泊松分布（Poisson distribution）

附近的用户
要鼓起勇气啊，喜欢就去表白呀～

ฅ •ﻌ• ♡ 安安静静的记录心理学，专栏只发表自己深刻理解后的知识，可添加关注私聊

会认真做好一件事，再次感谢大家的戳入

以下为正文：

❀ 泊松分布（Poisson distribution）

可参考关联选项：*统计与概率学*、*离散概率分布*、*正态分布*、*心理学*

可解释的现象：记录观察

泊松分布（法语：loi de Poisson，英语：Poisson distribution）又称Poisson分布、帕松分布、布瓦松分布、布阿松分布、普阿松分布、波以松分布、卜氏分布、帕松小数法则（Poisson law of small numbers），是一种统计与概率学里常见到的离散概率分布，由法国数学家西莫恩・德尼・泊松在1838年时发表。

泊松分布适合于描述单位时间内随机事件发生的次数的概率分布。如某一服务设施在一定时间内受到的服务请求的次数，电话交换机接到呼叫的次数、汽车站台的候客人数、机器出现的故障数、自然灾害发生的次数、DNA序列的变异数、放射性原子核的衰变数、激光的光子数分布等等。

泊松分布的概率质量函数为：

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

泊松分布的参数λ是单位时间（或单位面积）内随机事件的平均发生率。

记号

若X服从参数为λ的泊松分布，记为$X\sim\pi(\lambda)$，或记为$X\sim Pois(\lambda)$.

性质

1、服从泊松分布的随机变量，其数学期望与方差相等，同为参数$\lambda: E(X)=V(X)=\lambda$

2、两个独立且服从泊松分布的随机变量，其和仍然服从泊松分布。更精确地说，若 $X\sim Poisson(\lambda_1)$且$Y\sim Poisson(\lambda_2)$，则$X+Y\sim Poisson(\lambda_1+\lambda_2)$。

3、其矩生成函数为：

$$M_X(t)=E[e^{tX}]=\sum_{x=0}^{\infty}e^{tx}\frac{e^{-\lambda}\lambda^x}{x!}=e^{-\lambda}\sum_{x=0}^{\infty}\frac{(e^t\lambda)^x}{x!}=e^{\lambda(e^t-1)}$$

推导

期望值：(倒数第三至第二是使用泰勒展开式)

$$
\begin{aligned}
\mathrm{E}(X) &= \textstyle\sum_{i=0}^{\infty} iP(X=i) \\
&= \textstyle\sum_{i=1}^{\infty} i\frac{e^{-\lambda}\lambda^i}{i!} \\
&= \lambda e^{-\lambda}\textstyle\sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\
&= \lambda e^{-\lambda}\textstyle\sum_{i=0}^{\infty} \frac{\lambda^{i}}{i!} \\
&= \lambda e^{-\lambda}e^{\lambda} \\
&= \lambda
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}(X^2) &= \textstyle\sum_{i=0}^{\infty} i^2P(X=i) \\
&= \textstyle\sum_{i=1}^{\infty} i^2\frac{e^{-\lambda}\lambda^i}{i!} \\
&= \lambda e^{-\lambda}\textstyle\sum_{i=1}^{\infty} \frac{i\lambda^{i-1}}{(i-1)!} \\
&= \lambda e^{-\lambda}\textstyle\sum_{i=1}^{\infty} \frac{1}{(i-1)!}\frac{d}{d\lambda}(\lambda^i) \\
&= \lambda e^{-\lambda}\frac{d}{d\lambda}[\textstyle\sum_{i=1}^{\infty} \frac{\lambda^{i}}{(i-1)!}] \\
&= \lambda e^{-\lambda}\frac{d}{d\lambda}[\lambda\textstyle\sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!}] \\
&= \lambda e^{-\lambda}\frac{d}{d\lambda}(\lambda e^{\lambda}) = \lambda e^{-\lambda}(e^{\lambda}+\lambda e^{\lambda}) = \lambda+\lambda^2
\end{aligned}
$$

我们可以得到：$Var(X) = (\lambda+\lambda^2)-\lambda^2 = \lambda$

如同性质：$E(X) = Var(X) = \lambda$、$\sigma_X = \sqrt{\lambda}$

泊松分布的来源（泊松小数定律）

在二项分布的伯努利试验中，如果试验次数n很大，二项分布的概率p很小，且乘积λ= np比较适中，则事件出现的次数的概率可以用泊松分布来逼近。事实上，二项分布可以看作泊松分布在离散时间上的对应物。

证明如下。首先，回顾e的定义：

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda},$$

二项分布的定义：

$$P(X=k) = \binom{n}{k}p^k(1-p)^{n-k}。$$

如果令$p=\lambda/n$, n趋于无穷时P的极限：

$$
\begin{aligned}
\lim_{n\to\infty}P(X=k) &= \lim_{n\to\infty}\binom{n}{k}p^k(1-p)^{n-k} \\
&= \lim_{n\to\infty}\frac{n!}{(n-k)!k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \lim_{n\to\infty}\underbrace{\left[\frac{n!}{n^k\,(n-k)!}\right]}_{F}\left(\frac{\lambda^k}{k!}\right)\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\to\exp(-\lambda)}\underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{\to 1} \\
&= \lim_{n\to\infty}\underbrace{\left[\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{k-1}{n}\right)\right]}_{\to 1}\left(\frac{\lambda^k}{k!}\right)\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\to\exp(-\lambda)}\underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{\to 1} \\
&= \left(\frac{\lambda^k}{k!}\right)\exp(-\lambda)
\end{aligned}
$$

最大似然估计（MLE）

给定n个样本值ki，希望得到从中推测出总体的泊松分布参数λ的估计。为计算最大似然估计值，列出对数似然函数：

$$
\begin{aligned}
L(\lambda) &= \ln\prod_{i=1}^{n}f(k_i\mid\lambda) \\
&= \sum_{i=1}^{n}\ln\left(\frac{e^{-\lambda}\lambda^{k_i}}{k_i!}\right) \\
&= -n\lambda+\left(\sum_{i=1}^{n}k_i\right)\ln(\lambda)-\sum_{i=1}^{n}\ln(k_i!).
\end{aligned}
$$

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}L(\lambda)=0 \iff -n+\left(\sum_{i=1}^{n}k_i\right)\frac{1}{\lambda}=0.$$

解得λ从而得到一个驻点（stationary point）：

$$\widehat{\lambda}_{\mathrm{MLE}} = \frac{1}{n}\sum_{i=1}^{n}k_i.$$

检查函数L的二阶导数，发现对所有的λ与k_i大于零的情况二阶导数都为负。因此求得的驻点是对数似然函数L的极大值点：

$$\frac{\partial^2 L}{\partial\lambda^2} = \sum_{i=1}^{n}-\lambda^{-2}k_i$$

生成泊松分布的随机变量

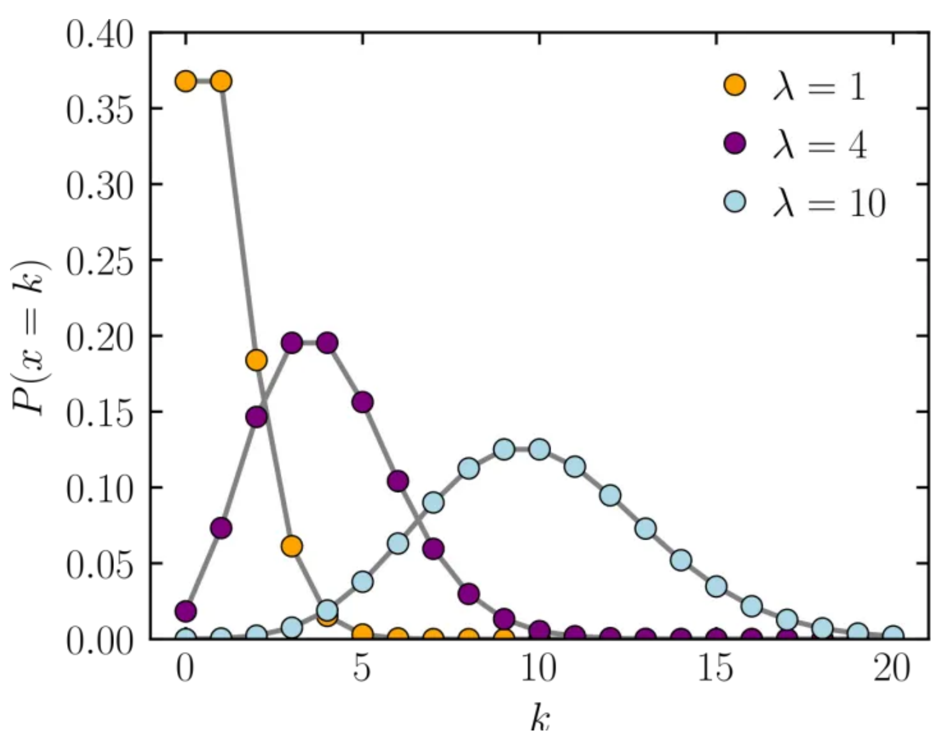
一个用来生成随机泊松分布的数字（伪随机数抽样）的简单算法，已经由高德纳给出（见下文参考）：

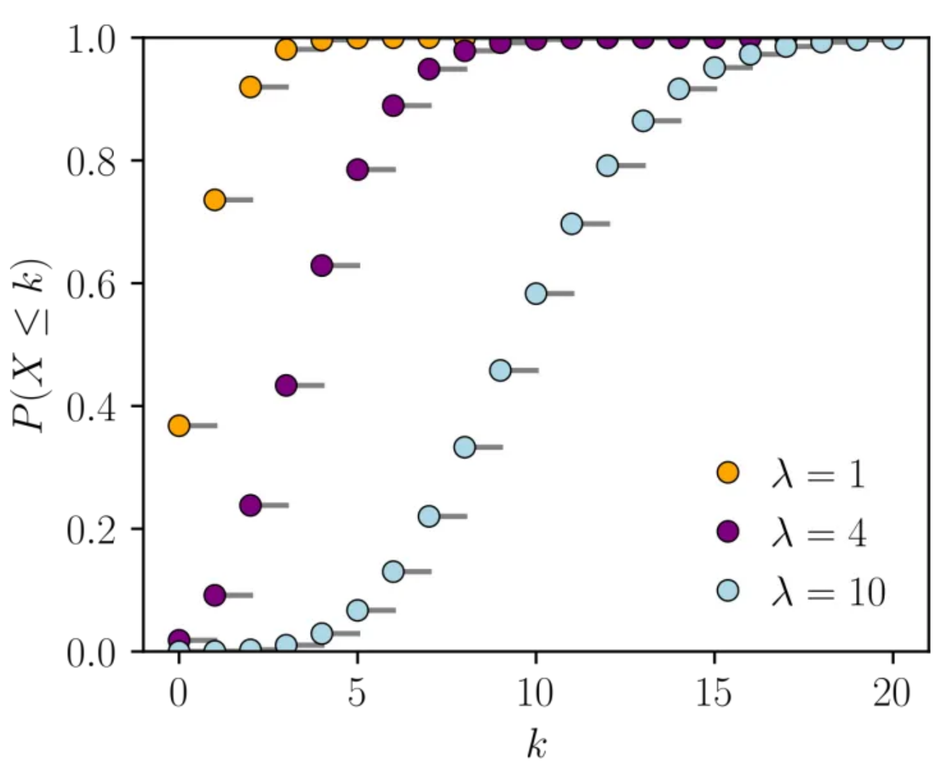
```
algorithm poisson random number (Knuth):
    init:
         Let L ← e^{−λ}, k ← 0 and p ← 1.
    do:
         k ← k + 1.
         Generate uniform random number u in [0,1] and let p ← p×u.
    while p > L.
    return k − 1.
```

尽管简单，但复杂度是线性的，在返回的值k，平均是λ。还有许多其他算法来克服这一点。有些人由Ahrens和Dieter给出，请参阅下面的参考资料。同样，对于较大的λ值，e-λ可能导致数值稳定性问题。对于较大λ值的一种解决方案是拒绝采样，另一种是采用泊松分布的高斯近似。

对于很小的λ值，逆变换取样简单而且高效，每个样本只需要一个均匀随机数u。直到有超过u的样本，才需要检查累积概率。

```
algorithm Poisson generator based upon the inversion by sequential search:
    init:
         Let x ← 0, p ← e^{−λ}, s ← p.
         Generate uniform random number u in [0,1].
    do:
         x ← x + 1.
         p ← p * λ / x.
         s ← s + p.
    while u > s.
    return x.
```





写下你的评论...

5 条评论　默认　最新

凌德
如果p*n=1，当n足够大时，应该服从λ=1的泊松分布，但由中心极限定理，它应该近似服从正态分布，这两者的差距显然是较大的，这是为什么啊
2021-12-31　2

Casanova
我的理解是,你说的这个p*n等于1那么这个分布的期望也是1，方差也是1对应的。中心极限定理的正态分布为N(1,1)，和这个泊松分布在正半轴是差不多的
07-17　赞

cfmt-bia
如果概率p较大，而实验次数N较少，那么就不适合用泊松分布来求解
是这样理解吗?
2021-06-21　赞

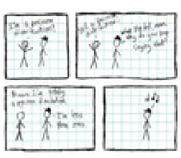
诛心之剑
如果p大n小，你还用得着泊松分布吗?
2021-11-12　2

Master
心理学还有这个🤔
06-07　赞

文章被以下专栏收录

心理学Psychology
人类关怀组织，收集黄昏者

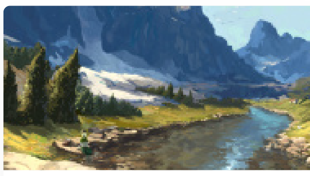
推荐阅读



泊松分布 (Poisson Distributions) 的推导
李家伟

关于泊松分布的一些应用--(i)
关键词: 泊松定理 泊松分布 二项分布的近似 特征函数引言: 概率论是一门处理随机现象统计规律性的基础课程，也是一门实用性很强的课程。我们除了熟练掌握课堂及书本上的知识外，还应该学会&…
Gyp007sinh

泊松分布
一个故事：你已经做了10年的自由职业者了。到目前为止，你的平均年收入约为8万美元。今年，你觉得自己陷入了困境，决定要达到6位数。要做到这一点，你需要先计算这一令人兴奋的成就发生的概…
deeph...　发表于deeph...



泊松分布
逆流君