2.2.1 多递归生成器 2020年5月22日10点41分

阶多递归生成器(MRG)由一系列维状态向量决定,其分量满足线性递归

对于某些模数m,乘数,以及给定的种子.该生成器的最大周期长度为.为了产生快速算法,除少数几个乘数外,其他所有乘数都应为0.当为大整数时，可通过获得随机数的输出流.

通过组合几个较小周期的MRG,可以有效地实现周期很大的MRG,从而产生组合的多递归生成器.

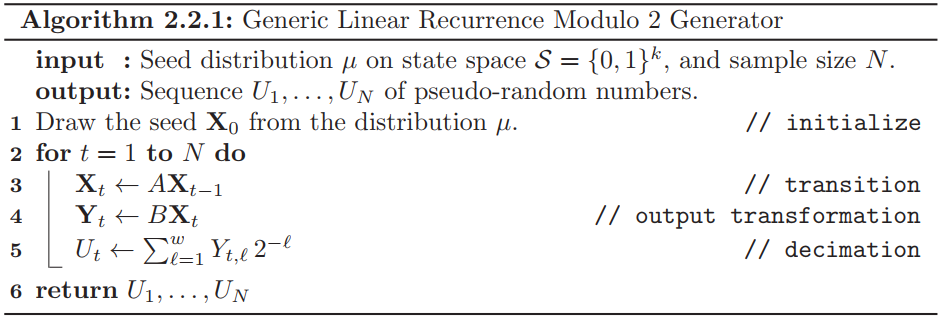
最成功的合并MRG之一是L’Ecuyer [9]的MRG32k3a,它使用了两个3阶MRG,并且重复发生

且输出为

周期长度约为.生成器MRG32k3a通过了当今最全面的测试套件TestU01[11]中的所有统计测试，并已在许多软件包中实现，包括Matlab，Mathematica，Intel的MKL库，SAS，VSL，Arena， 和Automod。 它也是L'Ecuyer的SSJ模拟程序包中的核心生成器，并且可以轻松扩展以生成多个随机流.

2.2.2 模2线性生成器 2020年5月22日11点23分

好的随机生成器必须具有非常大的状态空间.对于线性同余生成器,这意味着模数m必须为大整数.但是,对于多个递归生成器,由于周期长度可以高达mk-1,因此不必采用较大的模数.因为二进制操作通常比浮点操作快(而浮点操作又比整数操作快).考虑基于线性递归模2的MRG和其他随机数生成器是有意义的.[10]中给出了此类随机数生成器的通用框架,其中状态为位向量通过线性变换映射到位输出向量,从而随机数按位抽取如下获得:



在此,A和B分别是和二进制矩阵,并且所有运算都以2为模.特别是,加法对应于按位XOR运算(特别是1+1 = 0).整可以看作是计算机的字长(即w = 32或64).通常情况下(但有例外,请参见[10]),k比w大得多.

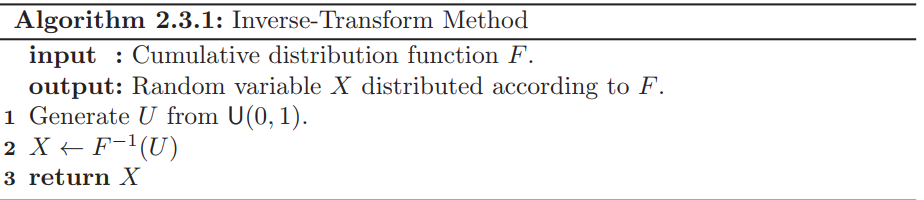
例2.3 马特赛特旋转演算法 过程比较复杂,只给出的算法过程,但没有给出原理.

2.3 随机变量生成

在本节中,我们讨论从规定的分布生成一维随机变量的各种通用方法.我们考虑逆变换方法,混叠方法,合成方法和接受拒绝方法.

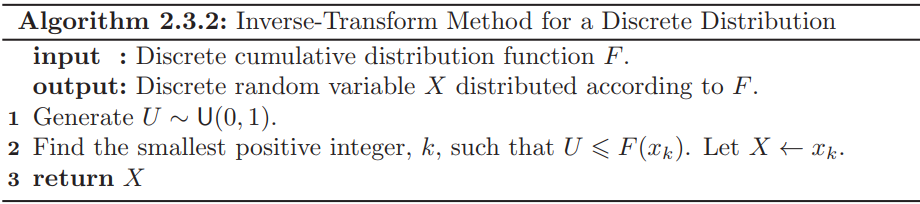
2.3.1 逆变换法

本小节内容给出了你变换法的原理,将均匀随机变量传递给CDF逆函数即可得到的相应随机变量X.



例2.5 有序统计量

例2.6 从离散分布中选取随机变量

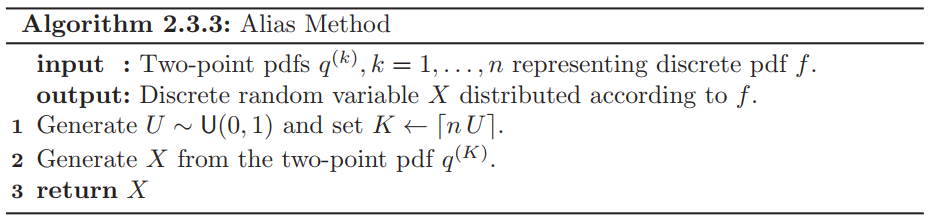


2.3.2 2020年5月29日15点11分

生成离散随机变量的逆变换方法的一种替代方法,即所谓的别名方法[19]，该方法不需要像算法2.3.2的步骤2那样费时的搜索技术.它基于以下事实:任意离散的n点pdf f:

可以表示为n个pdf的均等加权混合,,每个元素最多具有两个非零分量.也就是说，任何n点pdf f都可以表示为

对于两点pdf 的适当定义;参见[19].

别名方法相当通用且有效，但需要初始设置和n份pdf q（k）的额外存储。 在[2]中可以找到计算这两点pdf的过程。 一旦建立了表示（2.8），从f的生成就很简单，可以这样写：

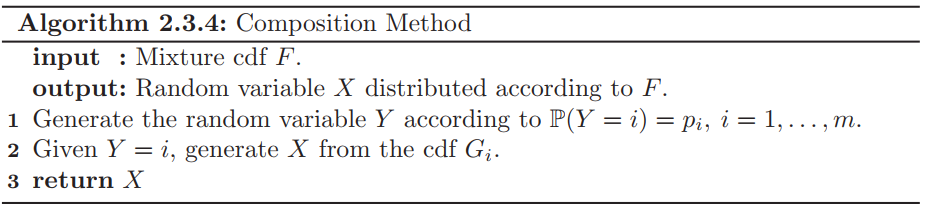
2.3.3 合成方法

此方法假定cdf F可以表示为多个cdf {Gi}的混合,也就是说,

其中

设,是离散随机变量,其且与独立,.则具有cdf F的随机变量X可表示为

因此,为了从F生成X，我们必须首先生成离散随机变量Y，然后在给定Y = i的情况下从Gi生成Xi。 因此，我们有以下方法：



2.3.4 接受拒绝方法

就它们直接处理要生成的随机变量的cdf而言，逆变换和合成方法是直接方法。 由于Stan Ulam和John von Neumann，接受/拒绝方法是一种间接方法。 当上述直接方法失败或证明计算效率低下时，可以应用该方法。

为了介绍这个想法，假设目标pdf f（我们要从中采样的pdf）以某个有限区间[a，b]为边界，并且在该间隔之外为零（见图2.3）。 设

在这种情况下,生成随机变量很简单,可以使用以下接受/拒绝步骤来完成:

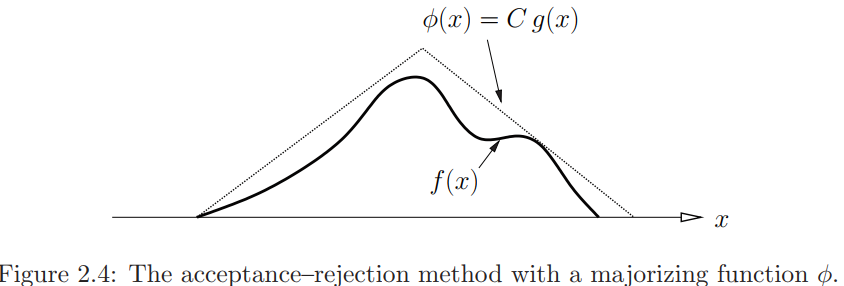
产生X〜U（a，b）。

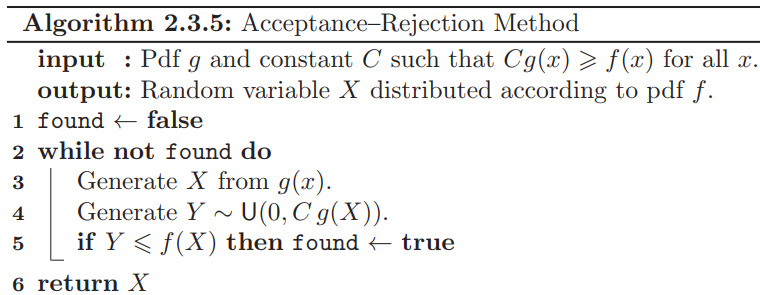
独立于X生成Y〜U（0，c）。

如果Y f（X），则返回Z =X。否则，返回到步骤1。

重要的是要注意,每个生成的向量（X，Y）在矩形[a，b]×[0，c]上均匀分布.因此,接受对（X，Y）在图f下均匀分布。 这意味着X的可接受值的分布具有所需的pdf f.

我们可以将其概括如下:令g为任意密度函数,使得φ（x）= C g（x）使f（x）集中于某个常数C（图2.4）； 即，对于所有x，φ（x）f（x）。 请注意，必然性C1。我们将提案pdf称为g（x），并假设很容易从中生成随机变量。

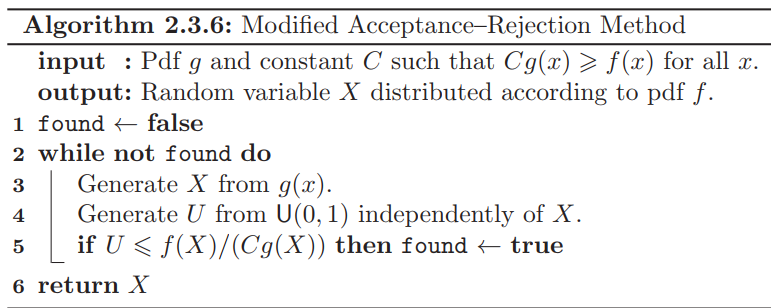




定理2.3.1 根据算法2.3.5生成的随机变量具有所需的pdf f（x）(**证明过程需看懂**)

算法2.3.5的效率定义为

通常,使用算法2.3.5的略微修改版本.这是因为第4行中的与相同,其中,然后我们可以写在第5行中以表示.换句话说,从生成并以概率接受它;否则,拒绝X,然后重试.因此,算法2.3.5的修改版可以重写如下:



2.4 从常用的分布产生随机变量 2020年6月19日13点18分

2.4.1 生成连续随机变量

**2.4.1.1 指数分布**

如果是指数分布,即它的PDF为

其CDF为

因此可以使用逆变换法得到随机变量X,

2.4.1.2 正态分布

如果,其pdf为

由于正态分布不存在解析CDF,逆变换方法不适用于生成正态随机变量,因此必须设计其他过程.我们只考虑从(标准正态分布)生成数据,因为任何随机都可以表示为,其中来自。Box和Muller开发了从生成变量的最早方法之一,如下所示:

令和为两个独立的标准正态随机变量;因此是平面中的随机点.令为对应的极坐标.因此和也可以看作是随机变量,由于

以及和的联合分布为

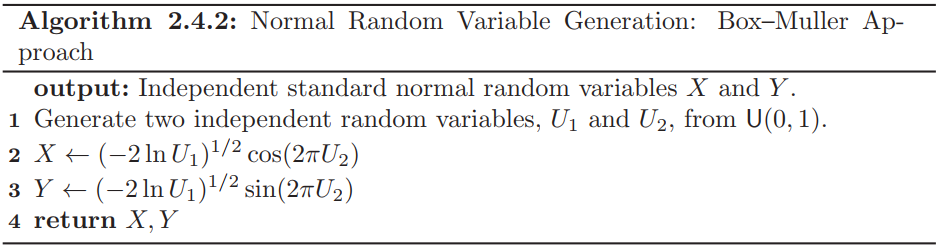
因此可以推导出和的联合分布

即

其中是的取值空间.只依赖值,根据独立随机变量的定义,和是独立随机变量.取常数项作为其pdf,取项作为其pdf,可以看出它们的pdf分别在区间和上各自成立.是在区间上的均匀分布.下面计算的CDF,

因此可以使用逆变换法得到随机变量,

整个生成标准正态分布的流程是: (1)首先利用均匀随机变量生成和随机变量;(2)其次根据公式(2.16)生成标准随机变量和;(3)利用正态分布与标准正态分布的关系来得到最后的结果.



的另一种生成方法是基于接受-拒绝方法.首先,请注意,为了从生成随机变量Y,首先可以从以下pdf生成正随机变量X

然后为X随机分配一个符号.该过程的有效性来自标准正态分布关于零的对称性.

为了从生成随机变量X,我们将与绑定,其中是分布的pdf.使得的最小常数C是.

接受条件,可以写成

其等价于

其中来自.由于也来自(**可以手动证明**),所以最后的不等式可以写成

其中,是独立的并且均为分布的随机变量.

2.4.1.3 伽马分布 2020年7月10日17点27分

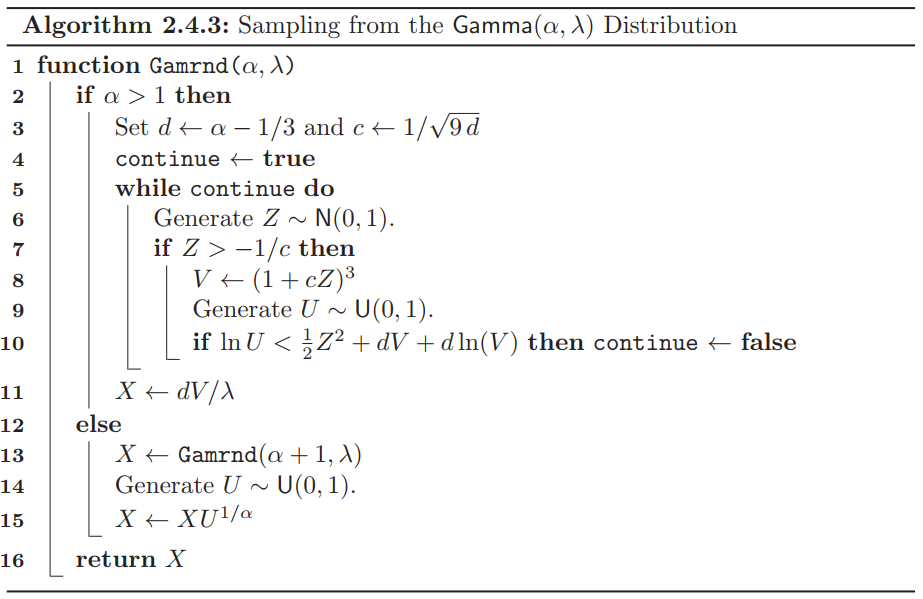
如果,则它的pdf的形式为

参数和分别称为形状和比例参数.由于仅改变比例,因此仅考虑的随机变量生成就足够了.特别是,如果，则(请参阅练习2.16).因为伽马分布的cdf通常不以显式形式存在,所以逆变换方法不能始终应用于从该分布生成随机变量。 因此需要替代方法.我们讨论一种针对情况的方法.令和,其中c和d是常量且为正数.注意是严格增加的函数.令的密度为,其中是归一化常数.那么的密度为.即,通过变换规则(1.16),我们获得

我们通过接受-拒绝方法使用标准正态分布作为建议分布来绘制Y.我们选择c和d,使得,其中接近1，是标准正态分布的pdf.为了找到这样的c和d,我们首先写,其中一些代数将表明

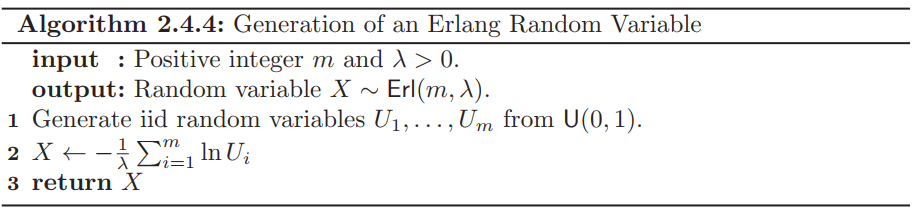
2020年8月7日10点42分

部分推导过程跳过。。。



具有整数形状参数(例如)的伽马分布也称为Erlang分布,表示为.在这种情况下,X可以表示为iid指数随机变量的总和.也就是说,,其中是iid指数变量,每个均值均值为;参见例1.9.使用算法2.4.1,我们可以写出,其中

因此,



2.4.1.4 贝塔分布 如果,则其pdf形式为

假定参数和都大于0.请注意,是(均匀)分布.

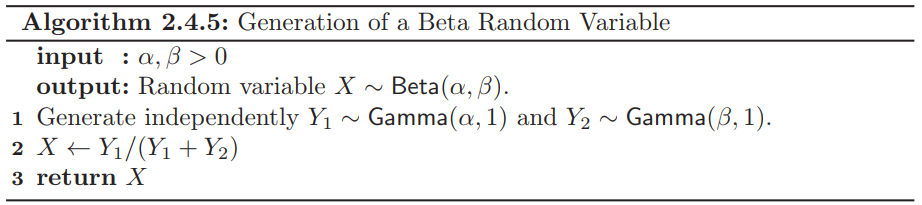
为了从beta分布中采样,让我们首先考虑或等于1的情况.在那种情况下,我们可以简单地使用逆变换方法.例如,对于, pdf为

并且相应的cdf为

因此,通过选择U〜U（0，1）并返回,可以从该分布生成随机变量X.

生成随机变量的一般过程基于以下事实:如果以及和是独立的,则

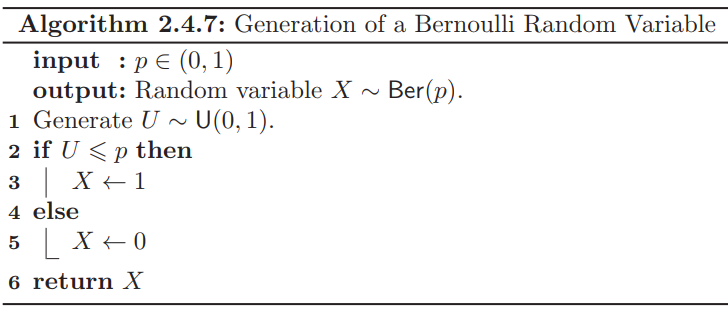
是分布.鼓励读者证明这一命题(见问题2.18).对应的算法如下:



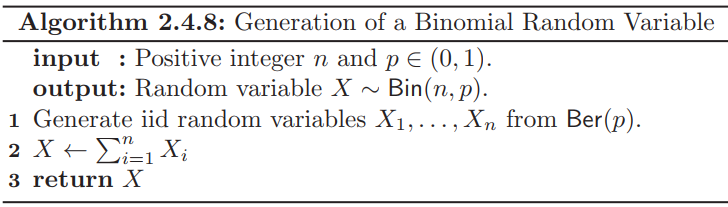
2.4.2 生成离散随机变量

2.4.2.1 伯努利分布 如果,则其pdf形式为

其中是成功概率.应用逆变换方法,我们可以很容易地获得以下生成算法:



**2.4.2.2 二项分布** 如果,则其pdf形式为

回想一下，二项式随机变量可以看作是个独立的伯努利实验的成功总数,每个实验的成功概率为p;参见例1.1.用Xi = 1（成功）或Xi = 0（失败）表示第i次试验的结果,我们可以写成X = X1 +··+ Xn，其中{Xi}是iid Ber（p）随机变量 .因此，最简单的生成算法可以编写如下:

由于算法2.4.8的执行时间与n成正比，因此我们可能会被鼓励使用较大n的替代方法。 例如，我们可以将正态分布视为二项式的近似值.尤其是,根据中心极限定理,随着n的增加,X的分布接近;见（1.26）.实际上,的cdf更好地接近X的cdf.这称为连续性校正.

因此,要获得二项式随机变量,我们可以从生成Y,然后截断为最接近的非负整数.等效地,我们可以生成并设置

作为分布的近似样本.在此,表示的整数部分.

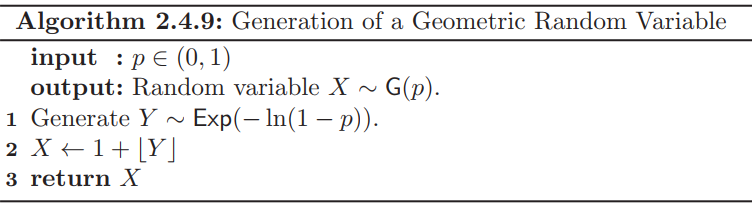
**2.4.2.3 几何分布** 2020年8月14日10点14分

如果,则其pdf形式为

随机变量可以解释为在成功参数为的一系列独立伯努利试验中,直到获得首次成功为止所需的试验次数.注意,.

现在,我们提出一种基于指数分布和几何分布之间关系的算法.设,其中使得.那么具有分布.这是因为

因此,要从生成随机变量,我们首先从具有的指数分布生成随机变量,将获得的值截断为最接近的整数,然后加1.

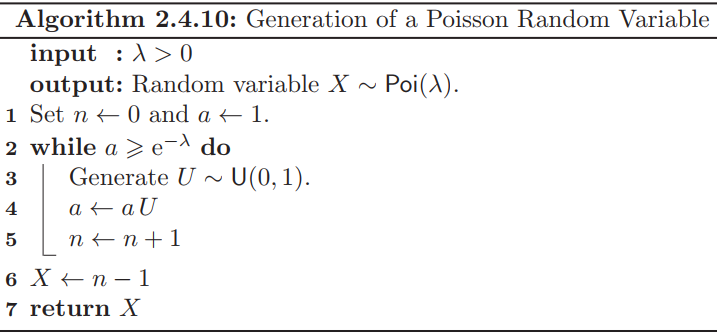


**2.4.2.4 泊松分布** 如果,则其pdf形式为

其中是速率参数.泊松过程和指数随机变量之间存在密切的关系,这一点在泊松过程的性质中得到了明显的体现.请参阅第1.12节.特别是,泊松随机变量可以解释为iid指数变量之和不大于1的最大值（参数为）,即

其中是独立的,并且是分布.由于我们可以将(2.29)重写为

因此,算法为

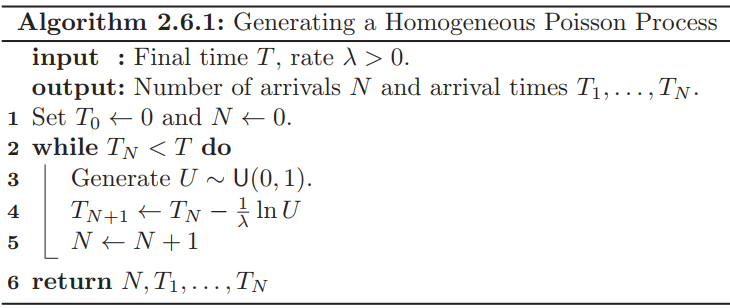


2.5 随机向量生成

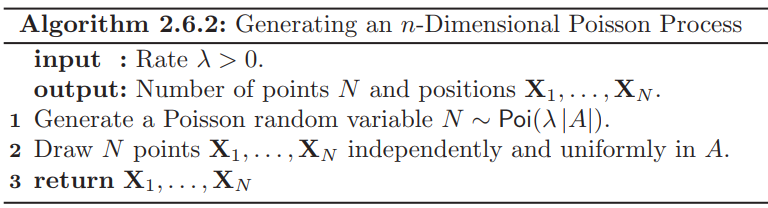
假设我们需要根据pdf f（x）和cdf F（x）从给定的n维分布生成随机向量X =（X1，...，Xn）。 当组件X1，... 。 。 ，Xn是独立的，情况很容易：我们只需对每个组件分别应用逆变换方法或我们选择的另一种生成方法

2.6 生成泊松过程 2020年8月21日11点27分

本节介绍泊松过程的产生.回顾第1.12节,泊松过程有两种不同(但等效)的表征. 在第一个中(请参见定义1.12.1),该过程被解释为一种计数度量,其中对到达的次数进行计数.第二个特征是的到达时间形成一个更新过程,即一个iid随机变量序列.在这种情况下,到达时间具有分布,我们可以写成,其中是iid 分布.使用第二个表征,我们可以生成间隔[0,T]期间的到达时间,如下所示:



泊松过程的第一个特征,即作为随机计数度量,提供了生成此类过程的另一种方法,该方法在多维情况下也有效.特别是(请参阅第1.12节的结尾),以下过程可用于以“体积” | A |的任何集合A生成速率为λ的齐次Poisson过程:



非均匀泊松过程是一个计数过程N = {Nt，t 0}，对于该过程，非重叠间隔中的点数是独立的（类似于普通的泊松过程），但是点到达的速率与时间有关。 如果λ（t）表示时间t的比率，则任何间隔（b，c）中的点数均具有泊松分布，其均值为bcλ（t）dt。 图2.9说明了构建此类过程的方法。 我们首先在带{（t，x），t 0，0 xλ}上以恒定速率λ= maxλ（t）生成二维齐次Poisson过程，然后简单地将所有点投影到λ（ t）到t轴上。

注意，二维泊松过程的点可被视为具有时间和空间维度。 到达历元以速率λ形成一维泊松过程，并且位置在间隔[0，λ]上是均匀的。 这建议了以下用于生成非均质Poisson过程的替代过程：一维均质Poisson过程的每个到达历元都以1 −λ（Tλn）的概率被拒绝（变细），其中Tn是第n个事件的到达时间。 尚存的时期定义了所需的非均匀泊松过程.

