期望和方差

主讲人：刘宏志
liuhz@ss.pku.edu.cn

北京大学

期望的起源

- 1654年，一赌徒向法国大数学家帕斯卡提出了如下问题：
  - 甲乙两人赌技相当，各自赌金100法郎，并约定先胜3局者为胜，取得全部200法郎。
  - 由于出现意外情况，在甲胜2局乙胜1局时，不得不终止赌博，如果要分赌金，该如何分配才算公平？

分法1：把钱分为3份，胜了2局的拿2份，胜了1局的拿1份

分法2：约定是胜3局者为胜，而谁都没达到，所以应一人一半

期望的起源

- 1654年，一赌徒向法国大数学家帕斯卡提出了如下问题：
  - 甲乙两人赌技相当，各自赌金100法郎，并约定先胜3局者为胜，取得全部200法郎。
  - 由于出现意外情况，在甲胜2局乙胜1局时，不得不终止赌博，如果要分赌金，该如何分配才算公平？
- 帕斯卡和费马通信讨论了此问题，并于1654年共同建立了概率论的第一个基本概念：数学期望

- 不考虑平局， 最多5局就能分出胜负（一方胜3局，另一方胜2局）
- 已经进行了3局，最多还需进行2局
- 甲胜的概率：$1/2 + 1/2 * 1/2 = 3/4$，乙胜的概率： $1/2 * 1/2 = 1/4$

期望的起源

- 甲胜的概率为3/4，乙胜的概率为1/4
- 甲乙获胜的可能性大小比值为3:1
- 甲应获得赌金的3/4，而乙只能获得赌金的1/4
- 甲能“期望”获得数目为：
  - $200 * 3/4 + 0 * 1/4 = 150$ (法郎)
- 乙能“期望”获得数目为：
  - $200 * 1/4 + 0 * 3/4 = 50$ (法郎)

期望的定义

- 离散型随机变量
  - 如果X是频率函数为$p(x)$的离散型随机变量，且满足$\sum_i |x_i| p(x_i) < \infty$，则X的期望为：

$$\mathrm{E(X)} = \sum_i x_i p(x_i)$$

  - 如果和式发散，则期望无定义

- 连续型随机变量
  - 如果X是密度函数为$f(x)$的连续型随机变量，且满足$\int_{-\infty}^{\infty} |x| f(x)\,\mathrm{d}x < \infty$，则X的期望为：

$$\mathrm{E(X)} = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x$$

  - 如果积分发散，则期望无定义

案例：圣彼得堡悖论

- 一个赌徒按照如下策略赌博：
  - 开始下注1美元，如果输了，就接着双倍下注
  - 连续双倍下注，直到最终获胜
  - 当他最终获胜时，他稳赢1美元
- 此策略是否合理？
- 令X表示最后一局（获胜的一局）的赌注
  - $P(X=2^k)=1/2^{k+1}$
  - $E(X)=\sum_{n=0}^{\infty} nP(X=n)=\sum_{k=0}^{\infty} 2^k \frac{1}{2^{k+1}}=\infty$

马尔可夫不等式

- 定理：如果随机变量X满足$P(X \geq 0)=1$，且$E(X)$存在，则对任意$t>0$,

$$P(X \geq t) \leq E(X)/t \text{ 。}$$

- 证明：$E(X)=\sum_{x} xp(x)=\sum_{x<t} xp(x)+\sum_{x \geqslant t} xp(x)$

$$E(X) \geqslant \sum_{x \geqslant t} xp(x) \geqslant \sum_{x \geqslant t} tp(x)=tP(X \geqslant t)$$

- 变形：令$t=kE(X)$，则$P(X \geq kE(X)) \leq 1/k$
- 应用：
  - 不超过1/5的人口会有超过5倍于人均收入的收入

随机变量函数的期望

- 假设$Y=g(X)$.
  - 如果X是频率函数为$p(x)$的离散型随机变量，且满足$\sum|g(x)|p(x)<\infty$，则：
  $$E(Y)=\sum_x g(x)p(x)$$
  - 如果X是密度函数为$f(x)$的连续型随机变量，且满足$\int_{-\infty}^{\infty}|g(x)|f(x)\,\mathrm{d}x<\infty$，则：
  $$E(Y)=\int_{-\infty}^{\infty}g(x)f(x)\,\mathrm{d}x$$

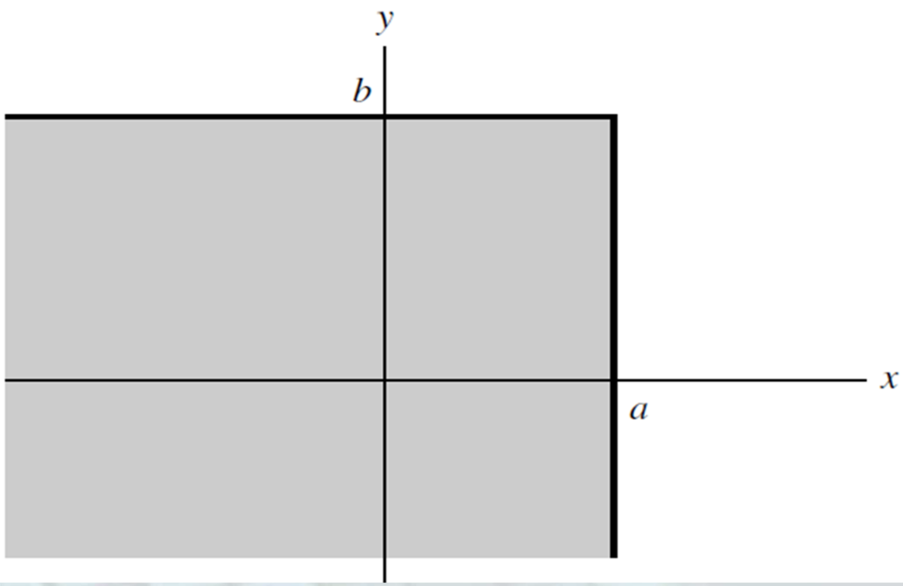
注意：$E[g(X)]\neq g[E(X)]$，即函数的期望不等于期望的函数

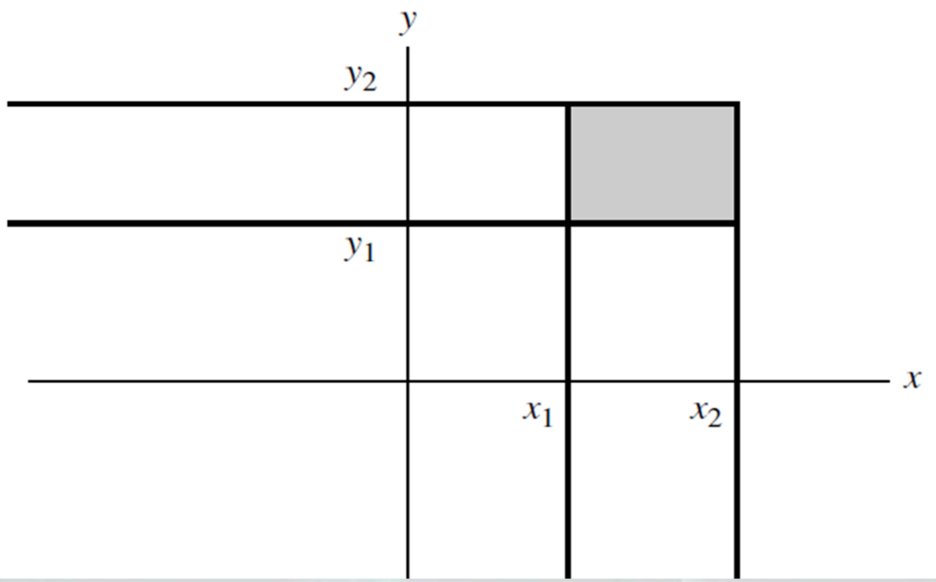
联合分布函数的期望

- 设$X_1, \ldots, X_n$是具有联合分布的随机变量, $Y=g(X_1, \ldots, X_n)$.
  - 如果X是频率函数为$p(x_1, \ldots, x_n)$的离散型随机变量，且满足$\sum |g(x_1, \ldots, x_n)| p(x_1, \ldots, x_n) < \infty$，则：
$$E(Y) = \sum g(x_1, \ldots, x_n) p(x_1, \ldots, x_n)$$
  - 如果X是密度函数为$f(x_1, \ldots, x_n)$的连续型随机变量，且满足$\int \int \cdots \int |g(x_1, \ldots, x_n)| f(x_1, \ldots, x_n)\, \mathrm{d}x_1, \ldots, \mathrm{d}x_n < \infty$，则：
$$E(Y) = \int \int \cdots \int g(x_1, \ldots, x_n) f(x_1, \ldots, x_n) \mathrm{d}x_1, \ldots, \mathrm{d}x_n$$

联合分布

- 定义在同一样本空间上的两个或两个以上的随机变量的联合概率结构
- 两个随机变量X和Y的联合性质由累积分布函数(cdf)决定: $F(x, y) = P(X \le x, Y \le y)$
  - 累积分布函数给出了点(X, Y)落入平面半无穷矩形区域的概率





联合频率函数

- 假设X和Y是定义在同一样本空间上的两个离散随机变量，分别取值$x_1, x_2, \ldots$和$y_1, y_2, \ldots$
- 它们的联合频率函数或联合概率质量函数为

$$p(x_i, y_j) = P(X=x_i, Y=y_j)$$

- 例：抛掷一枚均匀硬币3次，令X表示第一次抛掷是否正面朝上，Y表示所有正面的次数
  - 样本空间为：$\Omega=\{hhh, hht, hth, htt, thh, tht, tth, ttt\}$

x \ y	0	1	2	3
0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0
1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

随机变量线性组合的期望

- 定理：如果$X_1, \ldots, X_n$是具有期望$E(X_i)$的联合分布随机变量，Y是X_i的的线性函数 $Y=a+\sum_{i=1}^{n} b_i X_i$, 则：

$$E(Y)=a+\sum_{i=1}^{n} b_i E(X_i)$$

- 例：假设你收集购物券，共有n种不同类型的购物券，且每次试验你都可以等可能地得到任一类型的购物券。你期望多少次试验才能收集到完整的券集？
  - 令X_{i+1}表示在收集了i种购物券后出现第$i+1$类购物券所需的次数
  - X_r的分布为：几何分布
  - 每次试验成功的概率为：$p=(n-r+1)/n$
  - $E(X_r)=1/p = n/(n-r+1)$

$$
\begin{aligned}
E(X) &= \sum_{r=1}^{n} E(X_r) \\
&= \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + \frac{n}{1} \\
&= n \sum_{r=1}^{n} \frac{1}{r}
\end{aligned}
$$

几何分布：由无穷次伯努利试验构造而成，每次试验成功的概率为p，X表示直到第一次成功所做的试验次数

方差和标准差

- 随机变量的标准差描述分布关于中心的发散程度，度量随机变量偏离期望的平均幅度
- 定义：如果X是具有期望$E(X)$的随机变量，只要下述期望存在，则X的方差为：

$$\mathrm{Var}(X) = E\{[X-E(X)]^2\}$$

- X的标准差是方差的平方根
- 定理1：如果$\mathrm{Var}(X)$存在，$Y=a+bx$, 则

$$\mathrm{Var}(Y)=b^2\mathrm{Var}(X)$$

- 定理2：如果X的方差存在，也可计算如下：

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2$$

切比雪夫不等式

- 定理：令X是均值为μ，方差为σ^2的随机变量，则对任意$t>0$， $P(|X-\mu|\geq t)\leq \sigma^2/t^2$.
- 设定$t=k\sigma$，不等式变为

$$P(|X-\mu|\geq k\sigma)\leq 1/k^2$$

- 推论：如果$\mathrm{Var}(X)=0$, 则$P(X=\mu)=1$
- 应用：在所有数据中
  - 至少有3/4（或75%）的数据位于期望值2个标准差范围内
  - 至少有8/9（或88.9%）的数据位于期望值3个标准差范围内
  - 至少有24/25（或96%)的数据位于期望值5个标准差范围内

案例：投资组合

- 投资者计划分投资本C于两种资产
- 两种资产的期望收益率分别为μ_1=0.1和μ_2=0.03
- 设在第一种资产上的投资比率为π_1，则收益为

$$R = \pi_1 R_1 + (1-\pi_1) R_2$$

- 平均收益为：$E(R) = \pi_1 \mu_1 + (1-\pi_1)\mu_2$
- 设第一种资产为股票（有风险），标准差为σ_1=0.075
- 第二种资产为银行存款(无风险），标准差为σ_2=0
- 则投资收的标准差为：$\sigma_R = \pi_1 \sigma_1$

π_1越大，期望收益越大，但风险也会随之变大。

北京大学

协方差

- 两个随机变量的协方差（covariance)是它们联合变异性的度量，或是它们关联性的度量
- 定义：如果X和Y是分别具有期望μ_X和μ_Y的随机变量，只要下述期望存在，则X和Y的协方差为

$$\mathrm{Cov}(X, Y) = E[(X- \mu_X)(Y- \mu_Y)]$$

- 性质：
  - 两个随机变量的关联是正向的，则协方差为正
  - 两个随机变量的关联是反向的，则协方差为负
  - 两个随机变量是独立的，则协方差为0 (无关的)
  - $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$

协方差的双线性性质

- 定理：设$U=a+\sum_{i=1}^{n} b_i X_i$和$V=c+\sum_{j=1}^{m} d_j Y_j$, 则

$$\mathrm{Cov}(U, V)=\sum_{i=1}^{n}\sum_{j=1}^{m} b_i d_j \mathrm{Cov}(X_i, Y_j)$$

- 推论1：

$$\mathrm{Var}(a+\sum_{i=1}^{n} b_i X_i)=\sum_{i=1}^{n}\sum_{j=1}^{m} b_i d_j \mathrm{Cov}(X_i, X_j)$$

- 推论2：

如果X_i独立，则$\mathrm{Var}(\sum_{i=1}^{n} X_i)=\sum_{i=1}^{n}\mathrm{Var}(X_i)$

北京大学

案例：随机游走

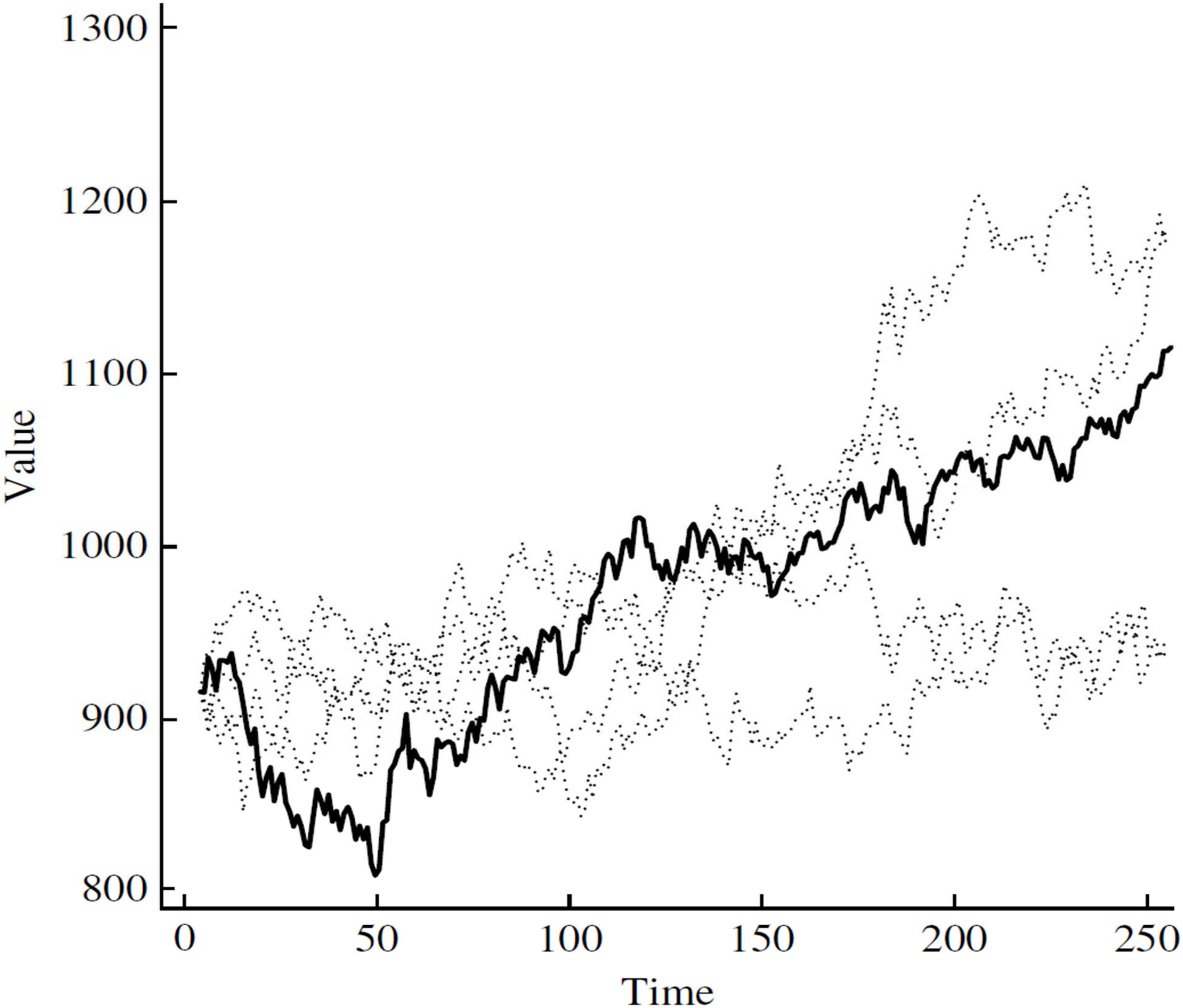
- 一个醉汉自实数线的点x_0开始行走，每一步步长是期望为μ，方差为σ^2的随机变量
- n步后的位置为：$S(n)= x_0 +\sum_{i=1}^{n} X_i$

$$E(S(n))= x_0 +E(\sum_{i=1}^{n} X_i) = x_0 +n\,\mu$$

$$\mathrm{Var}(S(n))= \mathrm{Var}(\sum_{i=1}^{n} X_i) = n\sigma^2$$

- 若μ>0，对于较大的n，他将以较高的概率处于x_0的右侧

北京大学



1300
1200
1100
1000
900
800
Value
0
50
100
150
200
250
Time

相关系数

- 定义：如果X和Y的方差和协方差都存在，且方差非零，则X和Y的相关系数为：

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

- 性质：
  - $-1 \le \rho \le 1$. $\rho = \pm 1$当且仅当$P(Y = a + bX) = 1$,其中a和b为常数
  - 相关系数无量纲
  - 如果X和Y进行线性变换，相关系数保持不变

北京大学

总结

- 期望：
  - 随机变量的平均值，可视为密度或频率函数的中心
  - 也称为随机变量的位置参数
- 标准差：
  - 描述概率分布关于中心的发散程度
  - 随机变量偏离期望的平均幅度
- 协方差：
  - 两个随机变量之间关联度的度量