Hand out

due	10 juillet 2014
cur	10 juillet 2014
ver	0
rev	0

確率論 (*Probability*) 第 14 週

w.yamamoto

1　確率変数の変換 (1 次元)

確率変数 X が確率密度関数 $f_x(x)$ を持つ確率分布に従うとする。

1.1　位置変換

確率分布の平均を μ だけ横にずらすことを、位置変換と呼ぶことがある。確率変数 Y の従う確率分布が、確率変数 X の従う確率分布から μ だけずれているとは、

$$Y \quad \sim \quad X+\mu \tag{1}$$

という関係にあることを言う。連続確率変数の場合には、X と Y の標本空間は共通でなければならない。このとき Y の密度関数は

$$y \quad = \quad x+\mu \tag{2}$$

かつ $x \in \mathcal{X} \Rightarrow y \in \mathcal{Y}$ から、

$$f_y(y) = f_x(x-\mu) \tag{3}$$

となる。

これは実は、確率変数の関数の期待値を計算する際に、$y = x+\mu$ と変数変換をすると、$dx = dy$ と積分領域が変わらないことから

$$\int_X g(y)\, f_y(y)\, dy \quad = \quad \int_X g(y)\, f_x(y-\mu)\, dy \tag{4}$$

と右辺で積分ができることと同等である。

このことを以て、確率分布の μ による位置変換後の確率変数 Y の確率密度関数は $f(y-\mu)$ である、という。

1.2　尺度変換

確率変数の原点からの距離 (あるいは値) を $1/\sigma$ 倍することを、尺度変換と呼ぶことがある。σ 倍することも、同じく尺度変換という。[1]

確率論では、

1. 期待値が μ の非負の確率変数 X について、期待値を 1 にする変換: $Y \sim X/\mu$
2. 期待値が 0 で分散が σ^2 の確率変数 X について、分散を 1 にする変換: $Y \sim X/\sigma$

という方向の尺度変換と

1. 非負で期待値が 1 の確率変数 Y について、期待値を μ にする変換: $XsimY\mu$
2. 期待値が 0 で分散が 1 の確率変数 Y について、分散を σ^2 にする変換: $X \sim Y\sigma$

[1]たとえば m 単位を km 単位に変換するときに 1/1000 倍することと、km 単位を m 単位に変換するときに 1000 倍すること、どちらく単位の変換と言われるのと同じこと。

という上とは逆方向の尺度変換をよく見かける。いずれにせよ変換のヤコビアンは

$$dx = \mu dy \quad dy = \frac{1}{\mu}dx$$

あるいは

$$dx = \frac{1}{\sigma}dy \quad dy = \sigma dx$$

となる。

例えば指数分布の密度関数には

$$f(x) = \frac{1}{\mu}e^{-x/\mu} = \frac{\lambda^{-\lambda x}}{e} \tag{5}$$

のふたつの表現がある。平均が 1 の指数分布の密度関数が

$$f_1(z) = e^{-z} \tag{6}$$

と見比べると、式 (5) の真ん中の表現は Z を μ 倍すると期待値も μ になるという意味であり、右側の表現は Z を $1/\lambda$ 倍すると平均が $1/\lambda$ の指数分布になるという意味である。参考までに両者が同じ確率分布であるためには、$\mu = 1/\lambda$ が成り立たなければならない。

確率変数 Y の尺度が、確率変数 X の尺度の σ 倍であるとは、

$$Y/\sigma \sim X \quad \text{or} \quad Y \sim X\sigma \tag{7}$$

という関係にあることを言う。このとき、連続確率変数の場合には、X と Y の標本空間は共通であるなら、Y の密度関数は

$$y = x\sigma \tag{8}$$

かつ $x \in \mathcal{X} \Rightarrow y \in \mathcal{Y}$ から、

$$f_y(y) = \frac{1}{\sigma}f_x\left(\frac{x}{\sigma}\right) \tag{9}$$

となる。

これは実は、確率変数の関数の期待値を計算する際に、$y = x\sigma$ と変数変換をすると、$dx = dy/\sigma$ であり、ここでも積分領域が変わらないことから

$$\int_X g(y) f_y(y)\, dy = \int_X g(y) \frac{1}{\sigma} f_x\left(\frac{y}{\sigma}\right) dy \tag{10}$$

と右辺で積分ができることと同等である。

このことを以て、確率分布の μ による位置変換後の確率変数 Y の確率密度関数は $f_x(y/\sigma)$ である、という。

1.3 一般の変換

一般の変換の場合には、確率変数 X の密度関数を $f_x(x)$ とすると、一階微分可能かつ狭義単調な関数 $h(\cdot)$ を用いて $Y \sim h(X)$ という関係にある確率変数 Y の密度関数は、

$$f_y(y) = \frac{1}{\left|\frac{\partial h^{-1}(y)}{\partial x}\right|} f_x\left(h^{-1}(y)\right) \tag{11}$$

となる。

2　正規分布

正規分布は「誤差の分布」として知られている。ある量 X を計るとする。重量計に乗せてメモリを読む、ということを何回か繰り返して測定値を複数個、$X_1, X_2, \ldots, X_n$ と得て、それらの平均 $\overline{X}_n$ を X の測定結果、とする実験を考える。このとき $\overline{X}_n$ は、n が大きくなればなるほど、真値 X を中心とし、分散が σ^2/n と測定回数に反比例して小さくなる、正規分布に従うようになる。このことは中心極限定理によって正当化されるが、詳細は後に譲る。正規分布の密度関数は、次のように位置変換と尺度変換を用いて、得る。

平均が 0、分散が 1 の正規分布の確率密度関数は

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{12}$$

で与えられる。$\frac{1}{\sqrt{2\pi}}$ は、全積分を 1 にするための乗数で、基準化定数と呼ばれることがある。

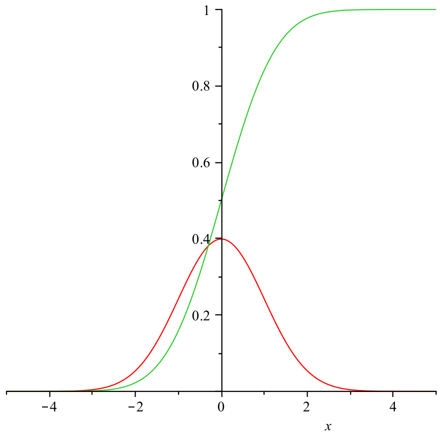


図 1: 標準正規分布の密度関数と累積分布関数

まず、平均が μ で分散が 1 の正規分布の密度関数は、位置変換を用いて、

$$f(x;\mu) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2}\right) \tag{13}$$

で与えられることが分かる。

次に、分散を σ^2 倍するということが尺度は σ 倍することに他ならない

$$V[aX] = a^2 V[X] \tag{14}$$

ことを思い出せば、平均が μ で分散が σ^2 の正規分布の密度関数は、

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{15}$$

で与えられることが分かる。指数の肩に入った $1/\sigma^2$ は $x-\mu$ を $(x-\mu)/\sigma$ に置き換えることによって出現し、指数の外の $1/\sigma$ は尺度変換のヤコビヤンとして出現していることに注意する。これが、平均 μ、分散 σ^2 の正規分布の密度関数である。

2.1　e^{-x^2} の積分

e^{-x^2} の曲線は美しいベル型を描くが、その不定積分は誤差関数と呼ばれ、解析的に陽な表現を持たない。また $(-\infty,\infty)$ 上の定積分は計算でき、 $\sqrt{\pi}$ となるが、計算には極座標に持ち込む意気込みが必要となる。

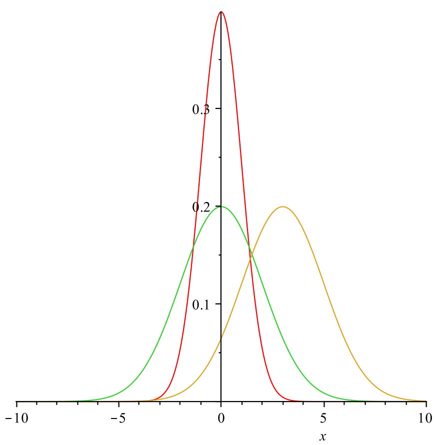


図 2: 標準正規分布の密度関数と累積分布関数

まず $I = \int_{-\infty}^{\infty} \exp\left(-x^2\right) dx$ と置く。指数の肩で円を描かせたいので、 $\int_{-\infty}^{\infty} \exp\left(-y^2\right) dx$ との積を考える。これも I に等しいので、

$$
\begin{aligned}
I^2 &= \left\{\int_{-\infty}^{\infty} \exp\left(-x^2\right) dx\right\}\left\{\int_{-\infty}^{\infty} \exp\left(-y^2\right) dy\right\} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-x^2\right)\exp\left(-y^2\right) dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-x^2-y^2\right) dxdy \qquad (16)
\end{aligned}
$$

2 行目は積分順序を交換しただけ、3 行目は指数の肩をまとめただけ。これを極座標表現 $(x, y) = (r\cos\theta, r\sin\theta)$ に書き換えると、

$$
I^2 = \int_0^{2\pi}\int_0^{\infty} \exp\left(-r^2\right) rdrd\theta \qquad (17)
$$

となる。ここで被積分関数に現れた r は、以下に示すように、 $dxdy$ から $drd\theta$ に変換するときの、ヤコビ行列の行列式である。

$$
\begin{aligned}
\begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} &= \begin{vmatrix} \frac{\partial r\cos\theta}{\partial r} & \frac{\partial r\cos\theta}{\partial \theta} \\ \frac{\partial r\sin\theta}{\partial r} & \frac{\partial r\sin\theta}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} \\
&= r\left(\cos^2\theta + \sin^2\theta\right) = r \qquad (18)
\end{aligned}
$$

これより、

$$
dxdy = \left|\frac{\partial (x, y)^t}{\partial (r, \theta)}\right| drd\theta = rdrd\theta \qquad (19)
$$

と分かる。

さて、積分を続ける。被積分関数に θ が含まれないので、

$$
\begin{aligned}
I^2 &= \int_0^{2pi} d\theta \int_0^{\infty} \exp\left(-r^2\right) rdr \\
&= 2\pi \int_0^{\infty} \exp\left(-r^2\right) rdr \qquad (20)
\end{aligned}
$$

ここで $u = r^2$ (つまり $r = \sqrt{u}$) と置換すると、

$$
du = 2rdr \quad \Leftrightarrow \quad dr = \frac{1}{2r}du \qquad (21)
$$

かつ

$$
u : 0 \to \infty \quad \text{when} \quad r : 0 \to \infty \qquad (22)
$$

よって、I^2 は

$$I^2 = 2\pi \int_0^\infty \exp(-u) \frac{1}{2} du = \pi \int_0^\infty \exp(-u)\, du \tag{23}$$

となる。この積分は簡単で、

$$\int \exp(-u)\, du = -\exp(-u) + C \tag{24}$$

より、

$$I^2 = \pi \left[-\exp(-u)\right]_0^\infty = \pi\left[0 - (-1)\right] = 2\pi \tag{25}$$

を得る。

以上より、

$$I = \int_{-\infty}^\infty \exp\left(-x^2\right) dx = \sqrt{\pi} \tag{26}$$

となる。

2.2 誤差関数

誤差関数 $erf(x)$ は $[0, x]$ 上の e^{-t^2} の定積分の $2/\sqrt{\pi}$ 倍として定義される。

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \tag{27}$$

数学では特殊関数の 1 つに数えられている。この関数は陽は表現を持たないが、被積分関数 e^{-t^2} を原点の回りでテイラー展開し、項別に定積分を求めれば、誤差関数自体のテイラー級数が得られる。

e^{-t^2} は偶関数であり、奇数次数の項はすべて消えることに注意しつつ、

$$\begin{aligned}
\exp\left(-t^2\right) &= \sum_{k=0}^\infty \frac{1}{k!} (-1)^k t^{2k} \\
&= 1 - t^2 + \frac{1}{2}t^4 - \frac{1}{6}t^6 + \frac{1}{24}t^8 + O\left(t^{10}\right)
\end{aligned} \tag{28}$$

となる。これを項別に積分すると、

$$\begin{aligned}
\int_0^x \exp\left(-t^2\right) &= \sum_{k=0}^x \frac{1}{k!} (-1)^k \int_0^x t^{2k} dt \\
&= \sum_{k=0}^x \frac{1}{k!} (-1)^k \left[\frac{t^{2k+1}}{2k+1}\right]_0^x \\
&= \sum_{k=0}^x \frac{1}{k!} (-1)^k \frac{x^{2k+1}}{2k+1}
\end{aligned} \tag{29}$$

となり、

$$\begin{aligned}
erf(x) &= \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \\
&= \frac{2}{\sqrt{\pi}} \sum_{k=0}^x \frac{1}{k!} (-1)^k \frac{x^{2k+1}}{2k+1} \\
&= \frac{2}{\sqrt{\pi}} \left\{ x - \frac{x^3}{3} + \frac{x^5}{10} - \frac{x^7}{42} + O\left(x^{11}\right) \right\}
\end{aligned} \tag{30}$$

を得る。

2.3　期待値の計算

一番簡単なのは、関数

$$(x-\mu)\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{31}$$

が、点 $(\mu,0)$ に対して奇関数となることを用いると、

$$\int_{-\infty}^{\mu}(x-\mu)\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx \;=\; -\int_{\mu}^{\infty}(x-\mu)\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx \tag{32}$$

が示せて、これより

$$\begin{aligned}\int_{-\infty}^{\infty}(x-\mu)\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx &= \int_{-\infty}^{\mu}(x-\mu)\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx+\int_{\mu}^{\infty}(x-\mu)\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx\\ &= \int_{-\infty}^{\mu}(x-\mu)\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx-\int_{-\infty}^{\mu}(x-\mu)\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx\\ &= 0\end{aligned} \tag{33}$$

となることから、

$$\int_{-\infty}\infty(x-\mu)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx \;=\; E[x-\mu]=0 \tag{34}$$

を得る。以上から

$$E[X] \;=\; \mu \tag{35}$$

を得る。

2.4　正規分布の分散

分散 σ^2 の計算は、部分積分を用いるものと、置換積分を用いるものと、あとモーメント母関数をテイラー展開するもの、があるらしい。以下は部分積分。

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx=1$$

と

$$\begin{aligned}E[X] &= \int_{-\infty}^{\infty}x\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx\\ &= \mu\end{aligned}$$

は所与とする。

$$\begin{aligned}V[X] &= E\left[(X-\mu)^2\right]\\ &= \int_{-\infty}^{\infty}(x-\mu)^2\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx\\ &= \int_{-\infty}^{\infty}\{-(x-\mu)\}\left[\{-(x-\mu)\}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right]dx\end{aligned}$$

ここで

$$\int\{-(x-\mu)\}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx \;=\; \sigma^2\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)+C$$

より、$\{-(x-\mu)\}$ と $\{-(x-\mu)\}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ について部分積分を用いれば

$$\begin{aligned}V[X] &= \left[\sigma^2(x-\mu)^2\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right]_{-\infty}^{\infty}+\int_{-\infty}^{\infty}\sigma^2\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx\\ &= \sigma^2\end{aligned}$$

を得る。

2.5　モーメント母関数

分散やそれ以上のモーメントの計算も可能である。しかし、正規分布のモーメント母関数は

$$
\begin{aligned}
M_X(t) &= E[\exp(tX)] = \int_{-\infty}^{\infty} \exp(tx) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} + tx\right) dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2 - 2\mu x + \mu^2 - 2\sigma^2 xt}{2\sigma^2}\right) dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2 - 2\left(\mu + \sigma^2 t\right) x + \mu^2}{2\sigma^2}\right) dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(x - \mu - \sigma^2\right)^2 - 2\mu\sigma^2 t - \sigma^4 t^2}{2\sigma^2}\right) dx \\
&= \exp\left(\mu t + \frac{\sigma^2}{2} t^2\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(x - \mu - \sigma^2\right)^2}{2\sigma^2}\right) dx \\
&= \exp\left(\mu t + \frac{\sigma^2}{2} t^2\right) \qquad (36)
\end{aligned}
$$

のように、積分を変形するだけで得られる。そのため実際には、モーメント母関数をテイラー展開する方が、速い。

2.6　正規分布の和の分布

モーメント母関数を用いて、次の関係が得られる。
$N\left(\mu_1, {\sigma_1}^2\right) + N\left(\mu_2, {\sigma_2}^2\right) \sim N\left(\mu_1 + \mu_2, {\sigma_1}^2 + {\sigma_2}^2\right)$

3　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。

$$f(\boldsymbol{x}) = f\left(x_1, \ldots, x_p\right) \qquad (37)$$

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。

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right\} \qquad (38)$$

離散分布でも、確率関数は p 変数関数。

$$p(\boldsymbol{k}) = p\left(k_1, \ldots, k_p\right) = Pr\left[X_1 = k_1 \quad \text{and} \quad \cdots \quad \text{and} \quad X_p = k_p\right] \qquad (39)$$

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。

$$p(\boldsymbol{k}) = \frac{n!}{k_1! k_2! \cdots k_p!} {p_1}^{k_1} {p_2}^{k_2} \cdots {p_p}^{k_p} \qquad (40)$$

[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。

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

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1{}^2} + \frac{(x_1-\mu_1)^2}{\sigma_1{}^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]\right\} \quad (41)$$

とある。これは

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \ \Sigma = \begin{pmatrix} \sigma_1{}^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2{}^2 \end{pmatrix} \quad (42)$$

と置くと、次のように表現できる。

$$f(x_1, x_2) = \frac{1}{\sqrt{|2pi\Sigma|}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} \quad (43)$$

更に両辺の対数をとれば

$$\log f(x_1, x_2) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) - \frac{1}{2}\log|2pi\Sigma| \quad (44)$$

となる。式 (44) 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

$$f(\boldsymbol{x}) = c \quad (45)$$

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

$$\frac{1}{\sqrt{|2pi\Sigma|}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} = c \quad (46)$$

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

$$-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) - \frac{1}{2}\log|2pi\Sigma| = \log c \quad (47)$$

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

$$(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = -2\log c - \frac{1}{2}\log|2pi\Sigma| \quad (48)$$

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

$$\boldsymbol{x}'A\boldsymbol{x} = c \quad (49)$$

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

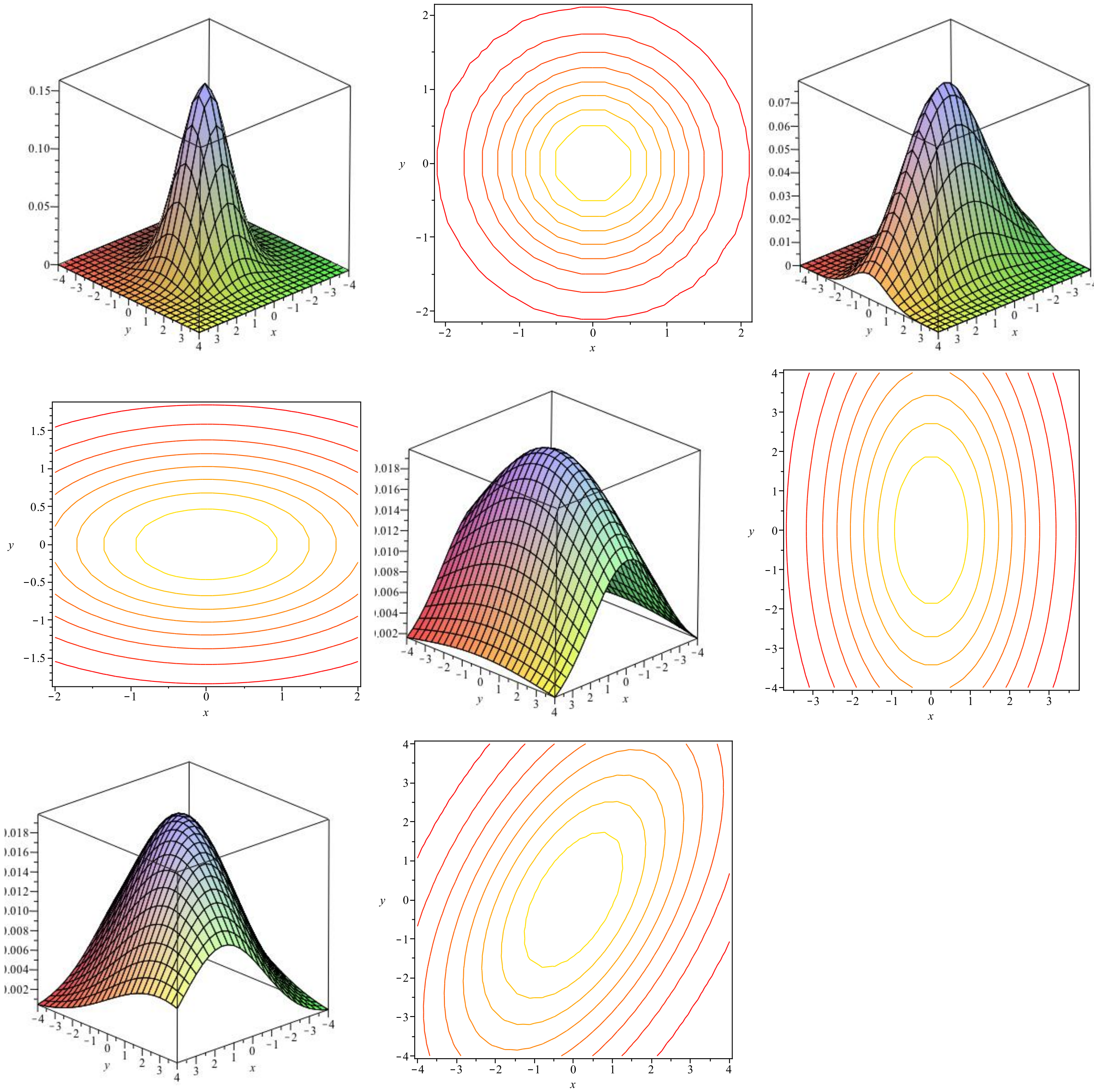
の曲線、曲面あるいは超曲面となる、ことが知られている。[3]

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

$$(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) = -2\log c - \frac{1}{2}\log|2pi\Sigma| \quad (50)$$

[3] これ、既習得ですよね …?

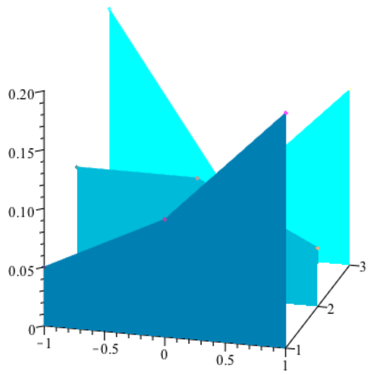
を満たす $\boldsymbol{x}$ は、$\boldsymbol{\mu}$ を中心とする楕円になることが分かる。平面の中心を $\boldsymbol{\mu}$ としたときに、第 1 象限と第 3 象限に多くの確率が含まれるか、第 2 象限と第 4 象限に多くの確率が含まれるかは、ρ の符号で決まる。



A

3.1　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

$$F(X_1 \leq x_1 \quad \text{and} \quad X_2 \leq x_2) \quad = \quad F(X_1 \leq x_1 | X_2 \leq x_2) F(X_2 \leq x_2) \tag{51}$$

から、確率関数あるいは確率密度関数についても

$$p(x_1, x_2) \quad = \quad p(x_1|x_2)\, p(x_2) \tag{52}$$

および

$$f(x_1, x_2) \quad = \quad f(x_1|x_2)\, f(x_2) \tag{53}$$

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

$$p(x_1|x_2) \quad = \quad \frac{p(x_1, x_2)}{p(x_2)} \tag{54}$$

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

$$f(x_1|x_2) \quad = \quad \frac{f(x_1, x_2)}{f(x_2)} \tag{55}$$

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

$$f(x_1, x_2) \quad = \quad \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{{\sigma_1}^2} + \frac{(x_1-\mu_1)^2}{{\sigma_1}^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]\right\} \tag{56}$$

では、x_1 について積分

$$\begin{aligned} f(x_2) &= \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1 \\ &= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{{\sigma_1}^2} + \frac{(x_1-\mu_1)^2}{{\sigma_1}^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]\right\} dx_1 \end{aligned} \tag{57}$$

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、

$$f(x_2) \quad = \quad \frac{1}{\sqrt{2\pi{\sigma_2}^2}} \exp\left\{-\frac{(x_2-\mu_2)^2}{2{\sigma_2}^2}\right\} \tag{58}$$

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。

$$
\begin{aligned}
f(x_1|x_2) &= \frac{f(x_1, x_2)}{f(x_2)} \\
&= \frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1{}^2} + \frac{(x_1-\mu_1)^2}{\sigma_1{}^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right]\right\}}{\frac{1}{\sqrt{2\pi\sigma_2{}^2}} \exp\left\{-\frac{(x_2-\mu_2)^2}{2\sigma_2{}^2}\right\}} \\
&= \frac{1}{\sqrt{2\sigma_1{}^2(1-\rho^2)}} \exp\left\{-\frac{\left(x_1 - \mu_1 - \frac{\sigma_1}{\sigma_2}\rho(x_2-\mu_2)\right)^2}{2\sigma_1{}^2(1-\rho^2)}\right\} \qquad (59)
\end{aligned}
$$

となる。

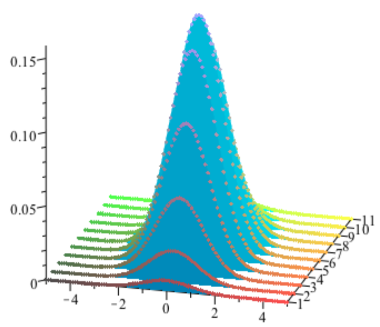
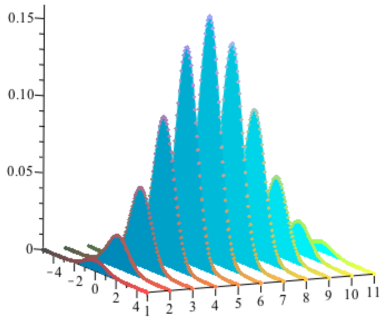
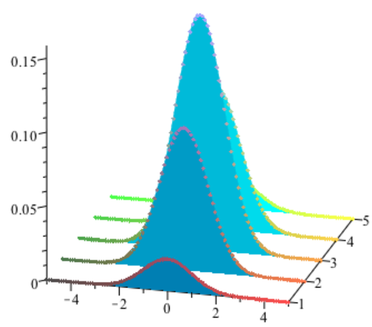
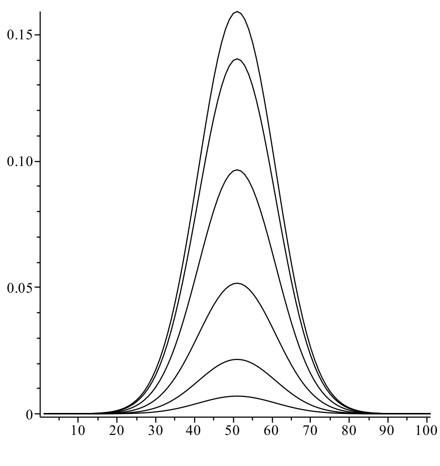
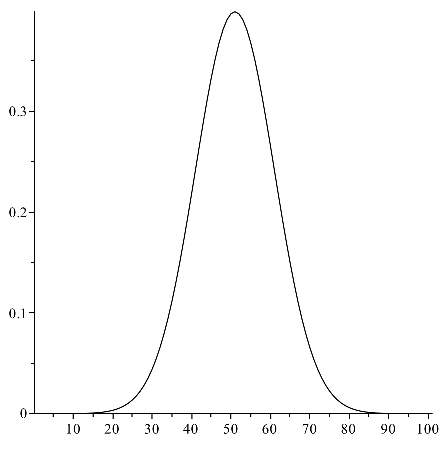
このことから条件付き確率分布も正規分布であり、の期待値は

$$E[X_1|X_2 = x_2] = \mu_1 + \frac{\sigma_1}{\sigma_2}\rho(x_2 - \mu_2) \qquad (60)$$

で、分散は

$$V[X_1|X_2 = x_2] = \sigma_1{}^2\left(1 - \rho^2\right) \qquad (61)$$

となることが分かる。

3.2　累積分布関数

連続分布の場合：

$$F(\boldsymbol{x}) = Pr\left[X_1 \le x_1 \quad \text{and} \quad \cdots \quad \text{and} \quad X_p \le x_p\right]$$
$$= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} f\left(x_1, \ldots, x_p\right) dx_1 \cdots dx_p \qquad (62)$$

離散分布の場合：

$$F(\boldsymbol{k}) = Pr\left[X_1 \le x_1 \quad \text{and} \quad \cdots \quad \text{and} \quad X_p \le x_p\right]$$
$$= \sum_{x_1=0}^{k_1} \cdots \sum_{x_p=0}^{k_p} p\left(x_1, \ldots, x_p\right) \qquad (63)$$

3.3 モーメント

多次元の確率変数の場合、モーメントは

$${X_1}^{k_1} {X_2}^{k_2} \cdots {X_p}^{k_p} \qquad (64)$$

の期待値、

$$m_{k_1 k_2 \cdots k_p} \qquad E\left[{X_1}^{k_1} {X_2}^{k_2} \cdots {X_p}^{k_p}\right] = \int \cdots \int_{X} {x_1}^{k_1} {x_2}^{k_2} \cdots {x_p}^{k_p} f\left(x_1, x_2, \ldots, x_p\right) dx_1 dx_2 \cdots dx_p \qquad (65)$$

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。

$$\mu_i = \int \cdots \int_{X} x_i f\left(x_1, x_2, \ldots, x_p\right) dx_1 dx_2 \cdots dx_p, \quad i = 1, \ldots, p. \qquad (66)$$

これは

$$\mu_i = \int x_i dx_i \int \cdots \int_{X} f\left(x_1, x_2, \ldots, x_p\right) dx_1 dx_2 \cdots dx_p, \quad i = 1, \ldots, p. \qquad (67)$$

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。

$$\mu_{ij} = \int \cdots \int_{X} x_i x_j f\left(x_1, x_2, \ldots, x_p\right) dx_1 dx_2 \cdots dx_p, \quad i = 1, \ldots, p, j = 1, \ldots, p. \qquad (68)$$

2 次の期待値まわりのモーメントも、同様に p^2 個ある。

$$\sigma_{ij} = \int \cdots \int_{X} (x_i - \mu_i)\left(x_j - \mu_j\right) f\left(x_1, x_2, \ldots, x_p\right) dx_1 dx_2 \cdots dx_p, \quad i = 1, \ldots, p, j = 1, \ldots, p. \qquad (69)$$

が、対称性から $i \ne j$ 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、

$$\sigma_{ii} = \int (x_i - \mu_i)^2 dx_i dx_i \int \cdots \int_{X} f\left(x_1, x_2, \ldots, x_p\right) dx_1 dx_2 \cdots dx_p, \quad i = 1, \ldots, p. \qquad (70)$$

より、周辺分布の分散と等しい。

3.4 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。

$$\boldsymbol{\mu} = \left(\mu_1, \mu_2, \ldots, \mu_p\right) \qquad (71)$$

3.5 分散共分散行列

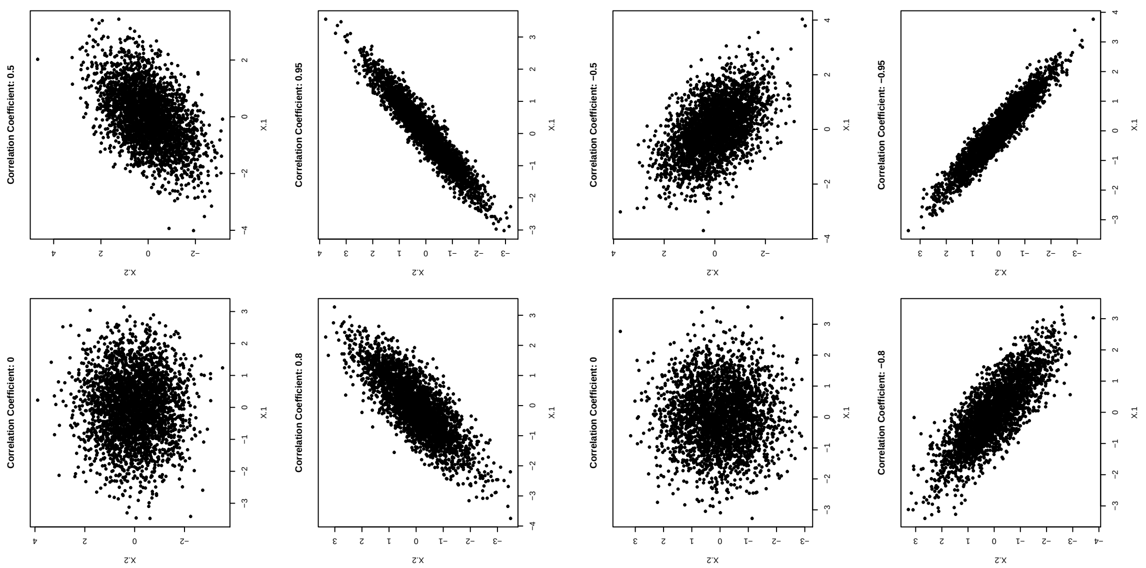
一次元の確率分布について、分散 $\sigma^2 u = V[X] = E\left[(X-\mu)^2\right]$ は確率分布の 散らばり具合(広がり具合、ばらつきとも) を表す量のひとつである、と説いた。多次元の確率分 p 布についても同様だが、各成分の分散だけでなく、成分間の関係を表す共分散も用いた、分散共分散行列 $\Sigma = V[\boldsymbol{X}] = E[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})']$ が確率分布の 散らばり具合 を表す。

$$\begin{aligned}
\Sigma &= \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & & \sigma_{2p} \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & & \sigma_{pp} \end{pmatrix} \\
&= \begin{pmatrix} V[X_1] & Cov[X_1, X_2] & \dots & & Cov\left[X_1, X_p\right] \\ Cov[X_2, X_1] & V[X_2] & \dots & & Cov\left[X_2, X_p\right] \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ Cov\left[X_p, X_1\right] & Cov\left[X_p, X_2\right] & \dots & & V\left[X_p\right] \end{pmatrix}
\end{aligned} \tag{72}$$

ここで共分散 $Cov\left[X_i, X_j\right]$ とは

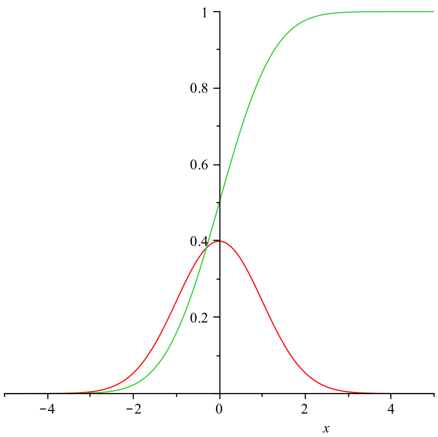
$$Cov\left[X_i, X_j\right] = E\left[(X_i-\mu_i)\left(X_j-\mu_j\right)\right] \tag{73}$$

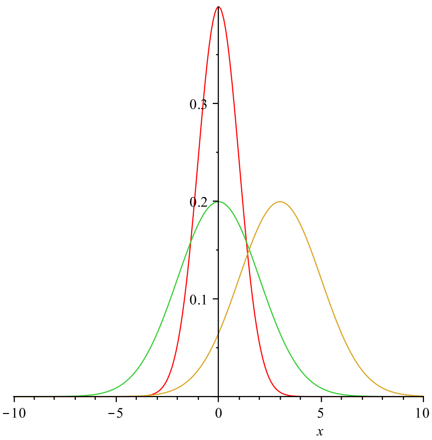
で定義される。$i = j$ であれば X_i の分散 $V[X_i] = E\left[(X_i-\mu_i)^2\right]$ と等しい。

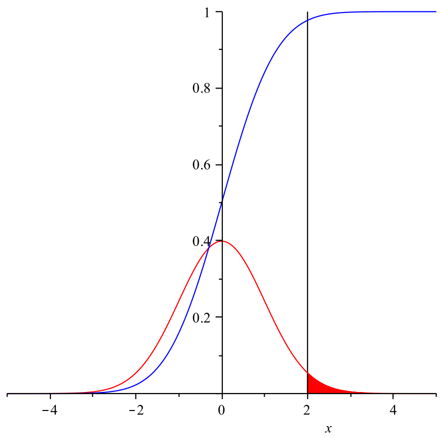


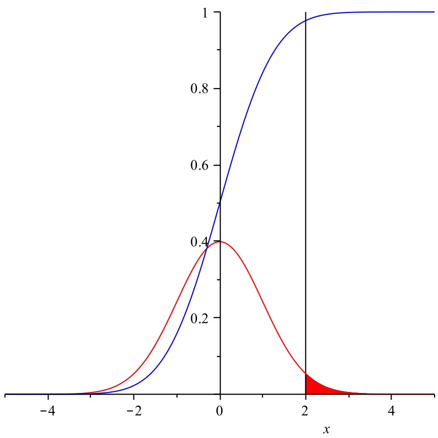
3.6 モーメント 母関数

$$E\left[\exp\left(X_1 t_1 + X_2 t_2 + \cdots X_p t_p\right)\right] = E\left[e^{X_1 t_1} e^{X_2 t_2} \cdots e^{X_p t_p}\right] \tag{74}$$









参考文献

[1] 微分積分学の教科書.

[2] 永田靖 (2005) 「統計学のための数学入門 30 講」朝倉書店.

[3] 宮川雅巳 (1988) 「統計技法」 共立出版.

[4] 稲垣宣生 (2003) 「数理統計学」改訂版, 裳華房.