Portfolio Construction and Analytics 读书笔记（第三章）

1　资产管理的介绍

2　随机变量、概率分布和重要的统计概念

3　常见的分布函数介绍

3.1　分布函数的样例

3.1.1记号说明

Gama 函数的定义如下：

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x}\, dx$$

Beta函数的定义如下：

$$\mathrm{B}(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1}\, dt$$

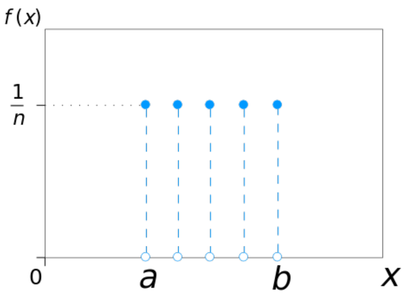
3.1.2离散型和连续型均匀分布

离散型均匀分布：

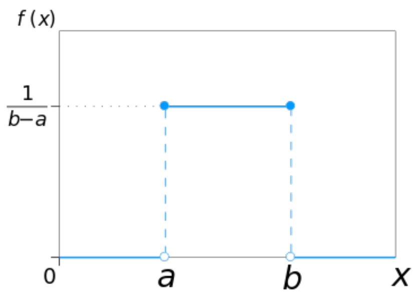
$$\Pr(X = x) = \frac{1}{N}$$

连续型均匀分布：

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b, \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$$



3-1 离散型均匀分布



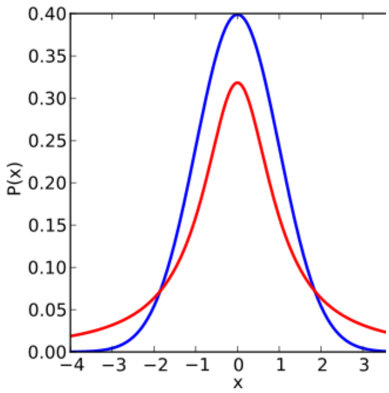
3-2 连续型均匀分布

3.1.3t分布

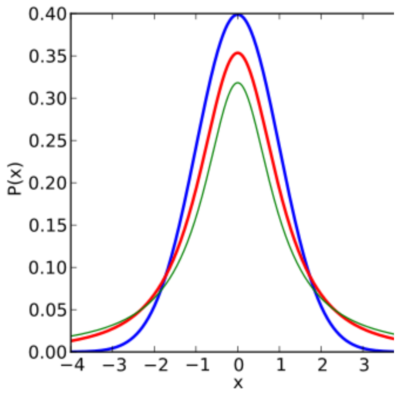
密度函数:

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})} \left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

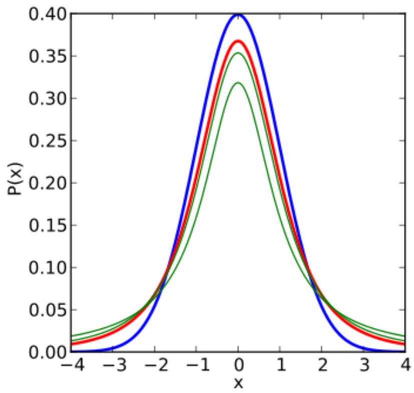
自由度分别为1，2，3的t分布的密度函数。其中蓝色的线条表示正态分布的密度函数。绿色的线条表示上一幅图中t分布的密度函数。



3-3 自由度为1



3-4 自由度为2



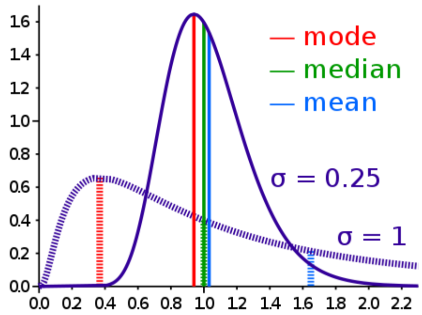
3-5 自由度为3

3.1.4对数正态分布

对数正态分布（logarithmic normal distribution）是指一个随机变量的对数服从正态分布，则该随机变量服从对数正态分布。对数正态分布从短期来看，与正态分布非常接近。但长期来看，对数正态分布向上分布的数值更多一些。

密度函数：

$$\begin{aligned} f_X(x) &= \frac{\mathrm{d}}{\mathrm{d}x}\Pr(X \le x) = \frac{\mathrm{d}}{\mathrm{d}x}\Pr(\ln X \le \ln x) \\ &= \frac{\mathrm{d}}{\mathrm{d}x}\Phi\left(\frac{\ln x-\mu}{\sigma}\right) \\ &= \varphi\left(\frac{\ln x-\mu}{\sigma}\right)\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\ln x-\mu}{\sigma}\right) \\ &= \varphi\left(\frac{\ln x-\mu}{\sigma}\right)\frac{1}{\sigma x} \\ &= \frac{1}{x}\cdot\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right). \end{aligned}$$

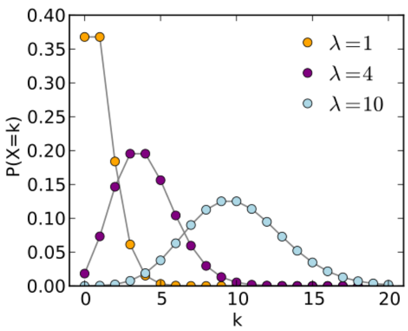


3-6 两种不同偏度的对数正态分布

3.1.5泊松分布

密度函数：

$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!}$$



3-7 不同λ值的泊松分布

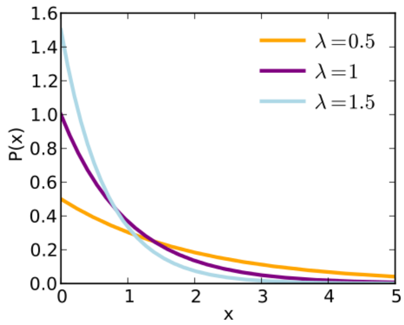
值得一提的是，随着λ的递增，泊松分布会不断逼近均值为λ，方差为$\sqrt{\lambda}$的正态分布，这一结论在金融模型中有重要的应用。

3.1.6指数分布

在概率理论和统计学中，指数分布（也称为负指数分布）是描述泊松过程中的事件之间的时间的概率分布，即事件以恒定平均速率连续且独立地发生的过程。这是伽马分布的一个特殊情况。

密度函数：

$$f(x;\lambda)=\begin{cases}\lambda e^{-\lambda x} & x\geq 0,\\ 0 & x<0.\end{cases}$$



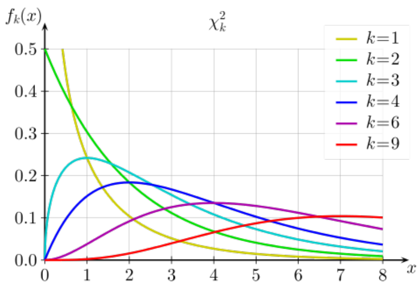
3-8 不同参数的指数分布

3.1.7卡方分布

若n个相互独立的随机变量$\xi_1, \xi_2, ..., \xi_n$，均服从标准正态分布（也称独立同分布于标准正态分布），则这n个服从标准正态分布的随机变量的平方和构成一新的随机变量，其分布规律称为卡方分布（chi-square distribution）。

密度函数：

$$f(x;\,k) = \begin{cases} \dfrac{x^{\frac{k}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)}, & x > 0; \\ 0, & \text{otherwise}. \end{cases}$$



3-9 不同参数的卡方分布

3.1.8伽玛分布

伽玛分布（Gamma Distribution）是统计学的一种连续概率函数，是概率统计中一种非常重要的分布。”指数分布”和”χ^2分布”都是伽马分布的特例。

Gamma分布中的参数α称为形状参数（shape parameter），β称为尺度参数（scale parameter）。

密度函数：

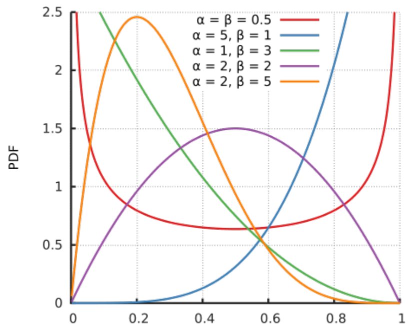
$$f(x;\alpha,\beta) = \frac{\beta^\alpha x^{\alpha-1}e^{-\beta x}}{\Gamma(\alpha)} \quad \text{for } x > 0 \text{ and } \alpha,\beta > 0,$$

3.1.9贝塔分布

在概率论中，贝塔分布，也称 B 分布，是指一组定义在(0,1) 区间的连续概率分布。

密度函数：

$$
\begin{aligned}
f(x;\alpha,\beta) &= \mathrm{constant}\cdot x^{\alpha-1}(1-x)^{\beta-1} \\
&= \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\,x^{\alpha-1}(1-x)^{\beta-1} \\
&= \frac{1}{\mathrm{B}(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}
\end{aligned}
$$



3-10 不同参数的贝塔分布

3.2　金融回报率的分布模型

3.2.1椭圆分布族

椭圆分布的密度函数具有如下形式：

$$
f(x) = \frac{c}{\sqrt{|\Sigma|}}\cdot g((x-\mu)'\Sigma^{-1}(x-\mu))
$$

例如，我们所熟知的多元正态分布就属于椭圆分布族：

$$
f_{\mathbf{x}}(x_1,\ldots,x_k) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}}
$$

3.2.2稳定Paretian分布族

稳定Paretian分布族主要包括如下三个分布：正态分布，柯西分布，列维分布。例

如我们所熟知的柯西分布，其密度函数为:

$$f(x) = \frac{1}{\pi\gamma\left[1+\left(\frac{x-\mu}{\gamma}\right)^2\right]} = \frac{1}{\pi\gamma}\left[\frac{\gamma^2}{(x-\mu)^2+\gamma^2}\right],$$

以及列维分布，其密度函数为：

$$f(x;\mu,\gamma) = \sqrt{\frac{\gamma}{2\pi}}\;\frac{e^{-\frac{\gamma}{2(x-\mu)}}}{(x-\mu)^{3/2}}$$

我们把μ称为位置参数，而把γ称为尺度参数。

3.2.3广义λ分布族

Tukey λ分布：

$$F^{-1}(p) = Q\left(p;\lambda\right) = \begin{cases} \frac{1}{\lambda}\left[p^\lambda-(1-p)^\lambda\right], & \text{if } \lambda \neq 0 \\ \log(\frac{p}{1-p}), & \text{if } \lambda = 0, \end{cases}$$

Tukeyλ分布可以近似一些常见的分布：

λ = -1:　接近柯西分布$C(0,\Pi)$

λ = 0:　logistic分布

λ = 0.14:　接近正态分布$N(0, 2.142)$

λ = 1:　均匀分布$U(-1, 1)$

利用Tukey分布我们可以定义广义λ分布(GLD)：

$$F^{-1}(p) = Q\left(p;\lambda\right) = \lambda_1 + \frac{1}{\lambda_2}\left[p_3^\lambda - (1-p)_4^\lambda\right]$$

上述分布的VaR和CVaR都是容易计算的。

3.3　金融回报率的尾部风险模型

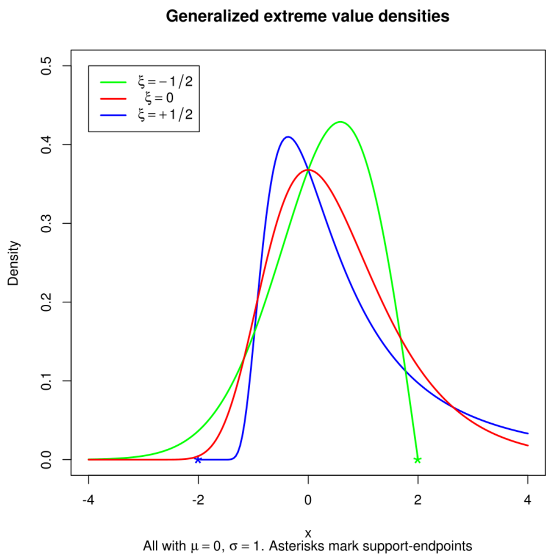
3.3.1广义极值分布

采用标准化的方法：

$$s = (x-\mu)/\sigma$$

可以得到：标准广义极值分布的密度函数：

$$f(s;\xi) = \begin{cases} (1+\xi s)^{(-1/\xi)-1}\exp(-(1+\xi s)^{-1/\xi}) & \xi \neq 0 \\ \exp(-s)\exp(-\exp(-s)) & \xi = 0 \end{cases}$$



3-11 不同参数下的极值分布

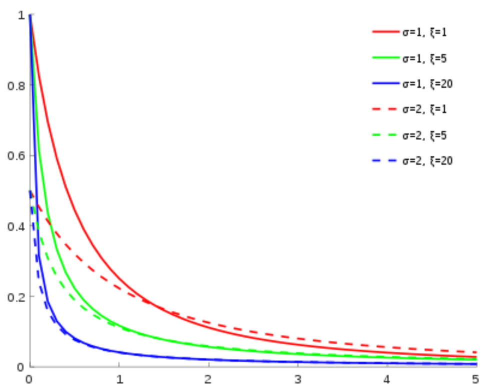
3.3.2广义帕累托分布

同样的采用标准化的方法：

$$s = (x - \mu)/\sigma$$

可以得到：标准广义帕累托分布的密度函数：

$$f_{(\xi,\mu,\sigma)}(x) = \frac{1}{\sigma}\left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{\left(-\frac{1}{\xi}-1\right)},$$



3-12 不同参数下的帕累托分布

3.3.3极值模型

为了拟合GDP模型，我们通常采用极大似然法估计参数（MLE）一旦GDP的参数

估计完成以后，我们就可以计算模型的VaR和CVaR了：

$$(100-\epsilon)\%VaR = u + \frac{\theta}{\xi}((\frac{n}{N_u}(1-\frac{\epsilon}{100}))^{-\xi}-1)$$