**5.2 伯努利和比努力分布** 2020年3月3日10点47分

**例题5.2.1** 临床试验 伯努利分布说明

**定义5.2.1** 如果一个随机变量只能在0和1取值,并且概率为

则被称为参数为的伯努利分布.

的p.f.可以被写为:

伯努利分布的一些属性

**定义5.2.2 伯努利实验/过程** 如果一组有限或无限的随机变量是i.i.d.,假如每一个随机变量都是参数为的伯努利分布,则称是参数为的伯努利实验.无限伯努利实验也被称为伯努利过程.

**例题5.2.2** 抛硬币 伯努利实验说明

**例题5.2.3** 缺陷零件 伯努利实验说明

**例题5.2.4** 临床试验 伯努利实验说明

**例题5.2.5** 缺陷零件 引出二项分布

**定义5.2.3 二项分布** 如果一个离散随机变量的p.f.如下:

则被称为参数为的二项分布.在该分布中,必须为正整数,且必须满足.

**定理5.2.1** 如果随机变量是个参数为的伯努利试验,并且,则是参数为和的二项分布.二项分布的期望,方差,边际m.f.g.分别为

**定理5.2.2** 如果是独立随机变量,如果是参数为和的二项分布,则是参数为和的二项分布.

如果我们将每个表示为参数的个伯努利试验的总和,则定理5.2.2也很容易理解.如果且所有次试验都是独立的,则总和将简单地是参数为的个伯努利试验的总和.因此,该和必须具有参数为和的二项式分布.

**例题5.2.6** 种族歧视 引出问题

每个仅采用两个值0和1的随机变量必须具有伯努利分布.但是,并非伯努利随机变量之和都具有二项式分布.应用定理5.2.1需要两个条件.伯努利随机变量必须相互独立,并且都必须具有相同的参数,如果这些条件中的任何一个失败,则总和的分布将不是二项式分布.当法院在示例5.2.6中进行二项式计算时,它定义“无歧视”是指独立选择陪审员,陪审员被选为墨西哥裔美国人的可能性为0.791.如果法院以其他方式定义“不歧视”,那么他们将需要进行另一种可能更复杂的概率计算.

**例题5.2.7** 群组测试 **非常复杂的一道题,没看懂**

**5.3 超几何分布** 2020年3月4日09点22分

**例题5.3.1** 不放回采样 引出非独立二项分布的概念

**定理5.3.1 概率函数** 假设一个盒子中包含个红球和个蓝球.同样假设从该盒子中不放回地随机选择个球,并且设为获得的红球数量.显然,.同样,如果,则.当时,如果第次选取的是红球则令,否则.显然.则每一个都是伯努利分布,但不是相互独立的.的分布为

其中

否则.(**证明过程需理解,注意:这里只能用组合而不能用排列**)

**定义5.3.1 超几何分布** 设和是非负整数且.如果一个随机变量是离散的且p.f.如公式(5.3.1)和(5.3.2)所示,则被称作是参数为和的超几何分布.

**例题5.3.2** 临床试验 从超几何分布观察临床实验数据,**结合例题5.3.3更能理解这种观点的可行性**.

**定理5.3.2** 设是超几何分布，则(**证明过程没有完全理解**)

**定理5.3.3** 设和是实数序列,如果收敛至0,同样收敛至0.则

特别地,如果收敛至,则收敛至.(**该定理需要手动证明**)

**定理5.3.4 二项式和超几何分布的接近程度** 设,设是一个正整数.设是参数为和的二项分布.对于每一个正整数使得.其中和是整数.设是参数为,和的超几何分布.对于每一个定值和每一个.

换句话说,定理5.3.4说,如果样本大小代表总量的可忽略部分,那么具有参数和的超几何分布将与具有参数为和的二项式分布非常接近.(**该定理的证明过不难理解,但是非常繁琐**)

**例题5.3.3** 未知组成的总量 **很有深度的一道例题,阐述二项分布和超几何分布之间差异,以及这些分布在应用时的严格前提条件**.

**5.4 泊松分布** 2020年3月6日10点30分

**例题5.4.1** 客人到店的频率 引出泊松分布 **非常重要,是理解泊松分布的关键**

**定义5.4.1 泊松分布** 设.如果随机变量的p.f.如下所示:

则称为均值为的泊松分布.

**定理5.4.1 均值** 泊松分布的均值为,即公式(5.4.2)的均值为.(如果是有限的,.)

**例题5.4.2** 客人到店的频率 **例题5.4.1**的扩展 **对比泊松分布和二项分布的异同**

**定理5.4.2 方差** 均值为的泊松分布的方差同样也是.(证明过程很有技巧性)

**定理5.4.3 矩量母函数** 均值为的泊松分布的m.g.f为

**定理5.4.4** 如果随机变量是独立的并且是均值为的泊松分布,则是均值为的泊松分布.(**证明过程看懂了,但是通过距量母函数反推分布的依据在哪里？**)

**例题5.4.3** 客人到店的频率 **例题5.4.1**的扩展 泊松分布的应用

**定理5.4.5 二项式和泊松分布的接近程度** 对每一个整数和每一个,设为参数为的二项分布,设为均值为的泊松分布.设是数值为0到1的序列使得.则

**例题5.4.4** 近似概率 用泊松分布计算二项分布的近似值

**定理5.4.6 超几何和泊松分布的接近程度** 设.设是均值为的泊松分布.对每一个正整数使得,,.其中, 和是整数.设是参数为,和的超几何分布.对于每一个定值和每一个.

**该定理需要手动证明,证明过程可参考定理5.3.4和5.4.5**.

**例题5.4.5** 客人到店的频率 **例题5.4.1**的扩展 提出新的问题:**假设店家对其它特定时间段(例如15分钟,4小时等等)的客流量感兴趣,那么一小时观察到的泊松分布是否依然有效**?

**定义5.4.2 泊松过程** 单位时间内频率为的泊松过程满足下列两个属性:

1. 在每个固定时间区间抵达的数量是均值为的泊松分布.
2. 在每个离散时间区间的抵达数量是相互独立的.

上述定义肯定地回答了**例题5.4.5**的**疑问**.

**例题5.4.6** 放射性粒子 泊松过程的应用

**注意:泊松过程的一般性**.尽管我们根据时间间隔内的到达次数介绍了泊松过程,但泊松过程实际上更为通用.例如,泊松过程可用于对空间和时间中的事件建模.泊松过程可用于对到达总机的电话呼叫,放射源发出的原子粒子,森林中患病的树木或制成品表面的缺陷进行建模.泊松过程模型之所以受欢迎,是由两个原因.首先,该模型在计算上很方便.第二,如果对现象如何发生做出三个合理的假设,则该模型具有数学上的依据.在另一个示例之后,我们将详细介绍这三个假设.

**例题5.4.7** 饮用水中的隐孢子虫 **非常精彩的一道综合例题,值得反复细品**

**泊松过程下的假设**

在下文中,我们将参考时间间隔,但是对于二维或三维区域的子区域或线性距离的子长度,这些假设同样适用.确实,泊松过程可用于对可细分为任意小块的任何区域中的事件进行建模.导致泊松过程模型的三个假设.第一个假设是任何不相交的时间间隔集合中的出现次数必须相互独立.例如，即使在特定时间间隔内总机接收到异常多的电话呼叫，在即将到来的时间间隔内将至少接收到一个电话的概率也保持不变。同样，即使在总机上没有收到异常长间隔的呼叫，下一个短间隔,内仍将接收到呼叫的概率仍然保持不变。第二个假设是，每个非常短的时间间隔内发生的概率必须与该时间间隔的长度近似成比例。为了更正式地表达该条件，我们将使用标准数学符号，其中o（t）表示t的任何具有以下性质的函数。根据（5.4.8），o（t）必须是当t→0时接近0的函数，此外，此函数必须以比t本身更快的速率接近0。这种函数的一个例子是o（t）=tα，其中α>1。可以证明该函数满足Eq。 （5.4.8）。现在可以将第二个假设表示为：存在一个常数λ> 0，使得对于每个长度为t的时间间隔，在该间隔内至少出现一次的概率为λt+ o（t）。因此，对于每个非常小的t值，在长度为t的时间间隔内至少发生一次的概率等于λt加一个数量级较小的量。第二个假设的后果之一是，被观察的过程在整个观察期间必须是静止的。也就是说，在整个期间内，发生的概率必须相同。既没有忙碌的时间间隔（在此之前我们无法事先知道发生的可能性会更频繁），也不会既没有安静的时间间隔（在这一间隔时间内我们事先知道的事件可能不会那么频繁）.这种情况反映在以下事实上：相同的常数λ表示整个观察周期内每个间隔中出现的可能性。可以以更复杂的数学为代价来放宽第二个假设，但是我们在这里不这样做。第三个假设是，对于每个非常短的时间间隔，在该时间间隔内将出现两次或多次出现的概率必须比仅发生一次出现的概率具有较小的数量级。在符号中，在长度为t的时间间隔内两次或多次出现的概率必须为o（t）。因此，与在该间隔中发生一次的概率相比，在一个小间隔中发生两次或更多次的概率必须微不足道。当然，从第二假设出发，与没有发生的概率相比，在相同间隔内发生一次的概率本身可以忽略不计。

在前面的三个假设下，可以证明该过程将满足速率为λ的泊松过程的定义。

**上述内容没有彻底理解,以后再做更全面的归纳总结**.

**5.5 负二项分布** 2020年3月7日13点35分

**例题5.5.1** 有缺陷的零件 引出问题

**定理5.5.1 采样直到固定次数成功为止** 假设一个无穷的伯努利实验,其成功率为.在第次成功前出现的失败次数的p.f.为:

**定义5.5.1 负二项分布** 如果离散随机变量的p.f.如公式(5.5.1)所示,则被称为参数为和的负二项分布.

**例题5.5.2** 有缺陷的零件 **例题5.5.1**的解是:的分布是参数为4和的负二项分布.

**定义5.5.2 几何分布** 如果离散随机变量的分布是如下定义:

则称是具有参数的几何分布.

**例题5.5.3** 三个彩票号码相同 对几何分布的实例说明

**定理5.5.2** 如果是i.i.d.随机变量并且每一个都是参数为的几何分布,则是参数为和的负二项分布.(**证明过程需要反复阅读,以加深理解**)

**定理5.5.3 矩量母函数** 如果被称为参数为和的负二项分布,则m.g.f.为

参数为的几何分布的m.g.f.是公式(5.5.4)时的特殊情况.(**证明过程需要反复阅读,以加深理解**)

**定理5.5.4 均值和方差** 如果具有参数为和的负二项分布,则均值和方差分别为

参数为的几何分布的均值和方差是公式(5.5.7)时的特殊情况.(**证明过程需要反复阅读,以加深理解**)

**例题5.5.4** 三个彩票号码相同 计算几何分布的期望 引出新的问题

**定理5.5.5 几何分布的无记忆属性** 设X具有参数为的几何分布,并且设.则对每一个整数,

定理5.5.5的直觉如下:将视为在伯努利试验序列中首次成功之前的失败次数.令为从第次试验开始直到下一次成功的失败次数.然后,具有与相同的分布，并且独立于前个试验.因此,以前次试验中发生的任何事情为条件(例如尚未成功)不会影响Y的分布,它仍然是相同的几何分布.练习8要求给出形式证明.在练习13中,您可以证明几何分布是唯一具有无记忆特性的离散分布.

**例题5.5.4** 三个彩票号码相同 **例题5.5.4**的分析,需要结合**定理5.5.5**理解,**需要深入思考**.

**5.6 正态分布** 2019年7月25日13点40分——2020年3月9日10点42分

**例题5.6.1** 汽车尾气 引出问题

**正态分布比较重要的三个原因**:

1. 它直接与正态分布的数学属性相关联.将随机变量的分布建模为正态分布会使得计算过程变得简单.
2. 基于物理实验的观测数据表面许多随机变量近似正态分布.例如人的身高和体重分布.
3. 从某些其它分布中取大量样本,观察到的许多重要功能具有近似的正态分布.

**定义5.6.1 定义和p.d.f.** 如果一个连续随机变量的p.d.f.为:

则称为均值,方差的正态分布.

**定理5.6.1** 公式所定义的函数是一个p.d.f..(**证明过程很有灵性**)

**例题5.6.2** 汽车尾气 **例题5.6.1**的扩展

**定理5.6.2 矩量母函数** 正态分布(5.6.1)的m.g.f.为(**证明过程很有灵性**)

**定理5.6.3 均值和方差** 由公式所定义p.d.f.的均值和方差分别为和.(证明过程需要结合定理4.4.2)

**例题5.6.3** 汽股票价格变化 正态分布的应用 **需要注意期望与距量母函数的关系**

**定理5.6.4** 如果是正态分布,均值和方差分别为和,如果,其中和是给定常数且.则是均值为,方差为的正态分布.(**证明过程需要结合的定理5.6.2证明步骤的逆向操作才能重构完整的证明**)

**定义5.6.2 标准正态分布** 均值为0,方差为1的正态分布被称为标准正态分布.标准正态分布的p.d.f.一般用符号表示,c.d.f.用符号表示.因此，

**定理5.6.5 对称结果** 对于所有的和(**证明过程容易使人困惑**)

**定理5.6.6 转换成标准正态分布** 设是正态分布,均值和方差分别为和.设为的c.d.f..则是标准正态分布,对所有的和

**例题5.6.4** 根据标准正态分布计算一般正态分布的概率

**例题5.6.5** 根据标准正态分布计算一般正态分布的分位函数

Comparisons of Normal Distributions,**该小节内容没有理解透彻**

**定理5.6.7** 如果随机变量是独立的并且是均值为,方差为的正态分布,则是均值为,方差为的正态分布.

**推论5.6.1** 如果随机变量是独立的并且是均值为,方差为的正态分布,假设和是常数并且这些值至少有一个不为0,则变量是均值为,方差为的正态分布.

**例题5.6.6** 男人和女人的身高 对**定理5.6.7**或**推论5.6.1**的应用,**值得引起注意**.

**定义5.6.3 采样均值** 设是个随机变量.这个随机变量的平均值,即,被称为采样均值且通常标记为.

**推论5.6.2** 假设随机变量是从均值和方差的正态分布得到的样本,设为它们的样点均值.则是均值为和方差为的正态分布.

**例题5.6.7** 确定样本大小 对**定义5.6.3**和**推论5.6.2**的应用**.例题的解答过程有一些跳步,直接对公式(5.6.11)作代数运算即可得到答案**.

**例题5.6.8** 均值区间 利用**例题5.6.7**的结果反推均值区间所对应的概率,该例题中存在**书写错误**:应该为.

**定义5.6.4 对数正态分布** 如果是均值和方差的正态分布,我们说是均值和方差的对数正态分布.对数正态分布的矩母函数m.g.f.,期望和方差分别为:

2020年3月13日10点25分

**例题5.6.9** 零件寿命 对数正态分布应用的案例 利用c.d.f.求解概率,**这道例题可以死记硬背,但是需要明白其原理**.

**例题5.6.10** 股票和期权价格 非常精彩的一道例题,结合**例题5.6.3**和**例题4.1.14**.特别需要注意**例题4.1.14**的内容,**有时间可以在下面写下完整的解答过程**.

**5.7 伽马分布** 2019年8月13日11点13分——2020年3月13日11点53分

**例题5.7.1** 灯泡寿命的均值和方差 引出问题

**定义5.7.1 伽马函数** 对于每一个正数值, 设定义为下列积分：

函数被称为伽马函数.

**定理5.7.1** 如果则

证明过程中使用的公式**需要证明**.

**定理5.7.2** 对于每一个正整数,

**例题5.7.2** 灯泡寿命的均值和方差 根据伽马函数(**定义5.7.1**和**定理5.7.2**)计算均值和方差

**定理5.7.3** 对于每一个和每一个,

**定理5.7.4** **斯特林(Stirling)公式**

**例题5.7.3** 队列中的服务时间 **定理5.7.3**的应用,与**例题3.7.5**具有相同的结果.

**例题5.7.4** 队列中的服务时间 **例题5.7.3**的扩展,引出伽马分布.

**定义5.7.2 伽马分布** 设和是正数.如果连续随机变量的分布如下:

则被称为参数为和的伽马分布.

**例题5.7.5** 队列中的服务时间 说明**例题5.7.4**分布是伽马分布.

**定理5.7.5 矩** 设是参数为和的伽马分布.对于

特别地,

**例题5.7.6** 队列中的服务时间 计算**例题5.7.4**分布的期望.

**定理5.7.6 距量母函数** 设是参数为和的伽马分布.则的m.g.f.为

**定理5.7.7** 如果随机变量是独立的,并假设是参数为和的伽马分布, 则之和是参数为和的伽马分布.

**定义5.7.3 指数分布** 设.如果随机变量的连续p.d.f.定义如下:

则被称为参数为的指数分布.

**定理5.7.8** 如果,则伽马分布退化为指数分布.如果是为参数为的指数分布,则

**定理5.7.9 指数分布的无记忆性** 设是参数为的指数分布,令.则对于每一个值,

**证明过程值得研究,貌似可以证明几何分布的无记忆性**(**定理5.5.5**).

为了说明无记忆属性,我们将假设表示某个事件发生之前经过的分钟数.根据等式,如果事件在最初的分钟内没有发生,那么事件在接下来的分钟内不会发生的概率就是.这与从时间0开始的分钟间隔内事件不会发生的概率相同.换句话说,在没有事件发生的情况下无论经过了多长时间,事件都会在接下来的分钟内发生的概率始终相同.

不是所有实际问题都严格满足无记忆性.例如,假设是灯泡失效前使用时长.灯泡在未来有望继续燃烧的时间长短取决于过去燃烧灯泡的时间长短.但是,指数分布已有效地用作各种产品寿命等变量的近似分布.

**例题5.7.7** 灯泡寿命 提出问题 给定分布最小值问题

**定理5.7.10** 假设一组随机变量是参数为的指数分布.则的分布是参数为的指数分布.(**证明过程需要理解**)

**定理5.7.11** 假设一组随机变量是参数为的指数分布.设是随机变量从小到大的排列.对于每一个,设,则的分布是参数为的指数分布.(**证明过程有点绕,需要结合无记忆性来理解**)

**例题5.7.8** 放射性粒子 **例题5.4.6**的延续 引出问题

**定理5.7.12 泊松过程中抵达时间** 假设抵达事件符合频率为的泊松过程.设是第次抵达的时间,.令,,.则是i.i.d并且每一个具有参数为的指数分布.(**证明过程很巧妙,但需要透彻理解**)

**推论5.7.1 第次抵达时间** 在定理5.7.12情况下,具有参数为和的伽马分布.

**证明**:由于

将上述个等式相加,得到.根据**定理5.7.7**得到具有参数为和的伽马分布.

**5.8 贝塔分布** 2019年8月19日10点25分——2020年3月16日09点31分

**例题5.8.1** 有缺陷的零件 **例题3.6.9**的扩展,引出新问题

**定义5.8.1 贝塔函数** 对于每一个正数和,定义

函数被称为贝塔函数.

**定理5.8.1** 对于所有的

**例题5.8.2** 有缺陷的零件 **例题5.8.1**所提出公式的进一步化简

**定义5.8.2 贝塔分布** 设,设是一个随机变量且p.d.f.为

则是被称为参数为贝塔分布.当时,该分布简化为区间上的均匀分布.

**例题5.8.3** 种族歧视 **5.2.6**的扩展和深化 **非常值得研究**.

**定理5.8.2** 假设具有参数为的贝塔分布,并且在给定时的条件分布是参数为和的二项分布.则在给定时的条件分布是参数为的贝塔分布.

**定理5.8.3 矩** 假设具有参数的贝塔分布.则对每一个正整数,

特别的，

**例题5.8.3** 种族歧视 **例题5.8.2**的解,**定理5.8.2**的应用 **非常值得研究**.

**例题5.8.4** 临床实验 **例题2.1.4**的扩展,与**例题5.8.3**类似,**定理5.8.2**的应用

**定理5.8.4** 设和是独立随机变量,其中是参数为和的伽马分布,是参数为和的伽马分布.则(**需要理解证明过程,雅可比变换**)

* 和是独立的.
* 是参数为和的贝塔分布,并且
* 是参数为和1的伽马分布.

同时，公式(5.8.2)成立.

**5.9 多项分布** 2019年8月20日10点10分——2020年3月16日10点10分

**例题5.9.1** 血型 引出问题

**定义5.9.1 多项分布** 一个离散随机向量的p.f.如下所示,则称为参数为和的多项分布.

**例题5.9.2** 参加棒球比赛的队员 多项分布的实例说明

**例题5.9.3** 血型 多项分布的实例说明

**定理5.9.1** 假设随机向量是参数为和的多项分布.则是参数为和的二项分布,并且.(**证明过程需要理解**)

**推论5.9.1** 假设随机向量是参数为和的多项分布.则每一个变量是参数为和的二项分布.

**推论5.9.2** 假设随机向量是参数为和的多项分布,且.设,并且假设是集合中不同的元素.则是参数为和的二项分布.

最后一点,伯努利分布与二项式分布之间的关系扩展到多项式分布.参数的伯努利分布与参数1和的二项式分布相同.但是,对于第一个参数的多项式分布,没有单独的名称.具有这种分布的随机向量其中一个分量为1,其余个分量均为0.第个坐标为1的概率为.维向量似乎是表示仅可以采用个不同值的随机对象的一种笨拙的方式.更常见的表示形式是将X作为单离散随机变量以概率在对应的个值之中选取一个.

**定理5.9.2 均值,方差和协方差** 假设随机向量是参数为和的多项分布.则

**注意:负协方差是多项分布的本质** 多项向量不同分量之间的负协方差是自然的,因为向量个分量共享个选择.如果一个分量较大,那么其它分量必须较小,因为分量之和固定为.

**5.10 二元正态分布** 2019年8月20日11点55分——2020年3月16日11点02分

**例题5.9.1** 甲状腺激素 提出问题

**定理5.10.1** 假设和是独立随机变量,并且每一个都是标准正态分布.设和均为常量并且满足.定义两个新的随机变量如下:

则和的联合p.d.f.为(**证明过程需要理解,雅可比变换**)

**定理5.10.2** 假设和的联合p.d.f.如公式(5.10.2)所示.则存在独立标准正态分布随机变量和使得公式(5.10.1)成立.同样,的均值为,方差为.和之间的相关性为.则的边际分布是均值为方差为的正态分布.其中.(**证明过程需要理解**)

**定义5.10.1 二元正态分布** 当两个随机变量和的联合p.d.f.形如公式(5.10.2),则称和是参数为,方差为,相关性为的二元正态分布.

**例题5.9.2** 蚤甲虫的人体测量学 二元正态分布实例,但是这个例子不太好

**定理5.10.3 独立性和相关性** 两个随机变量和是二元正态分布独立的当且仅当它们是非相关的.

**定理5.10.4 条件分布** 假设和具有二元正态分布,其联合p.d.f.如公式(5.10.2)所示.在给定时的条件分布是正态分布,其均值和方差分别为(**证明过程稍微复杂,需要理解**)

**例题5.9.3** 估计人体体重 该例题涉及估计量 二元正态分布比单个正态分布具有更小的方差,所以是更好的估计量.

**例题5.9.4** 估确定边际分布 该例题有点绕,需要注意

**例题5.9.5** 丈夫和妻子的身高 提出问题

**定理5.10.5 二元正态分布的线性组合** 假设和具有二元正态分布,其联合p.d.f.如公式(5.10.2)所示.设,其中是任意常量.则是一个正态分布,其均值为,方差为

**例题5.9.6** 丈夫和妻子的身高 根据**定理5.10.5**解答**例题5.9.5**的问题