

R4 常见概率分布

七种分布——均匀分布、二项分布、正态分布、对数正态分布、学生t分布、卡方分布和F分布——广泛用于投资分析。

学习目标

1. 定义概率分布并比较和对比离散和连续随机变量及其概率函数
2. 给定累积分布函数，计算和解释随机变量的概率
3. 描述离散均匀随机变量的性质，并计算和解释给定离散均匀分布函数的概率
4. 描述连续均匀分布的性质，并计算和解释给定连续均匀分布的概率
5. 描述伯努利随机变量和二项式随机变量的性质，并计算和解释给定二项式分布函数的概率
6. 解释正态分布的关键性质
7. 对比多元分布和单变量分布，并解释相关性在多元正态分布中的作用

8. 计算正态分布随机变量位于给定区间内的概率
9. 解释如何标准化随机变量
10. 使用标准正态分布计算和解释概率
11. 定义短缺风险，计算安全第一比率，并使用 Roy 的安全第一标准确定最佳投资组合
12. 解释正态分布和对数正态分布之间的关系以及为什么使用对数正态分布来模拟资产价格
13. 在给定特定持有期回报的情况下，计算和解释连续复合回报率
14. 描述学生t分布的性质，并计算和解释其自由度
15. 描述卡方分布和F分布的性质，并计算和解释它们的自由度
16. 描述蒙特卡罗模拟

介绍离散随机变量

- 正态分布和二项分布用于诸如 Black-Scholes-Merton 期权定价模型、二项期权定价模型和资本资产定价模型等基本估值模型。
- 学生的t -、卡方和F-分布用于验证统计显着性和假设检验。
- random variable随机变量：一个未来结果不确定的量
- discrete random variable离散随机变量：最多可以取可数（可能无限）数量的可能值。
- continuous random variable连续随机变量：结果的范围是实线(所有在 $-\infty$ 和 $+\infty$ 之间的实数)或实线的某个子集。
- probability function概率函数：一种指定随机变量取某一特定值的概率的函数。
- probability density function (pdf)概率密度函数：一种具有非负值的函数，这样概率就可以用函数曲线下的区域来描述。
- cumulative distribution function累积分布函数：给出随机变量小于或等于指定值的

概率的函数

离散和连续均匀分布

- 离散均匀随机变量的Probability Function $p(x) = P(X = x)$
- 离散均匀随机变量的Cumulative Distribution Function $F(x) = P(X \leq x)$
- 连续均匀随机变量的概率分布函数 pdf 是

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

- 连续均匀随机变量的累积概率分布cdf 是

$$F(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } a \leq x \leq b \\ 1 & \text{for } x > b \end{cases}$$

二项分布

- 二项分布的构建块是伯努利随机变量Bernoulli random variable
- 一个试验（一个可能重复的事件）产生两个结果之一。这样的试验就是伯努利试验Bernoulli trial
- binomial random variable二项随机变量
- $X \sim B(n, p)$: X 具有参数 n 和 p 的二项分布，均值是 np ，方差 $np(1 - p)$
- n 次试验中 x 次成功的概率

$$p(x) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}$$

- binomial tree: 资产价格动态模型的图形表示，在每个时期，资产以概率 p 上升或以概率 $(1 - p)$ 下降。
- up transition probability 向上转移概率； down transition probability 向下转移概率

正态分布

- 中心极限定理指出，大量独立随机变量（具有有限方差）的总和（和均值）近似正态分布。
- $X \sim N(\mu, \sigma^2)$: X服从均值 μ 和方差 σ^2 的正态分布。
- skewness偏度为 0（它是对称的）。正态分布的
- kurtosis峰度为 3
- excess kurtosis超峰态 (kurtosis - 3.0) 等于 0
- 密度函数

$$f(x) = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \text{ for } -\infty < x < \infty$$

- 正态分布随机变量位于给定区间内的概率:
 1. 大约 50% 的观测值落在 $\mu \pm (2/3)\sigma$ 区间内。
 2. 大约 68% 的观测值落在区间 $\mu \pm \sigma$ 中。
 3. 大约 95% 的观测值落在 $\mu \pm 2\sigma$ 区间内。
 4. 大约 99% 的观测值落在 $\mu \pm 3\sigma$ 区间内。
- 标准化standardizing正态分布

$$Z = \frac{X - \mu}{\sigma}$$

- central limit theorem中心极限定理：一组具有有限方差的独立同分布随机变量的和(和均值)是正态分布的定理，无论随机变量遵循何种分布

正态分布的应用

- mean–variance analysis均值方差分析：一种利用资产收益的预期均值、方差和协方差进行投资组合分析的方法。
- Safety-first rules安全第一规则：关注的是差额风险，即投资组合价值(或投资组合回报)在一段时间内低于某个最低可接受水平的风险。
- safety-first安全第一最优投资组合使安全第一比率safety-first ratio(SFRatio)最大化
$$SFRatio = \frac{[E(R_P) - R_L]}{\sigma_P}$$
- expected portfolio return $E(R_P)$
- a level of RL as unacceptable R_L
- SFRatio 与夏普比率的相似之处:用无风险利率 R_F 代替临界水平 R_L ,SFRatio 就变成了夏普比率

对数正态分布与连续复合

- lognormal distribution对数正态分布以0为界并且向右倾斜（它有一个长的右尾）
- 正态随机变量 X 有均值 μ 和方差 σ^2 . 定义 $Y = \exp(X)$ 是对数正态分布
- Mean of a lognormal random variable $\mu_L = \exp(\mu + 0.50\sigma^2)$.
- Variance of a lognormal random variable $\sigma_L^2 = \exp(2\mu + \sigma^2) * [\exp(\sigma^2) - 1]$.
- Continuously Compounded Rates of Return连续复合回报率
- price relative相对价格：期末价格与期初价格之比;它等于1加上资产的持有期回报率。
- 连续复合收益从 t 到 $t + 1$ 是 $r_{t,t+1} = \ln(S_{t+1}/S_t) = \ln(1 + R_{t,t+1})$.
- Volatility挥发性衡量标的资产连续复利的标准差

Student's t-, Chi-Square, and F-Distributions

- t分布具有“长尾”，它可以提供更可靠、更保守的下行风险估计。
- 假设样本来自正态分布。The ratio $z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ 服从正态分布均值为0标准差为1;ratio $t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$ 服从t分布均值为0自由度 $n - 1$.
- Chi-Square卡方分布是不对称的
- 自由度为 k 的卡方分布是 k 个独立标准正态分布随机变量的平方和的分布
- F分布是一系列以 0 为界的非对称分布。 $F = (\chi_1^2/m)/(\chi_2^2/n)$ 具有 m 分子和 n 分母自由度的 F 分布

Monte Carlo Simulation蒙特卡洛模拟

- 蒙特卡罗模拟的一个特征是从一个或多个指定的概率分布中生成大量随机样本来表示风险在系统中的作用。
- contingent claim或有债权：一种价值基于其他基础证券的证券。
- inverse transformation method逆变换：来自均匀分布的随机数 (PDF) 映射到标准正态分布的CDF和PDF以产生随机观察