Ch 3.4 常用的离散型随机变量

回顾前一次课

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

性质：$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}\,(X)$，$E\big(X - E(X)\big)^2 \le E(X - a)^2$

对$X \in [a, b]$有 $\mathrm{Var}(X) \le \big(b - E(X)\big)\big(E(X) - a\big) \le (b - a)^2/4$

- 0/1分布：$\mathrm{X} \sim \mathrm{Ber}(p)$，$E(X) = p$ $\quad\quad \mathrm{Var}(X) = p(1 - p)$
- 二项分布：$X \sim B(n, p)$，$E(X) = np$ $\quad\quad \mathrm{Var}(X) = np(1 - p)$

泊松分布

泊松分布用于描述大量试验中稀有事件出现次数的概率模型.

- 一个月内一个网站的访问量
- 一个小时内公共汽车站来到的乘客数
- 一天中银行办理业务的顾客数
- 一年内中国发生的地震次数

若随机变量X的分布列为

$$P(X=k)=\frac{\lambda^k}{k!}\,e^{-\lambda}\qquad (k\geq 0)$$

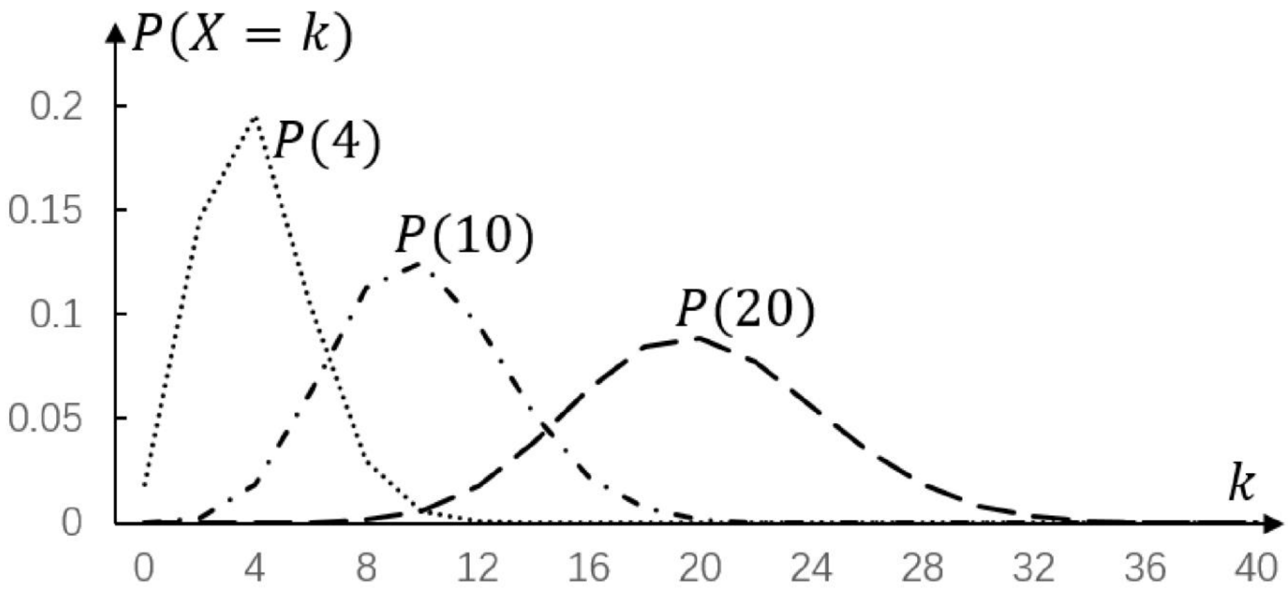
其中$\lambda>0$是一个常数, 称随机变量X服从**参数为 λ 的泊松分布**, 记为 $\boldsymbol{X \sim P(\lambda)}$

泊松分布的数字特征

若随机变量 $X \sim P(\lambda)$, 则有

$$E(X) = \lambda \qquad 和 \qquad \mathrm{Var}\,(X) = \lambda$$

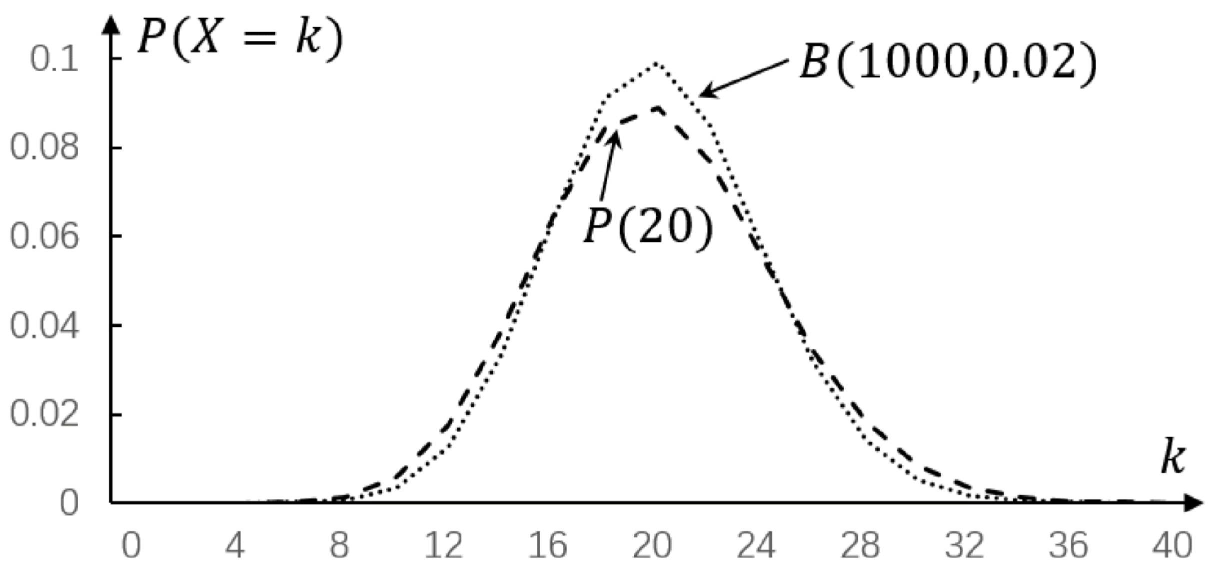
泊松分布可由期望或方差唯一确定



泊松定理

对任意常数$\lambda > 0$, n为任意正整数, 设$np_n = \lambda$, 则对任意给定的非负整数k, 有

$$\lim_{n\to+\infty}\binom{n}{k}p_n^k(1-p_n)^{n-k} = \frac{\lambda^k}{k!}e^{-\lambda}$$



泊松定理的应用

泊松定理的应用: 若随机变量$X \sim B(n,p)$, 当n比较大而p比较小时, 令$\lambda = np$, 有

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$$

即利用泊松分布近似计算二项分布

针对彩票中奖、火山爆发、洪水泛滥、意外事故等小概率事件, 当试验的次数较多时, 可将n重伯努利试验中小概率事件发生的次数近似服从泊松分布

例

有80台同类型设备独立工作，发生故障的概率是0.01，一台设备发生故障时只能由一人处理, 考虑方案

I)　由四人维护, 每人单独负责20台

II)　由三人共同维护80台

方案I)或方案II)哪种更可取?

例

一个公共汽车站有很多路公交车，若一个时间段内到站乘客数$X \sim P(\lambda)$ $(\lambda > 0)$，所有到站的乘客是相互独立的、且选择D1路公交车的概率为p $(p > 0)$，求乘坐D1路公交车的乘客数Y的分布

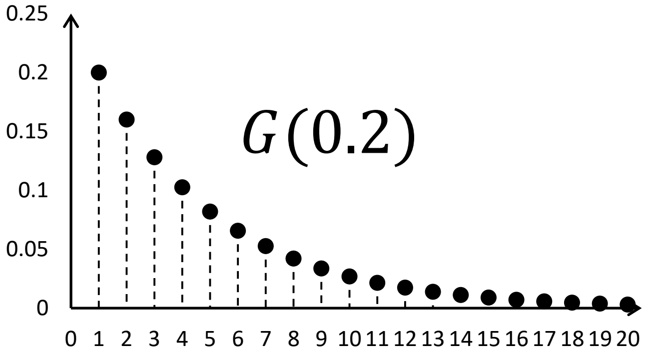
几何分布

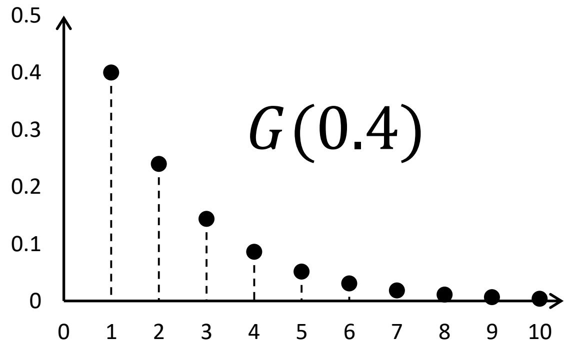
在多重Bernoulli试验, 设事件A发生的概率为p.

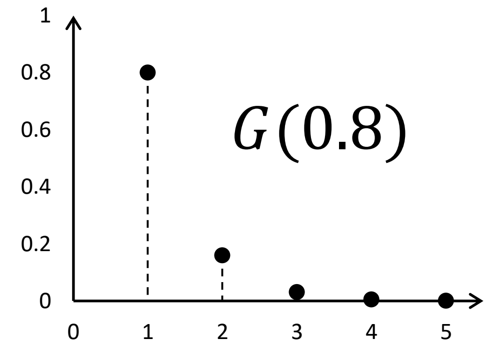
用随机变量X表示事件A首次发生时的试验次数，则X的取值为$1,2,\cdots$, 其分布列为

$$P(X=k)=(1-p)^{k-1}p \qquad (k\geq 1)$$

称**X服从参数为p的几何分布**, 记为 $\boldsymbol{X\sim G(p)}$







几何分布：无记忆性 (memoryless property)

设随机变量$X \sim G(p)$, 对任意正整数m, n, 有

$$P(X > m + n \mid X > m) = P(X > n)$$

直观解释: 现已经历m次失败, 从当前起到成功的次数与m无关

例：一赌徒在赌博时前面总是输, 总觉得下一次应该赢了

几何分布的无记忆性：下一次是否赢与前面输了多少次无关

几何分布的期望与方差

若随机变量$X \sim G(p)$, 则有

$$\mathrm{E}(X) = \frac{1}{p} \qquad 和 \qquad \mathrm{Var}(X) = \frac{1-p}{p^2}$$

例题

古人非常重视生男孩且资源有限，规定每个家庭可生一个男孩，如果没男孩则可以继续生育直至有一个男孩；若已有一个男孩，则不再生育. 不妨假设每个家庭生男孩的概率为$p = 1/2$, 问:

1) 一个家庭恰好有n个小孩的概率;
2) 一个家庭至少有n个小孩的概率;
3) 男女比例是否会失衡?

负二项分布 (Pascal分布)

在多重Bernoulli试验中, 随机事件A发生的概率为p.

用X表示事件A第r次成功时发生的试验次数，则X取值$r, r+1, r+2, \cdots$, 其分布列为

$$P(X=n)=\binom{n-1}{r-1}p^{r-1}(1-p)^{n-r}\cdot p \quad (n\geq r)$$

称***X*服从参数为*r*和*p*的负二项分布，又称 Pascal分布**

案例分析：德国坦克问题

在二战期间，同盟国一直在努力确定德国坦克的生产数量，有助于对德国战力的评估

问题：德国生产了n辆坦克，编号分别为$1,2,\cdots,n$. 盟军在战斗中任意击毁了k辆坦克，被击毁的坦克编号为$x_1, x_2, \dots, x_k$，能否通过被击毁的坦克编号来估计n的大小, 估计德国生产了多少辆坦克

观察被击毁坦克编号分别为17, 68, 94, 127, 135, 212, 估计n

对比情况

时间	统计估计	英国情报估计	德国实际产量
1940-06	169	1000	122
1941-06	244	1550	271
1942-08	327	1550	342

案例分析：集卡活动

小朋友喜欢参加各种集卡活动, 如奥特曼卡和叶罗丽卡等. 事实上很多成年人也对集卡游戏并不陌生，例如80年代的葫芦娃洋画、或90年代的小虎队旋风卡等

问题：市场上有n种不同类型的卡片，假设一个小朋友每次都能以等可能概率、独立地收集一张卡片，问一个小朋友在平均情况下至少要收集多少张卡才能收集齐n种不同类型的卡片

引理: 对任意的随机变量$X_1, X_2, \cdots, X_n$有

$$\boldsymbol{E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)}$$