多元/多维高斯/正态分布概率密度函数推导 (Derivation of the Multivariate/Multidimensional Normal/Gaussian Density)

作者：凯鲁嘎吉 - 博客园 http://www.cnblogs.com/kailugaji/

当年在学《概率论与数理统计》时，遇到二元正态分布的概率密度函数，那个公式特别长，当时只是要求记住，并未深究其原因，今天终于有机会好好回顾一下了。二元/二维只是多元的一个特例，现在将问题延伸到多元/多维高斯/正态分布概率密度函数的推导上。多元高斯分布在很多场景下都有用，比如高斯混合模型(Gaussian Mixture Model)中，每个组件都是单个多元高斯分布，样本不仅是一维的，现实中大多是数据样本都是多维的。只有真正弄清楚公式的来龙去脉，来能更好的编写程序，进行实现(虽然很多包都是现成的，不需要自己从头编写)。想要推导概率密度函数公式，需要知道线性代数中矩阵论的一些基础知识，从单变量到二元/二维再延伸到多元/多维，本身就涉及到从标量到向量再到矩阵的一个过程。这篇博客详细推导了多元/多维高斯/正态分布概率密度函数公式，并应用到二维高斯分布中，进行进一步分析。也给出了当维度之间独立同分布(Independent identically distributed, i.i.d.)情况下多维高斯分布的概率密度函数的特例。值得注意的是，整个过程仅是对一个样本进行计算，该样本无论是一个标量，还是一个多维向量，最终出来的概率密度函数都是一个数(标量)。如果有N个样本(按列排开)的话，其概率密度函数就是N维列向量。注意：多元就是多维，高斯分布就是正态分布。(只是大概推导，过程可能并不严谨，望海涵)

1. 前提基础

包括连续随机变量变换法(Transformations of Continuous Random Variables)，单变量正态分布的概率密度函数(Univariate Normal Density)，以及随机变量间的独立性(Independence of Random Variables)。

➢ 前提基础

- 连续随机变量变换法
  (Transformations of Continuous Random Variables)

这里以二维随机变量为例，该方法可推广到多维连续随机变量上。设二维随机变量(X, Y)的联合密度函数为$p(x, y)$，若函数

$$\begin{cases} u = g_1(x,y), \\ v = g_2(x,y) \end{cases}$$

有连续偏导数，且存在唯一的反函数

$$\begin{cases} x = x(u,v), \\ y = y(u,v) \end{cases}$$

其变换的雅可比(Jacobian)行列式为

$$J = \frac{\partial(x,y)}{\partial(u,v)} = \det\begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{pmatrix} = \left(\frac{\partial(u,v)}{\partial(x,y)}\right)^{-1} = \left(\det\begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}\right)^{-1} \neq 0$$

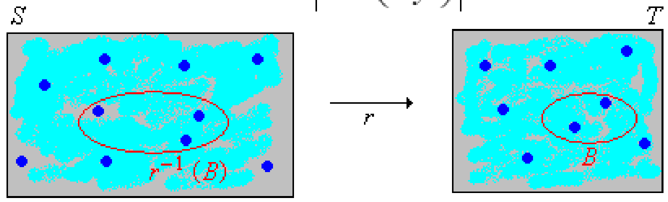
若 $\begin{cases} U = g_1(X,Y), \\ V = g_2(X,Y) \end{cases}$

则(U, V)的联合密度函数为

$$p(u,v) = p(x(u,v), y(u,v))\,|J|$$

- 定理：若X在S上有一个连续分布，其概率密度函数为f，则$Y=r(X)$在T上有一个连续分布，其概率密度函数g为

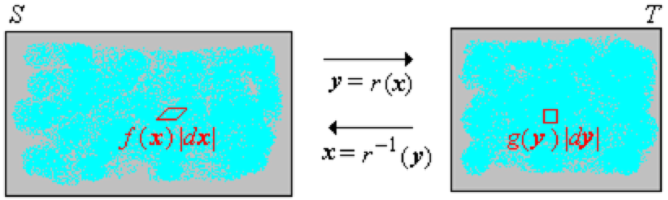
$$g(y) = f(x)\left|\det\left(\frac{dx}{dy}\right)\right|, y \in T.$$



- 简要论述：该定理仅仅是连续随机变量变换法的一维形式。已知$y=r(x)$有连续偏导数，且存在唯一的反函数$x=r^{-1}(y)$，其变换的雅可比行列式为

$$J = \det\left(\frac{dx}{dy}\right) = \left(\det\left(\frac{dy}{dx}\right)\right)^{-1}$$

则Y的概率密度函数为$g(y)=f(x)|J|$，定理得证。



茆诗松, 程依明, 濮晓龙. 概率论与数理统计教程. 高等教育出版社, 2011.
The Multivariate Normal Distribution http://www.randomservices.org/random/special/MultiNormal.html

多元/多维高斯/正态分布概率密度函数推导 (Derivation of the Multivariate/Multidimensional Normal/Gaussian Density)

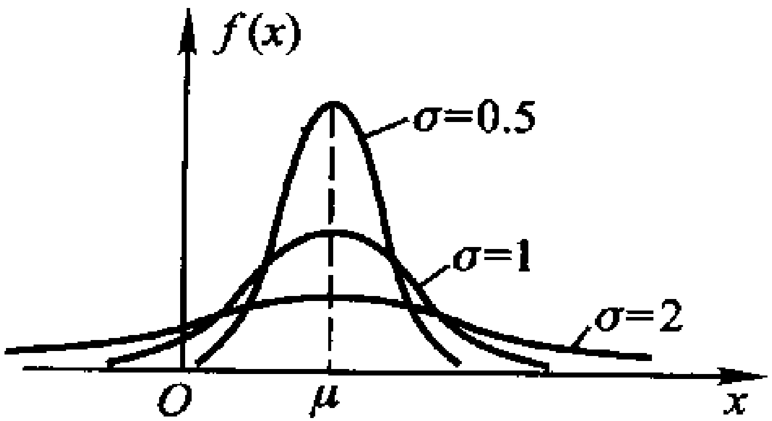
➢ 前提基础

- 单变量正态分布的密度函数(Univariate Normal Density)

若随机变量X的密度函数为

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<+\infty$$

则称X服从正态分布，称X为正态变量，记作$X\sim N(\mu,\ \sigma^2)$。



特别地，$X\sim N(0,\ 1)$，则

$$p(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},-\infty<x<+\infty$$

茆诗松, 程依明, 濮晓龙. 概率论与数理统计教程. 高等教育出版社, 2011.

- 随机变量间的独立性(Independence of Random Variables)

在连续随机变量场合，如果对任意D个实数 $x_1,x_2,\ldots,x_D$，都有

$$p(x_1,x_2,\ldots,x_D)=\prod_{d=1}^{D}p_d(x_d)$$

则称 $X_1,X_2,\ldots,X_D$ 相互独立。

若$Z=(Z_1,Z_2,\ldots,Z_D)^T\sim N(0,I)$，$D$个维度的变量之间相互独立，则Z的概率密度函数$g$为

$$g(z_1,z_2,\ldots,z_D)=\prod_{d=1}^{D}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z_d^2\right)$$

$$=\left(\frac{1}{\sqrt{2\pi}}\right)^D\prod_{d=1}^{D}\exp\left(-\frac{1}{2}z_d^2\right)=\left(\frac{1}{\sqrt{2\pi}}\right)^D\exp\left(-\frac{1}{2}\sum_{d=1}^{D}z_d^2\right)$$

$$=\frac{1}{(2\pi)^{D/2}}\exp\left(-\frac{1}{2}z^Tz\right)$$

其中 $z=(z_1,z_2,\ldots,z_D)^T$ $\quad (z_1,z_2,\ldots,z_D)\begin{pmatrix}z_1\\z_2\\\vdots\\z_D\end{pmatrix}=\sum_{d=1}^{D}z_d^2$

技巧：$e^ae^be^c=e^{a+b+c}$

2．多维高斯分布的概率密度函数定义及其推导

➢ 多维高斯分布的概率密度函数定义

设D维随机变量$X=(X_1,X_2,\dots,X_D)^T$的协方差矩阵为 $\Sigma=Cov(X)$，数学期望向量为 $\mu=(\mu_1,\mu_2,\dots,\mu_D)^T$.
又记$x=(x_1,x_2,\dots,x_D)^T$，则由密度函数

$$p(x)=p(x_1,x_2,\dots,x_D)=\frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

定义的分布称为D元正态分布，记为 $X\sim\mathcal{N}(\mu,\Sigma)$其中 $|\Sigma|$表示 Σ的行列式，Σ^{-1}表示 Σ的逆矩阵，
$(x-\mu)^T$表示向量$(x-\mu)$的转置. 注： Σ为对称正定矩阵 .
若记$\Sigma^{-1}=(\sigma_{ij}^{-1})$存在,则上式可写为

$$p(x)=p(x_1,x_2,\dots,x_D)=\frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}\sum_{i=1}^{D}\sum_{j=1}^{D}\sigma_{ij}^{-1}(x_i-\mu_i)(x_j-\mu_j)\right)$$

茆诗松, 程依明, 濮晓龙. 概率论与数理统计教程. 高等教育出版社, 2011.

➢ 多维高斯分布的概率密度函数推导

假设$X \sim \mathcal{N}(\mu,\Sigma)$，令$X=\mu+AZ$，其中$AA^T=\Sigma$，$A$是非奇异矩阵(由于$\Sigma$为正定阵)

$\det(\Sigma)=(\det(A))^2$. $Z=(Z_1,Z_2,\ldots,Z_D)^T$的概率密度函数为：

$\boxed{X \sim \mathcal{N}(\mu,\Sigma) \Rightarrow \frac{X-\mu}{\sqrt{\Sigma}} \sim \mathcal{N}(0,I)}$

$$g(z_1,z_2,\ldots,z_D)=\frac{1}{(2\pi)^{D/2}}\exp\left(-\frac{1}{2}z^Tz\right), z=(z_1,z_2,\ldots,z_D)^T$$

由$X=\mu+AZ \Rightarrow Z=A^{-1}(X-\mu) \Rightarrow$ 雅可比矩阵$J=\det\left(\frac{dZ}{dX}\right)=\det\left(A^{-1}\right)=\det\left(A\right)^{-1}$

随机变量变换法

$$f(x)=g(z)\left|\det\left(\frac{dz}{dx}\right)\right|=\left|(\det(A))^{-1}\right|g(A^{-1}(x-\mu))=\left|(\det(A))\right|^{-1}\frac{1}{(2\pi)^{D/2}}\exp\left(-\frac{1}{2}\left(A^{-1}(x-\mu)\right)^T\left(A^{-1}(x-\mu)\right)\right)$$

$$=\frac{1}{(2\pi)^{D/2}}(\det(\Sigma))^{-1/2}\exp\left(-\frac{1}{2}(x-\mu)^T(A^T)^{-1}A^{-1}(x-\mu)\right)=\frac{1}{(2\pi)^{D/2}}(\det(\Sigma))^{-1/2}\exp\left(-\frac{1}{2}(x-\mu)^T(AA^T)^{-1}(x-\mu)\right)$$

$$=\frac{1}{(2\pi)^{D/2}}(\det(\Sigma))^{-1/2}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

原式得证。　　技巧：$\left|A^{-1}\right|=\left|A\right|^{-1}, (AB)^{-1}=B^{-1}A^{-1}, (A^T)^{-1}=(A^{-1})^T$

Basic Multivariate Normal Theory http://www2.stat.duke.edu/~st118/sta732/mvnormal.pdf

3. 多维高斯分布的概率密度函数(维度之间独立同分布)

➢ 多维高斯分布的概率密度函数(维度之间独立同分布)

设D维$i.i.d$随机变量$X=(X_1,X_2,\ldots,X_D)^T$的协方差矩阵为对角阵 $\Lambda=(\sigma_d^2)$ 仅对角线上有元素，其余全是 0.

数学期望向量为$\mu=(\mu_1,\mu_2,\ldots,\mu_D)^T$ 又记$x=(x_1,x_2,\ldots,x_D)^T$ 则密度函数为

$$p(x)=p(x_1,x_2,\ldots,x_D)=\prod_{d=1}^{D}p(x_d)=\prod_{d=1}^{D}\frac{1}{\sqrt{2\pi}\sigma_d}\exp\left(-\frac{(x_d-\mu_d)^2}{2\sigma_d^2}\right)$$

$$=\left(\frac{1}{\sqrt{2\pi}}\right)^D\prod_{d=1}^{D}\frac{1}{\sigma_d}\exp\left(-\frac{1}{2}\sum_{d=1}^{D}\frac{(x_d-\mu_d)^2}{\sigma_d^2}\right)=\frac{1}{(2\pi)^{D/2}|\Lambda|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu)^T\Lambda^{-1}(x-\mu)\right)$$

$$(x-\mu)^T\Lambda^{-1}(x-\mu)=\left((x_1-\mu_1),(x_2-\mu_2),\ldots,(x_D-\mu_D)\right)\begin{pmatrix}\sigma_1^{-2}&0&0&0\\0&\sigma_2^{-2}&0&0\\0&0&\ddots&0\\0&0&0&\sigma_D^{-2}\end{pmatrix}\begin{pmatrix}x_1-\mu_1\\x_2-\mu_2\\\vdots\\x_D-\mu_D\end{pmatrix}$$

$$\Lambda=\begin{pmatrix}\sigma_1^{2}&0&0&0\\0&\sigma_2^{2}&0&0\\0&0&\ddots&0\\0&0&0&\sigma_D^{2}\end{pmatrix},|\Lambda|=\prod_{d=1}^{D}\sigma_d^2$$

4. 二维高斯分布的概率密度函数定义及其推导

➢ 二维高斯分布的概率密度函数

令$\rho = \dfrac{\sigma_{12}}{\sigma_1 \sigma_2}$

由多维高斯分布的概率密度函数定义，二维高斯分布的概率密度函数推导如下

$$p(x) = p(x_1, x_2) = \frac{1}{2\pi |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \right)$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \Rightarrow |\Sigma| = \sigma_1^2\sigma_2^2 - \sigma_{12}^2 = \sigma_1^2\sigma_2^2\left(1 - \boxed{\frac{\sigma_{12}^2}{\sigma_1^2\sigma_2^2}}\right) = \sigma_1^2\sigma_2^2\left(1 - \boxed{\rho^2}\right), \Sigma^{-1} = \frac{1}{|\Sigma|}\begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{pmatrix}$$

利用伴随矩阵求逆矩阵

$$\begin{aligned}
(x-\mu)^T \Sigma^{-1} (x-\mu) &= \left((x_1-\mu_1),(x_2-\mu_2)\right) \frac{1}{|\Sigma|}\begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{pmatrix}\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix} \\
&= \frac{1}{|\Sigma|}\left((x_1-\mu_1)\sigma_2^2 - (x_2-\mu_2)\sigma_{12}, -(x_1-\mu_1)\sigma_{12} + (x_2-\mu_2)\sigma_1^2\right)\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix} \\
&= \frac{1}{|\Sigma|}\left((x_1-\mu_1)^2\sigma_2^2 - (x_1-\mu_1)(x_2-\mu_2)\sigma_{12} - (x_1-\mu_1)(x_2-\mu_2)\sigma_{12} + (x_2-\mu_2)^2\sigma_1^2\right) \\
&= \frac{(x_1-\mu_1)^2\sigma_2^2 - 2(x_1-\mu_1)(x_2-\mu_2)\sigma_{12} + (x_2-\mu_2)^2\sigma_1^2}{\sigma_1^2\sigma_2^2\left(1-\rho^2\right)}
\end{aligned}$$

➢ 二维高斯分布的概率密度函数

其中 $\rho = \dfrac{\sigma_{12}}{\sigma_1 \sigma_2}$

因此，二维高斯分布的概率密度函数为

$$p(x) = p(x_1, x_2) = \frac{1}{2\pi\left(\sigma_1^2\sigma_2^2\left(1-\rho^2\right)\right)^{1/2}} \exp\left(-\frac{(x_1-\mu_1)^2\sigma_2^2 - 2(x_1-\mu_1)(x_2-\mu_2)\sigma_{12} + (x_2-\mu_2)^2\sigma_1^2}{2\sigma_1^2\sigma_2^2\left(1-\rho^2\right)}\right)$$

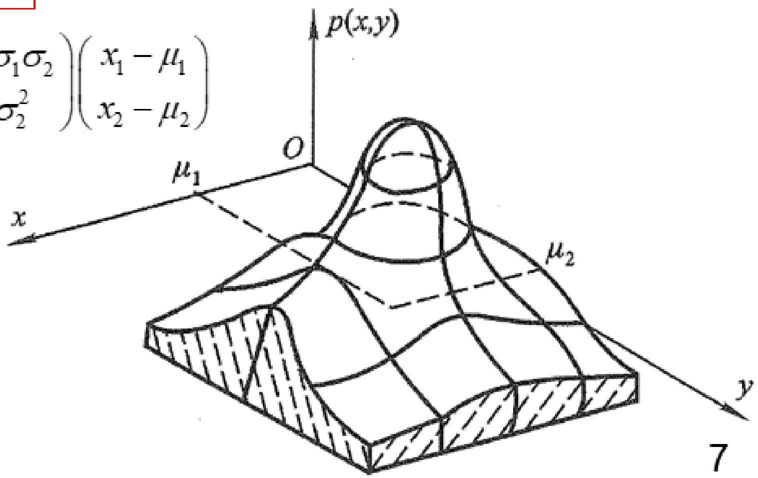
上下同除以 $\sigma_1^2\sigma_2^2$

$$= \boxed{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\frac{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2}{1-\rho^2}\right)}$$

其中 $\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 = \left((x_1-\mu_1),(x_2-\mu_2)\right)\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\begin{pmatrix}x_1-\mu_1 \\ x_2-\mu_2\end{pmatrix}$

$$\therefore \Sigma = \begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}, \sigma_{12} = \mathrm{cov}(X_1, X_2) = \rho\sigma_1\sigma_2$$

$$corr(X_1, X_2) = \frac{\mathrm{cov}(X_1, X_2)}{\sigma_1\sigma_2} = \rho(\rho\text{就是相关系数})$$



茆诗松, 程依明, 濮晓龙. 概率论与数理统计教程. 高等教育出版社, 2011.

5. 参考文献

[1] 茆诗松, 程依明, 濮晓龙. 概率论与数理统计教程. 高等教育出版社, 2011.

[2] The Multivariate Normal Distribution http://www.randomservices.org/random/special/MultiNormal.html

[3] Basic Multivariate Normal Theory http://www2.stat.duke.edu/~st118/sta732/mvnormal.pdf

[4] 凯鲁嘎吉 - 博客园 - 左边栏搜索"高斯"相关博文 https://zzk.cnblogs.com/my/s/blogpost-p?Keywords=%E9%AB%98%E6%96%AF