

多次元ガウス分布の定義式

利用パッケージ

格子点の作成

2次元ガウス分布の作図

分布の設定と計算

分布の作図

# 2次元ガウス分布のグラフの作成

@anemptyarchive (<https://www.anarchive-beta.com/>)

2022/02/06

## 多次元ガウス分布の定義式

多次元ガウス分布は、次の式で定義される確率分布です。

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{D}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^{\top} \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\}$$

$D$ は次元数です。ここでは、2次元のグラフで表現するため $D = 2$ とします。

$D = 2$ のとき、確率変数の値 $\mathbf{x}$ 、平均ベクトル $\boldsymbol{\mu}$ 、分散共分散行列 $\boldsymbol{\Sigma}$ は、次の形状になります。

$$\mathbf{x} = (x_1, x_2), \boldsymbol{\mu} = (\mu_1, \mu_2), \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_2^2 \end{pmatrix}$$

$\sigma_1^2$ は $x_1$ の分散、 $\sigma_2^2$ は $x_2$ の分散、 $\sigma_{1,2} = \sigma_{2,1}$ は $x_1, x_2$ の共分散です。 $\sigma_1^2 > 0, \sigma_2^2 > 0$ を満たす必要があります。

この2次元ガウス分布のグラフを作成します。

## 利用パッケージ

利用するパッケージを読み込みます。

```
# パッケージの読み込み
library(tidyverse)
library(mvnfast)
```

`mvnfast` は、多次元ガウス分布に関するパッケージです。

## 格子点の作成

等高線図を描画するのに格子状の点を利用します。作図の前に、格子点の作成方法を確認します。

`seq()` で、x軸とy軸の値を作成します。

```
# x軸の値を作成
x_vals <- seq(from = 1, to = 5, by = 1)

# y軸の値を作成
y_vals <- seq(from = 11, to = 15, by = 1)
x_vals; y_vals
```

```
## [1] 1 2 3 4 5
```

```
## [1] 11 12 13 14 15
```

`seq()` は、第1引数(`from`)から第2引数(`to`)までの値を第3引数(`by`)に指定した間隔で作成します。

x軸の値を `x_vals`、y軸の値を `y_vals` とします。

作成した値をデータフレームに格納します。

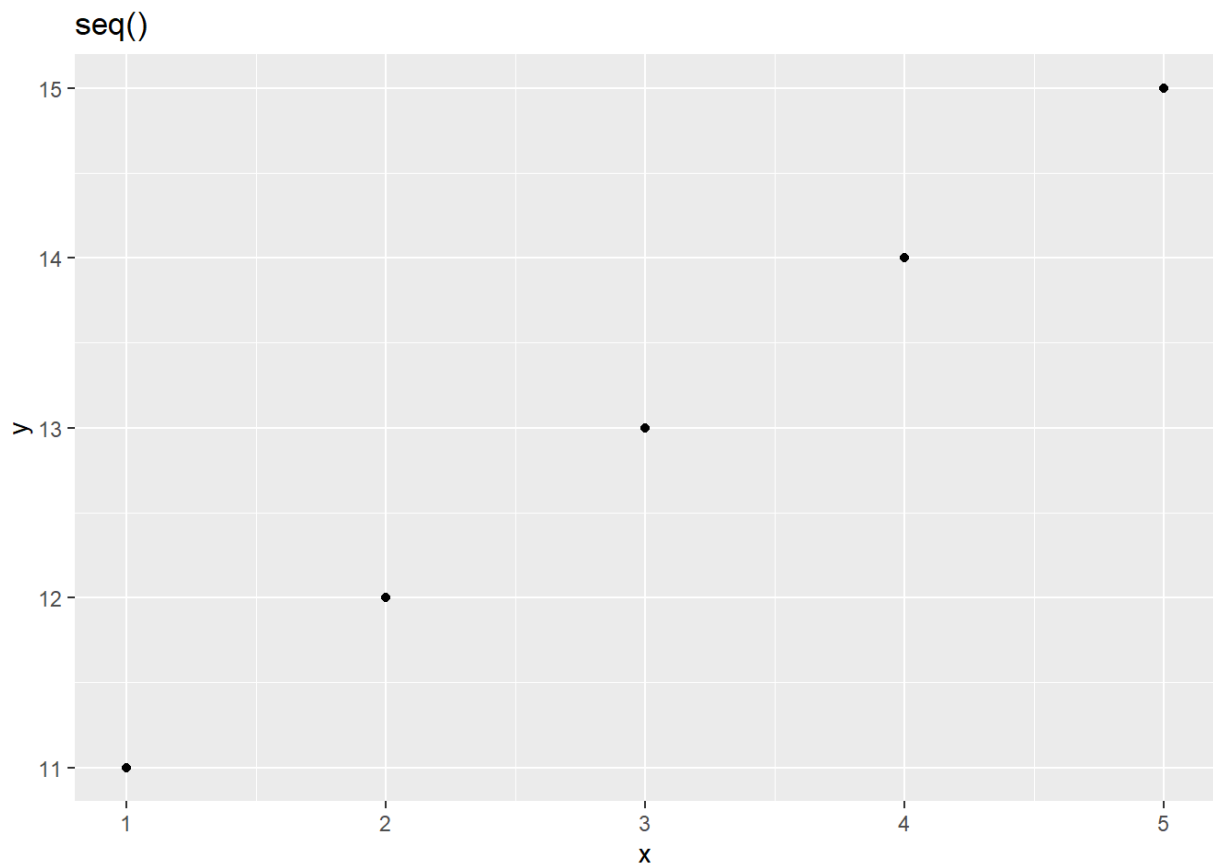
```
# 値をデータフレームに格納
x_vals_df <- tibble::tibble(
  x = x_vals, # x軸の値
  y = y_vals  # y軸の値
)
head(x_vals_df)
```

```
## # A tibble: 5 x 2
##       x     y
##   <dbl> <dbl>
## 1     1    11
## 2     2    12
## 3     3    13
## 4     4    14
## 5     5    15
```

`ggplot2` パッケージで作図するには、作図に用いる値をデータフレームにしておく必要があります。

散布図を作成して値を確認します。

```
# 散布図を作成
ggplot() +
  geom_point(data = x_vals_df, mapping = aes(x = x, y = y)) + # 散布図
  scale_x_continuous(breaks = x_vals) + # x軸目盛
  scale_y_continuous(breaks = y_vals) + # y軸目盛
  labs(title = "seq()")
```



`geom_point()` で散布図を作成できます。

`expand.grid()` で、格子点を作成します。

```
# 格子点を作成
x_grid_df <- expand.grid(
  x = x_vals, # x軸の値
  y = y_vals # y軸の値
)
head(x_grid_df); tail(x_grid_df)
```

```
##    x  y
##  1 1 11
##  2 2 11
##  3 3 11
##  4 4 11
##  5 5 11
##  6 1 12
```

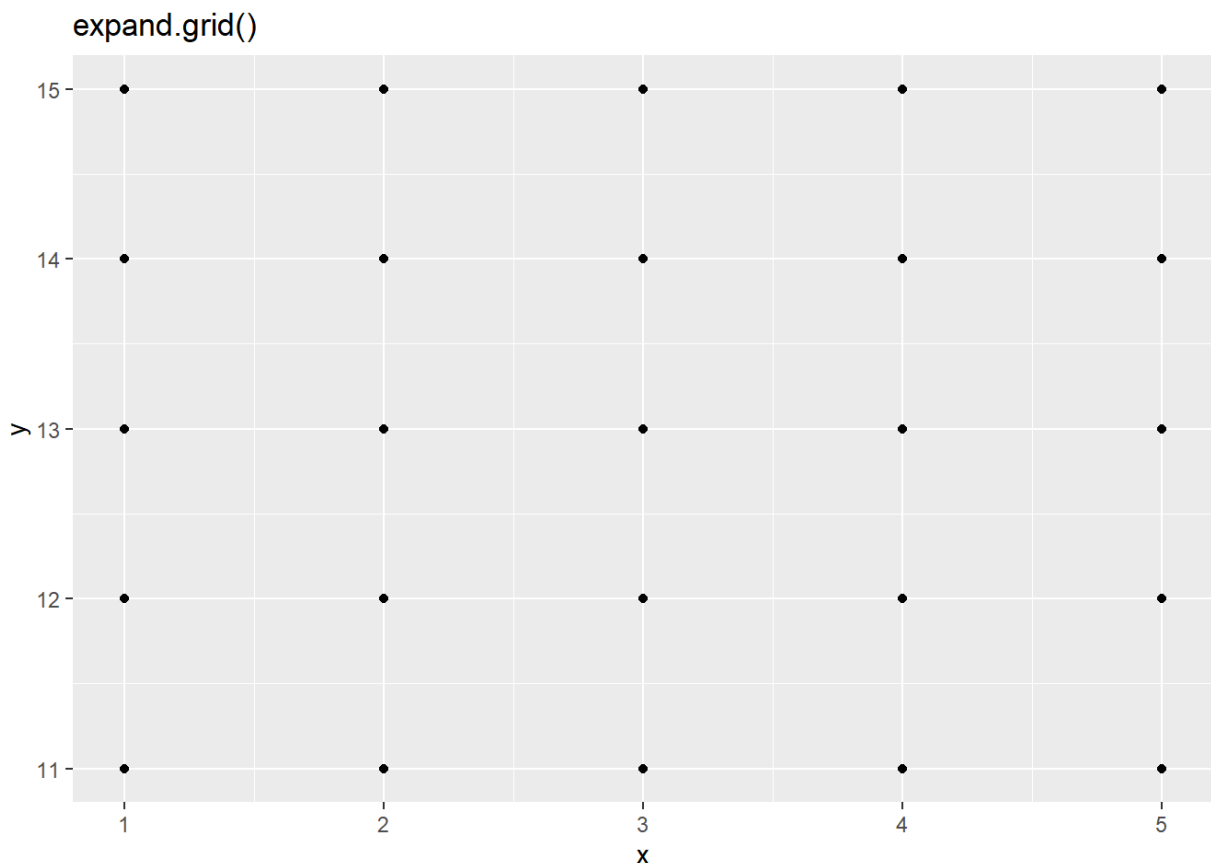
```
##    x  y
## 20 5 14
## 21 1 15
## 22 2 15
## 23 3 15
## 24 4 15
## 25 5 15
```

`expand.grid()` は、引数に指定した値を使って、全ての組み合わせを持つデータフレームを作成します。引数に 列名 = 値 として指定します。

各行が1つの点に対応します。「`x_vals` の要素数」掛ける「`y_vals` の要素数」個の点(組み合わせ)ができます。2つのベクトルの要素数が同じである必要はありません。

作成した点を確認します。

```
# 散布図を作成
ggplot() +
  geom_point(data = x_grid_df, mapping = aes(x = x, y = y)) + # 散布図
  scale_x_continuous(breaks = x_vals) + # x軸目盛
  scale_y_continuous(breaks = y_vals) + # y軸目盛
  labs(title = "expand.grid()") # ラベル
```



`x_vals`, `y_vals` の全ての組み合わせになっているのを確認できます。

## 2次元ガウス分布の作図

格子点の確認ができたので、2次元ガウス分布のグラフを作成します。

## 分布の設定と計算

作図に利用する  $\mathbf{x} = (x_1, x_2)$  の値を作成します。

```
# xの値を作成
x_vals <- seq(from = -5, to = 5, length.out = 51)

# xの点を作成
x_points <- expand.grid(x1 = x_vals, x2 = x_vals) %>%
  as.matrix()
head(x_points)
```

```
##           x1 x2
## [1,] -5.0 -5
## [2,] -4.8 -5
## [3,] -4.6 -5
## [4,] -4.4 -5
## [5,] -4.2 -5
## [6,] -4.0 -5
```

`seq()` で  $x_1, x_2$  の値を作成して、`x_vals` とします。等高線が粗い(点の数が少ない)場合や処理が重い(点の数が多)場合は、この設定を調整してください。

`expand.grid()` で  $\mathbf{x}$  の値(組み合わせ)を作成して、`x_points` とします。ただし、確率密度の計算のために、`as.matrix()` でデータフレームからマトリクスに変換しておきます。

平均ベクトル  $\boldsymbol{\mu}$  と分散共分散行列  $\boldsymbol{\Sigma}$  を作成します。

```
# 平均ベクトルを指定
mu_d <- c(0, 0)

# 分散共分散行列を指定
sigma_dd <- matrix(c(1, 0, 0, 1), nrow = 2, ncol = 2)
```

$\boldsymbol{\mu}$  の値を指定して `mu_d` とします。

$\boldsymbol{\Sigma}$  の値を指定して `sigma_dd` とします。`matrix()` に  $(\sigma_1^2, \sigma_{2,1}, \sigma_{1,2}, \sigma_2^2)$  の順に値を指定します。

設定したパラメータを使って確率密度を計算します。

```
# 2次元ガウス分布を計算
dens_df <- tibble(
  x1 = x_points[, 1], # x軸の値
  x2 = x_points[, 2], # y軸の値
  density = mvnfast::dmvn(X = x_points, mu = mu_d, sigma = sigma_dd) # 確率密度
)
head(dens_df)
```

```
## # A tibble: 6 x 3
##       x1      x2 density
##   <dbl> <dbl>   <dbl>
## 1  -5      -5 2.21e-12
## 2  -4.8    -5 5.89e-12
## 3  -4.6    -5 1.51e-11
## 4  -4.4    -5 3.71e-11
## 5  -4.2    -5 8.76e-11
## 6   -4     -5 1.99e-10
```

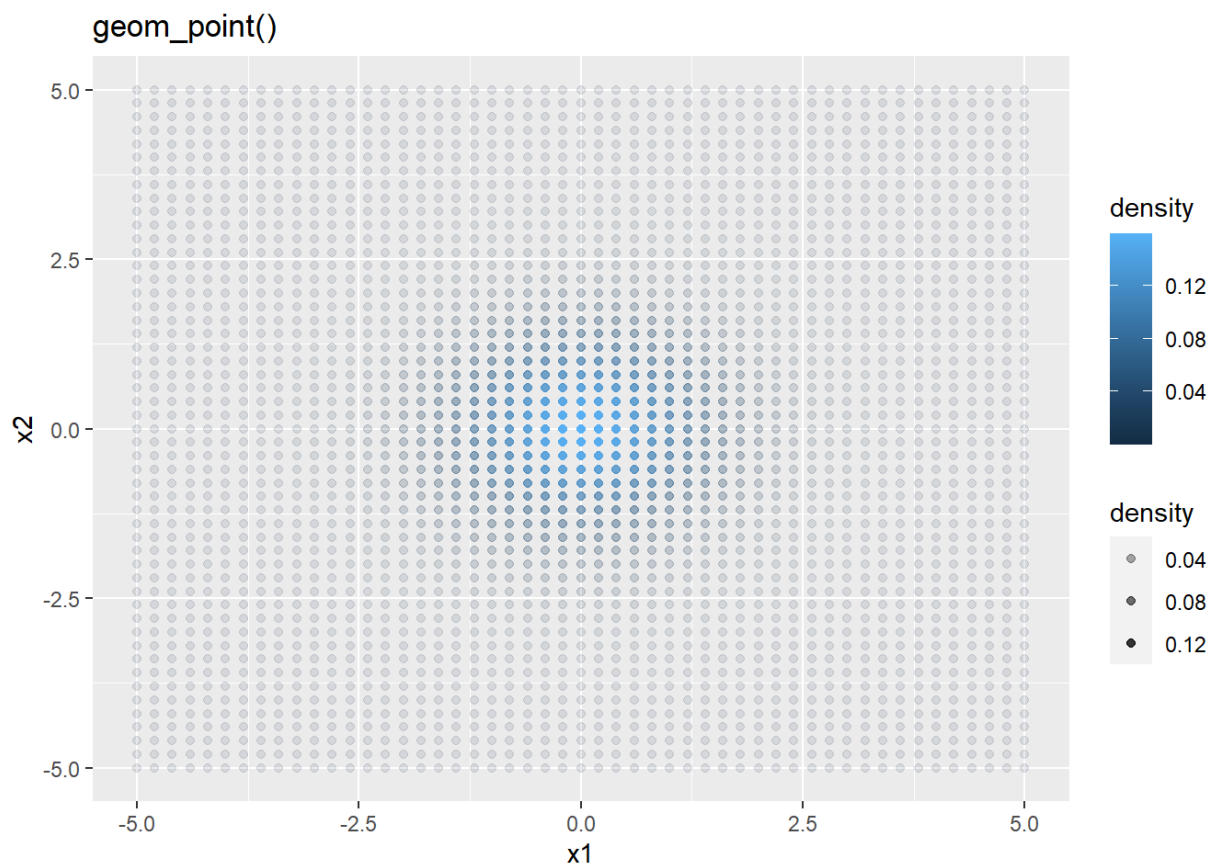
mvnfast パッケージの `dmvn()` で多次元ガウス分布の確率密度を計算できます。変数の引数  $x$  に `x_points`、平均ベクトルの引数 `mu` に `mu_d`、分散共分散行列の引数 `sigma` に `sigma_dd` を指定します。

$x_1, x_2$  の値と、対応する確率密度(計算結果)をデータフレームに格納します。

## 分布の作図

まずは、散布図で確認します。

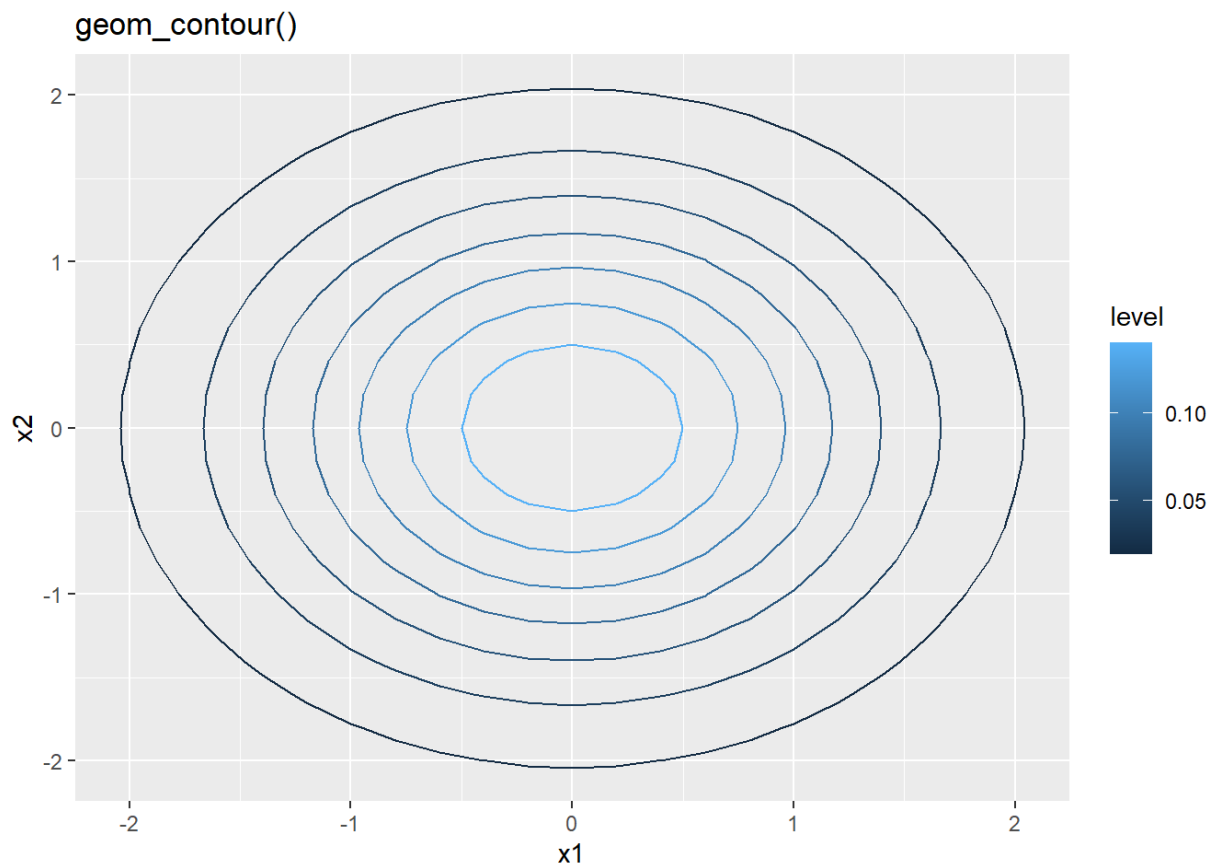
```
# 散布図を作成
ggplot() +
  geom_point(data = dens_df, mapping = aes(x = x1, y = x2, color = density, alpha = density)) + # 散布図
  labs(title = "geom_point()") # ラベル
```



色の引数 `color` に `density` 列を指定すると、点ごとに確率密度に応じた色が付きます。透過度の引数 `alpha` にも指定して、点ごとに濃淡を変更します。

では、等高線図を作成しましょう。

```
# 等高線図を作成
ggplot() +
  geom_contour(data = dens_df, mapping = aes(x = x1, y = x2, z = density, color = ..l
    evel..)) + # 等高線図
  labs(title = "geom_contour()") # ラベル
```



`geom_contour()` で等高線図を描画できます。

z軸の引数 `z` に確率密度( `density` 列)を指定します。

等高線の色を引数 `color` に `..level..` を指定すると、`z` 引数の値に応じて色付けされます。

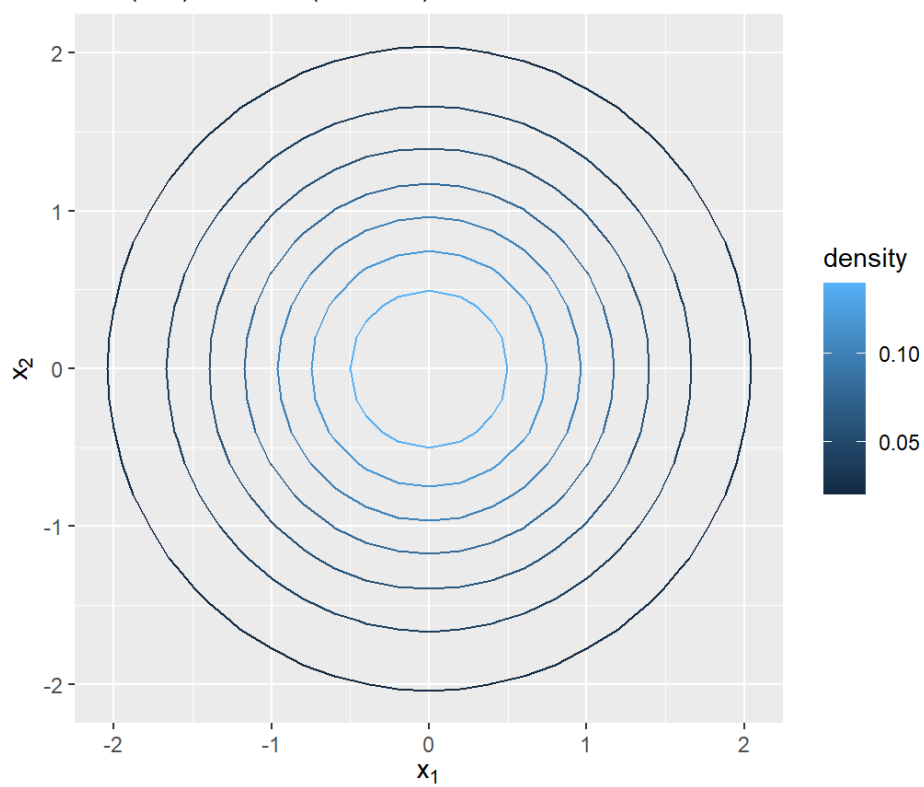
先ほどの散布図に関して、確率密度が同じ点を線で結んだイメージです。

最後に、色々装飾を施します。

```
# 2次元ガウス分布を作図
ggplot() +
  geom_contour(data = dens_df, mapping = aes(x = x1, y = x2, z = density, color = ..level..)) + # 等高線図
  coord_equal() + # アスペクト比
  labs(title = "Multivariate Gaussian Distribution",
        subtitle = paste0("mu=(", paste(mu_d, collapse = ", "), ")",
                           ", lambda=(", paste(sigma_dd, collapse = ", "), ")",
        x = expression(x[1]), y = expression(x[2]),
        color = "density") # ラベル
```

## Multivariate Gaussian Distribution

$\mu=(0, 0)$ ,  $\lambda=(1, 0, 0, 1)$



Enjoy!



