

A, α	alpha	B, β	beta	Γ, γ	gamma	Δ, δ	delta
$E, \epsilon\ (\varepsilon)$	epsilon	Z, ζ	zeta	H, η	eta	$\Theta, \theta\ (\vartheta)$	theta
I, ι	iota	K, κ	kappa	Λ, λ	lambda	M, μ	mu
N, ν	nu	Ξ, ξ	xi	O, o	omicron	$\Pi, \pi\ (\varpi)$	pi
$R, \rho\ (\varrho)$	rho	$\Sigma, \sigma\ (\varsigma)$	sigma	T, τ	tau	Υ, υ	upsilon
$\Phi, \phi\ (\varphi)$	phi	X, χ	chi	Ψ, ψ	psi	Ω, ω	omega

ギリシャ文字一覧 (日本語表記、括弧内の読み仮名はそのように読む場合もあるとの参考)

A, α	アルファ	B, β	ベータ
Γ, γ	ガンマ	Δ, δ	デルタ
$E, \epsilon\ (\varepsilon)$	イプシロン (エプシロン)	Z, ζ	ゼータ
H, η	イータ	$\Theta, \theta\ (\vartheta)$	シータ
I, ι	イオタ (アイオタ)	K, κ	カッパ
Λ, λ	ラムダ	M, μ	ミュー
N, ν	ニュー	Ξ, ξ	グザイ (ザイ), クシー
O, o	オミクロン	$\Pi, \pi\ (\varpi)$	パイ
$R, \rho\ (\varrho)$	ロー	$\Sigma, \sigma\ (\varsigma)$	シグマ
T, τ	タウ	Υ, υ	ウプシロン (ユプシロン)
$\Phi, \phi\ (\varphi)$	ファイ	X, χ	カイ
Ψ, ψ	プサイ (サイ)	Ω, ω	オメガ

ギリシャ文字大文字

$$AB\Gamma\Delta EZH\Theta IK\Lambda MN\Xi O\Pi R\Sigma T\Upsilon\Phi X\Psi\Omega \tag{39}$$

ギリシャ文字小文字

$$\alpha\beta\gamma\delta\epsilon\zeta\eta\theta\iota\kappa\lambda\mu\nu\xi o\pi\rho\sigma\tau\upsilon\phi\chi\psi\omega \tag{40}$$