第四周　　常见随机变量

4.2. 泊松（Poisson）分布

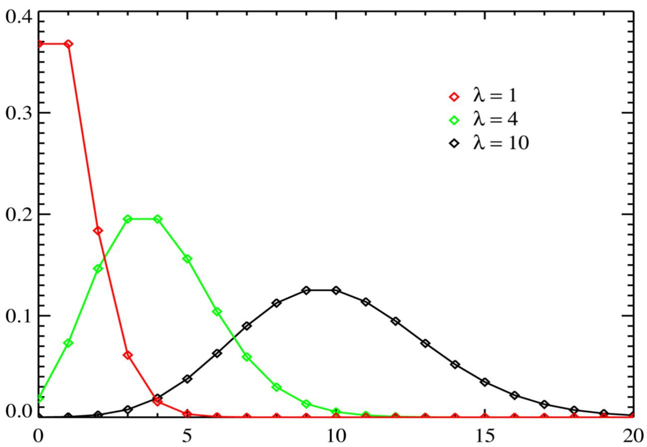
泊松分布（刻画稀有事件的发生规律）

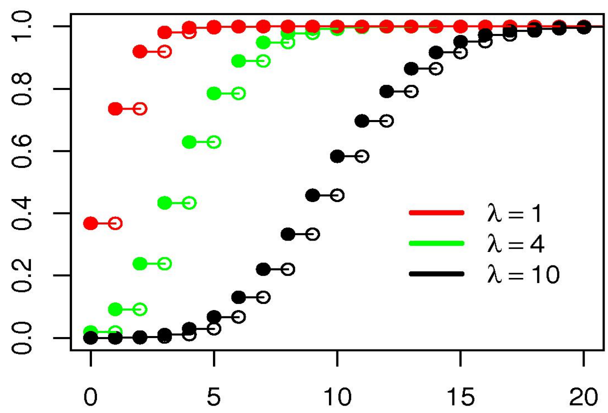
对给定的常数$\lambda>0$，如果随机变量X的分布律为

$$\begin{pmatrix} 0 & 1 & 2 & \cdots \\ p_0 & p_1 & p_2 & \cdots \end{pmatrix},\quad 其中\quad p_k=P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!},\quad k=0,1,2,\cdots$$

则称随机变量X服从参数为λ的泊松分布，记为$X\sim P(\lambda)$

验证：$\forall k\in Z^+, p_k>0$，$\quad \sum_{k=0}^{\infty}p_k=\sum_{k=0}^{\infty}e^{-\lambda}\frac{\lambda^k}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=e^{-\lambda}e^{\lambda}=1$





****

泊松分布与二项分布的关系

考虑二项分布$B(n,p)$，当p很小n很大时，$B(n,p)$与$P(np)$非常接近，可相互近似

若$X\sim B(n,p)$，$Y\sim P(np)$，　则　$P(X=k)\approx P(Y=k)$

令$\lambda=np$，则$\dfrac{B_k(n,p)}{B_{k-1}(n,p)}=\dfrac{\lambda-(k-1)p}{k(1-p)}\approx\dfrac{\lambda}{k}$，

同时 $P(X=0)=B_0(n,p)=(1-p)^n=\left(1-\frac{\lambda}{n}\right)^n=\left(1+\left(-\frac{\lambda}{n}\right)\right)^{\left(-\frac{n}{\lambda}\right)\cdot(-\lambda)}\approx e^{-\lambda}$，所以有

$B_1(n,p)\approx\lambda B_0(n,p)\approx\lambda e^{-\lambda}$，$B_2(n,p)\approx\frac{\lambda}{2}B_1(n,p)\approx\frac{\lambda^2}{2}e^{-\lambda},\cdots$，$B_k(n,p)\approx\frac{\lambda^k}{k!}e^{-\lambda},\cdots$

一个业余选手射箭射中靶心的概率是 0.0001，重复 30000 次，射中靶心 3 次的概率大约为 $B_3(30000,0.0001)\approx P_3(3)=e^{-3}\frac{3^3}{3!}=0.224$。

泊松分布刻画了小概率事件，或是说稀有事件，多次重复时，其发生概率的规律。

****

泊松分布与二项分布严格意义下的相互关系。

泊松定理 设 $X_n\sim B(n,p_n)$，其中 p_n 与 n 有关，且满足极限关系 $\lim\limits_{n\to+\infty}np_n=\lambda$，$\lambda$ 是一个与 n 无关的常数。则对任意固定的非负整数 k，$\lim\limits_{n\to+\infty}P(X_n=k)=\frac{\lambda^k}{k!}e^{-\lambda}$。

证明： $P(X_n=k)=C_n^k p_n{}^k(1-p_n)^{n-k}=\frac{n(n-1)\cdots(n-k+1)}{k!}p_n{}^k(1-p_n)^{n-k}$

$$=\frac{(np_n)^k}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)(1-p_n)^{n-k}$$

令 $np_n=\lambda_n$，则 $\lim\limits_{n\to+\infty}(1-p_n)^{n-k}=\lim\limits_{n\to+\infty}\left(1-\frac{\lambda_n}{n}\right)^{n-k}=\lim\limits_{n\to+\infty}\left(1-\frac{\lambda_n}{n}\right)^n=e^{-\lambda}$，

又 $\lim\limits_{n\to+\infty}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)=1$，所以 $\lim\limits_{n\to+\infty}C_n^k p_n{}^k(1-p_n)^{n-k}=\frac{\lambda^k}{k!}e^{-\lambda}$。

****

例 4.2.1 卢瑟福和盖革在 1910 年观察了放射性物质放出 α 粒子的个数的情况，共观察 $N=2608$ 次，每次观察间隔 7.5 秒，记录到达指定区域的 α 粒子数，共记录下 10094 个粒子，N_k 表示恰好记录到 k 个 α 粒子的观察次数。

k	0	1	2	3	4	5	6	7	8	9	≥ 10
N_k	57	203	383	525	532	408	273	139	45	27	16
$N \cdot p_k(3.87)$	54	211	407	525	508	394	254	140	68	29	17

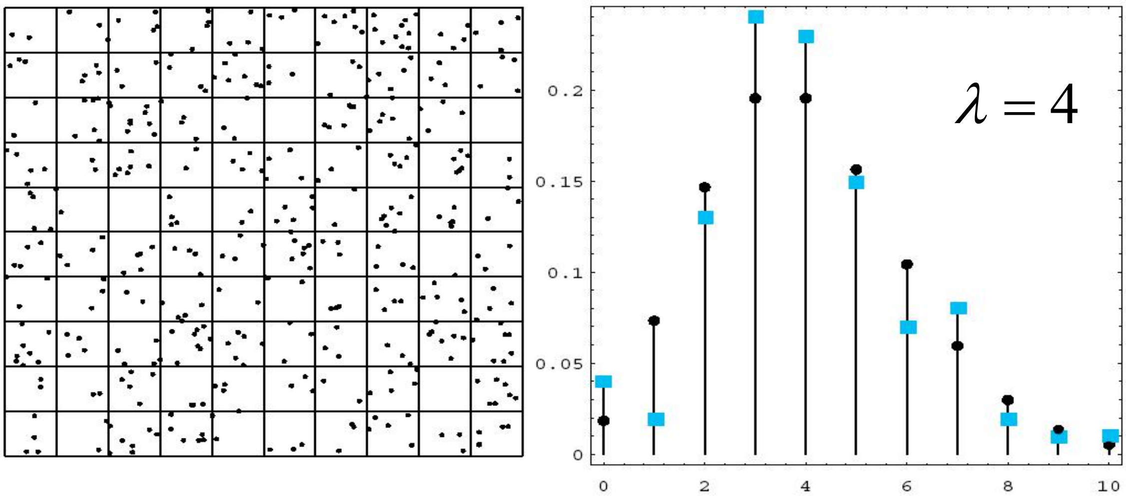
设 X 为 1 次观察中到达的粒子数，则 $X \sim B\left(10094, \frac{1}{2608}\right)$， $10094 \times \frac{1}{2608} \approx 3.87$

X 近似服从 $P(3.87)$。

例 4.2.2 伦敦飞弹。二战时伦敦遭到很多次炸弹袭击，将整个面积分为 $N = 567$ 小块，其中发现 k 枚炸弹的小块数为 N_k，总共发现炸弹 537 枚。

k	0	1	2	3	4	≥ 5
N_k	229	211	93	35	7	1
$N \cdot p(k, 0.9323)$	226.7	211.4	98.6	31.6	7.1	1.6

设 X 为 1 个小块区域的炸弹个数，则 $X \sim B\left(537, \frac{1}{567}\right) \dot{\sim} P\left(\frac{537}{567}\right) = P(0.9323)$



图中给出我们做的一项模拟试验，将 400 个点随机地投入到一个 10 乘 10 等分的包含 100 个小正方形的正方形区域中，右侧图中横坐标表示点数，彩色点表示落入相应点数的小正方形的比例，黑色点则表示参数为 4 的泊松分布的相应概率值。

****

例 4.2.3 上海市在 1875—1955 年中间有 63 年的夏季（5 月—9 月）暴雨记录，共计 180 次。每年夏季共有 $n = 31 + 30 + 31 + 31 + 30 = 153$ 天，每次暴雨如果以一天计算，则每

天发生暴雨的概率为 $p=\frac{180}{63\times 153}=0.0187$，其值很小，同时 $n=153$ 较大。如果暴雨可看成服从二项分布 $X\sim B(153,0.0187)$， $\lambda=np=\frac{180}{63}\approx 2.9$， $X\,\tilde{\sim}\,P(2.9)$

暴雨次数	0	1	2	3	4	5	6	7	≥ 8
实际年份数	4	8	14	19	10	4	2	1	1
理论年数 $\left(63\times p_k(2.9)\right)$	3.5	10.1	14.6	14.1	10.2	6.0	2.9	1.2	0.6

1898 年一位英国学者给出了一个应用泊松分布的经典例子，今天说起来多少有些荒诞。他列出一组数据表明一年中被马踢死的骑兵人数近似服从泊松分布。这也是一个典型的小概率事件被大量重复的例子。一个骑兵在一年中或者被马踢死或者不被马踢死，假定这一小概率事件发生的机会对所有骑兵来说都是一样的，并且士兵们被踢死的机会是独立的。这一假定如果合理的，则一年中被踢死的骑兵数是一个二项分布随机变量。但是被踢死的概率 p 很小，而试验次数，即骑兵人数很大，因此泊松分布对这些数据给出一个很好的描述。泊松分布刻画了稀有事件多次重复时的发生规律。书籍的印刷错误，在给定时间内某耐用设备大批量使用时的故障次数等等都近似服从泊松分布。

****