# 指数分布

服从参数为θ的指数分布

概率密度：

分布函数：

无记忆性

某元件已经使用了s小时，它总共能使用至少s+t小时的概率与从开始算起它能使用t小时的概率相等，也就是元件对他已经使用过s小时这件事没有记忆

# 自然分布

又称高斯分布

概率密度：

也就说明了x=μ是函数的对称轴，所以μ的位置决定了整个函数的位置，又称为位置参数，而σ越小函数形状越尖锐

3σ准则：尽管x的取值是无穷，但实际上几乎都会落在(μ-3σ,μ+3σ)的范围内

在机器学习中，当我们sample样本点的时候，μ决定了sample的中心位置，σ决定了样本点出现在中心位置的概率，也可以理解为σ越小，我们sample的点的数值范围越集中

# 随机变量的函数的分布

设随机变量X的概率密度<x<，又设函数g(x)处处可导且恒有,则Y=g(x)是连续型随机变量，其概率密度为

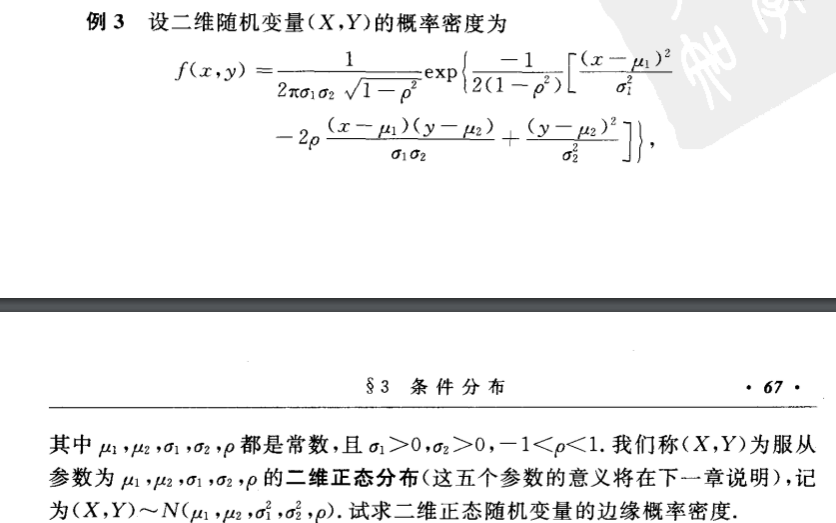
# 边缘分布

二维随机变量(X, Y)作为一个整体，其分布函数为F(x, y)， 而x, y也有各自的分布函数，设连续型随机变量(X, Y)的概率密度为f(x, y)，其边缘概率密度如下：

即为(X,Y)关于X和关于Y的边缘概率密度

单由关于X和关于Y的边缘分布，一般来说不能确定随机变量X和Y的联合分布

举例如下：



# 条件概率密度

设二维随机变量(X,Y)的概率密度为f(x,y),关于Y的边缘概率密度为,若对于固定的y，其边缘概率密度大于0，则在Y=y的条件下X的条件概率密度记为：

而Y=y的条件下X的条件分布函数为：

# 两个随机变量的函数的分布

设(X,Y)是二维连续型随机变量，它具有概率密度f(x,y)，则Z=X+Y仍然为连续型随机变量，其概率密度为

如果X和Y相互独立，设(X,Y)关于X和Y的边缘密度分别为和，则上述公式可变为

这两个公式称为和的卷积公式，记为\*

# 相互独立的随机变量

设()和 (相互独立，则和相互独立；如果h和g是连续函数，则h()和 g(相互独立

# 两个随机变量的函数的分布

Z=X+Y

设（X,Y）是二维连续型随机变量，它具有概率密度f(x,y)，则Z=X+Y仍然为连续型随机变量，其概率密度为

Z=Y/X

概率密度为

Z=XY

概率密度为

M=max{X, Y}

由于M=max{X, Y}不大于z等价于X和Y都不大于z，所以

又由于X和Y相互独立，得到M=max{X,Y}的分布函数为

# 期望

设离散型随机变量X的分布律为

如果级数绝对收敛，就说这个级数的和是随机变量X的数学期望：

而对于连续型随机变量X的概率密度为f(x)，如果积分绝对收敛，责成积分的值为随机变量X的数学期望：

数学期望简称为期望，也称为均值

设Y是随机变量X的函数：Y=g(X)（g是连续型随机变量）

如果X是离散型随机变量，它的分布律是如果绝对收敛，则有：

如果X是连续型随机变量，它的概率密度为f(x)，若绝对收敛，则有：

此公式的意义在于在求E(Y)的时候不必算出Y的分布律或概率密度

# 方差

设X是一个随机变量，若存在，则称为X的方差，记做D(X)或Var(X)，即

随机变量X的方差表达了X的取值与其数学期望的偏离程度，若D(X)较小意味着X的取值比较集中在E(X)附近，反之说明X的取值较为分散

# 协方差

对于二维随机变量称为随机变量X和Y的协方差，记做Cov（X,Y）

即

称为随机变量X和Y的相关系数