**4.1 随机变量的期望** 2020年2月25日09点30分

**例题4.1.1** 股票的公平价格

**例题4.1.2** 股票价格变化 **例题4.1.1**的解

**定义4.1.1 有界离散随机变量的均值** 设是一个有界离散随机变量,其p.f.为.的期望,标记未,定义如下:

的期望也被称为的均值[mean]或的期望值.

**例题4.1.3** 伯努利随机变量

**定义4.1.2 一般随机变量的均值** 设是一个一般离散随机变量,其p.f.为.假设下列表达式至少有一个是有限的:

则的均值或期望存在,定义如下

如果中的两个和式都是无限的,则不存在.

如果中的两个和都是无限大,则期望不存在的原因是,在这种情况下中的和没有得到很好的定义.从微积分可以知道,正负项都加到无穷大的无穷级数之和要么不能收敛,要么以不同顺序重新排列这些项来使其收敛到许多不同的值.我们不希望期望值的含义取决于加号顺序的任意选择.如果在中两个和只有一个是无限的,则期望值也是无限的,其符号与无限的和相同.如果两个总和都是有限的,则只有中的总和收敛,并且不依赖于项相加的顺序.

**例题4.1.4** 随机变量期望不存在的示例

**例题4.1.5** 随机变量期望是无穷的示例

**定义4.1.3 有界连续随机变量均值** 设是一个有界连续随机变量,其p.d.f.为.的期望标记为,定义为:

的期望也被称为的均值或的期望值.

**例题4.1.6** 预期故障时间

**定义4.1.4 一般有界连续随机变量均值** 设是一个一般有界连续随机变量,其p.d.f.为.假设下列积分至少有一个是有限的:

则的均值或期望存在,定义如下

如果中的两个积分都是无限的,则不存在.

**例题4.1.7** 保修期时长

**例题4.1.8** 连续分布均值不存在 柯西分布

**均值与重心的关系**:对随机变量的期望或等效地,对其分布的均值可以视为该分布的重心.

**例题4.1.9** 柯西分布 柯西分布在原点对称,但是期望不存在

**例题4.1.10** 故障率和故障时间

**例题4.1.11** 故障率和故障时间 **例题4.1.10**的解 根据随机变量函数求解

**定理4.1.1 佚名统计学家定理(Law of unconscious statistician)** 设是一个随机变量,设是一个实变量的实函数.如果是连续分布并且均值存在,则

如果是离散分布且均值存在,则

**例题4.1.12** 故障率和故障时间 **例题4.1.10**的解 根据**定理4.1.1**求解

**例题4.1.13** 随机变量函数的期望 **例题4.1.6**的扩展

**例题4.1.14** 期权价格 涉及一些经济学知识

**例题4.1.15** 两个随机变量的期望

**定理4.1.2 佚名统计学家定理(Law of unconscious statistician)** 假设是随机变量,其联合p.d.f.为.设是个实变量的实值函数,假设.则从下列关系中可以直接得出

如果均值存在.相似地,如果具有离散联合分布,其p.f.为并且存在均值,则的均值为

**例题4.1.15** 确定多变量随机函数的期望值 **定理4.1.2**的示例 均匀分布

4.2 期望的属性 2020年2月25日11点11分

**定理4.2.1 线性函数** 如果,其中和都是有限常数,则

**例题4.2.2** 计算线性函数的期望

**推论 4.2.1** 如果且概率为1,则.

**例题4.2.3** 投资 依赖**例题3.3.7**

**定理4.2.2** 如果存在一个常数使得,则.如果存在一个常数使得,则.

根据定理4.2.2得到如果,则.

**定理4.2.3** 假设,并且或.则**该定理的证明有点难,尤其是连续分布的证明需要注意**.

**定理4.2.4** 如果设是个随机变量且每一个期望是有限的(),则

**推论4.2.2** 假设期望是有限的().对所有常数和b存在下列关系:

**例题4.2.3** 投资组合 **例题4.2.1**的扩展

**例题4.2.4** 不放回抽样 该例题比较**古怪**

**定义4.2.1** **凸函数** 一个以向量为参数的函数是凸的,如果对于每一个,和每一个，,满足下列不等式

**定理4.2.5 Jensen不等式** 设是一个凸函数,设是一个有限均值的随机变量.则

**例题4.2.5** 放回抽样 **例题4.2.4**的扩展,两者期望相同,概率不同 二项分布

**例题4.2.6** 匹配的期望数值 **例题1.10.X**的新解 伯努利分布 **值得深度思考**

**定理4.2.6** 如果是个独立随机变量且每一个期望是有限的(),则

**证明过程需要看懂**.

**例题4.2.7** 计算随机变量组合的期望 **定理4.2.4**和**定理4.2.6**的示例

**例题4.2.8** 重复过滤 独立随机变量

**定理4.2.7 整型随机变量** 设是随机变量只能取值则

**证明过程非常有趣**.

**例题4.2.9** 期望的实验数 **定理4.2.7**的示例 实验直到成功的期望

**定理4.2.8 一般非负随机变量** 设是一个非负随机变量,其c.d.f.为.则

**例题4.2.9** 期望等待时间 **定理4.2.8**的示例 频率分布

**4.3 方差** 2020年2月26日09点22分

**例题4.3.1** 股票价格变化 均匀分布,独立连续随机变量

**定义4.3.1 方差/标准偏差** 设是一个随机变量,其有限均值为.的方差,标记为,定义如下:

如果具有无限均值或者如果的均值不存在,则我们说不存在.如果方差存在,则的标准偏差是的非负平方根.

当仅讨论一个随机变量时,通常用符号表示其标准偏差,方差用表示.当讨论多个随机变量时,随机变量的名称作为符号的下标包括在内,例如，将是的标准偏差，而将是的方差.

**例题4.3.2** 股票价格变化 **例题4.3.1**的方差计算

**注意:方差仅取决于分布.**随机变量的方差和标准差仅取决于的分布,正如的期望仅取决于分布.事实上,所有可以从p.f.或者p.d.f.计算的东西仅取决于分布.具有相同分布的两个随机变量将具有相同的方差,即使它们彼此无关.

**例题4.3.3** 离散随机变量的方差和标准差

**定理4.3.1 计算方差的另一种方法** 对于每一个随机变量,.

**例题4.3.4** 离散随机变量的方差和标准差 **定理4.3.1**对**例题4.3.3**的应用

方差(以及标准偏差)提供了围绕其均值分布的扩散或分散的确定性.较小的方差值表示概率分布紧密集中在附近;较大的方差值通常表示概率分布在附近具有较宽的范围.然而,将非常小但正的概率放置在距离实线上原点足够远的地方,可以使分布的方差及其均值任意大.

**例题4.3.5** 伯努利分布的改版 上述文字的示例

**定理4.3.2** 对每一个,.如果是一个有界随机变量,则必定存在并且是有限的.

**定理4.3.3** 当且仅当存在一个常数使得.**证明过程需要理解**.

**定理4.3.4** 对于常数和,设.则

且..

**例题4.3.6** 离散随机变量的方差和标准差 **定理4.3.4**对**例题4.3.3**的应用

**定理4.3.5** 如果是个独立随机变量且具有有限均值,则(**证明过程需要理解**)

**推论4.3.1** 如果是个独立随机变量且具有有限均值,设和b是任意常数,则

**例题4.3.7** 投资组合 投资股票的分析 有使用价值

**例题4.3.8** 投柯西分布 由于柯西分布的均值不存在,那么方差也必定不存在,那如何量化分布呢?

**定义4.3.2 四分位距[Interquartile Range IQR]** 设是一个随机变量且分位函数为.则四分位距(IQR)定义为.

换句话说,IQR是包含分布中间一半区间的长度.

**例题4.3.9** 投柯西分布 **例题4.3.8**的解

**4.4 矩** 2020年2月27日13点08分

**定义4.1.0 矩** 对每一个随机变量和每个正整数,期望值称为的阶矩.特别地,根据该定义,的均值是的一阶矩.

**定理4.4.1** 如果对于某个正整数成立,则对任意满足的正整数成立.

**中心距**：假设是一个随机变量且.对于每一个正整数，期望被称为的阶中心距,或关于均值的阶距.特别的,根据该定义,的方差就是的二阶中心距.

对于任意分布,一阶中心距必须为0,因为

**例题4.4.1** 对称p.d.f. 矩和中心距 **稍微有点难**

**定义4.4.1 偏度(skewness)** 设是一个随机变量,其,标准差为，并且具是有限三阶距.则的偏度定义为.

**例题4.4.2**  二项分布的偏度

**定义4.4.2 矩量母函数[Moment Generating Function]** 设是一个随机变量.对每一个实值,定义

函数被称为的距量母函数.

**注意:的距量母函数仅取决于的分布.**由于m.g.f.是函数的期望值,它必须仅取决于的分布.如果和具有相同的分布,它们必须具有相同的m.g.f.

**定理4.4.2** 设是一个随机变量,其m.g.f.在点附近的开区间是有限的.则对于每一个整数,的阶矩,是有限的且在处等于阶导数.也就是,

**例题4.4.3**  计算m.g.f

**定理4.4.3** 设是一个随机变量, m.g.f.为;设,其中和是给定的常数;设为的m.g.f..对于每一个值使得是有限的,则

**例题4.4.4**  计算线性函数的m.g.f

**定理4.4.4** 假设是个独立随机变量,且,设表示的m.g.f.设,设为的m.g.f..则对每一个值使得在是有限的,则

**参数为和的二项分布的矩量母函数**:

**定理4.4.5** 如果随机变量和的m.g.f.是有限的且在点附近的开区间对所有的值都是相等的,则和的概率分布也是相等的.

**定理4.4.6** 如果和是独立随机变量,如果具有参数为和的二项分布(),则也是参数为和的二项分布.

**4.5 均值和中位数** 2020年2月27日15点10分

**定义4.5.1 中位数** 设是一个随机变量,每一个满足下列属性的被称为分布的中位数:

理解该定义的另一种方式是,中位数是满足以下两个要求的点:首先,如果被左边的值包含,则

其次,如果被右边的值包含,则

如果存在一个使得,即,如果实际上将总概率分成了两个相等的部分,则当然将是 的分布(请参阅练习16).

**例题4.5.1**  离散分布的中位数

**例题4.5.2**  离散分布的多个中位数

**例题4.5.3**  连续分布的中位数

**例题4.5.4**  连续分布的多个中位数

**例题4.5.5** 州彩票 讨论中位数与均值的区别 **建模过程有点难懂**

**例题4.5.6** 平均收入 用家庭平均收入示例说明均值和中位数的差别

**定理4.5.1 一对一函数** 设是一个随机变量,其取值范围再实数区间I上.设是一个定义在区间I上的一对一函数.如果是的中位数,则是的中位数.(**证明过程很有价值**)

**定义4.5.2 均方误差/M.S.E.** 数值被称为估计量的均方误差.

**定理4.5.2** 设是一个随机变量,其有限方差为,设.对于每一个值,下列不等式成立

**例题4.5.7** 州彩票 **例题4.5.5**的解 **解题过程用了技巧**

**定义4.5.3 平均绝对误差/M.A.E.** 数值被称为估计量的平均绝对误差.

**定理4.5.3** 设是一个有限均值的随机变量,设是分布的中位数.对于每一个值,下列不等式成立(**证明过程很精彩**)

**例题4.5.8** 州彩票 **例题4.5.5**的解 说明均值和中位数的差异 **解题过程用了技巧**

**例题4.5.9** 预测离散均匀随机变量 说明均值和中位数的差异

**4.6 协方差和相关性** 2020年2月28日09点33分

**例题4.6.1** 成绩测试 引出协方差和相关性的概念

**定义4.6.1 协方差** 设和是有限均值的随机变量.设.和的协方差标记为,定义为

如果公式(4.6.1)期望存在.

**例题4.6.2** 成绩测试 **例题4.6.1**的解 计算协方差和相关性

**定理4.6.1** 对所有随机变量和满足和条件时,下列等式成立

**定义4.6.2 相关性** 设和是有限方差的随机变量,其方差分别为和.和的相关性标记为,定义为:

**定理4.6.2 施瓦茨不等式** 对于所有随机变量和,当存在时,下列不等式成立:

除此之外,如果等式右边是有限的,则等号成立时当且仅当存在非零常数和使得且概率为1.(**证明过程没有理解透彻**)

**定理4.6.3 柯西-施瓦茨不等式** 设和是有限方差的随机变量,则

且

除此之外,等式等号成立当且仅当存在非零常数和使得且概率为1.

**定义4.6.3 正相关/负相关/非相关** 如果则称和是正相关,如果则称和是负相关,如果则称和是非相关.

**例题4.6.3** 成绩测试 **例题4.6.2**相关性说明

**定理4.6.4** 如果和是独立随机变量且,则

**例题4.6.4** 非独立但非相关的随机变量

**例题4.6.5** 圆盘内的均匀分布

**定理4.6.5** 假设是一个随机变量且具有,设,为常量,.如果,则.如果,则.

**注意:相关性只能测量线性关系**.的大值表示和接近线性相关,因此紧密相关.但是的小值并不意味着和不太接近.

**定理4.6.6** 如果和是随机变量且具有.则

**推论4.6.1** 设为常数.在定理4.6.6条件下

另一个特别有用的推论是

**例题4.6.6** 投资组合 **例题4.3.7**加入相关性后的结果

**定理4.6.7** 如果是随机变量且具有则

**推论4.6.2** 如果是非相关的随机变量(当时,和是非相关的),则

如果是独立随机变量,该推论也适用.

**4.6节内容和习题出现的证明特别多,有时间可以看看**.

**4.7 条件期望** 2020年2月29日13点33分

**例题4.7.1** 住户调查 图表法

**定义4.7.1 条件期望/均值** 设和是随机变量,具有有限均值.时的条件期望(或条件均值)标记为,定义为时的条件分布的期望.

例如,如果是连续分布且条件p.d.f.为,则

类似的,如果是离散分布且条件p.f.为,则

**定义4.7.2 条件均值作为随机变量** 设代表公式(4.7.1)或(4.7.2)关于的函数.定于符号为的均值,该函数被称为给定时的条件均值.换句话说,是一个随机变量其值当时等于.显然,我们可以类似地定义和.

**例题4.7.2** 住户调查 **例题4.7.1**的解 图表法

**例题4.7.3** 临床试验 回顾二项分布的均值

**注意:给定时的条件均值是随机变量**.由于是随机变量的函数,因此它本身是具有自己概率分布的随机变量,可以从的分布中导出.另一方面,是的一个函数,可以像其它任何函数一样进行操纵.两者之间的联系是,当用中的代替随机变量时,结果为.

**定理4.7.1 总概率期望定律** 设和是随机变量,具有有限均值.则(**证明过程需要理解**)

**例题4.7.4** 住户调查 **例题4.7.2**的扩展,从中得到的分布等于x的边际分布. 图表法

**例题4.7.5** 临床试验 **例题4.7.3**的扩展,提出新的问题

**例题4.7.6** 从均匀分布选取点

**定理4.7.2** 设和是随机变量,设.在给定时的条件分布等于在时的条件分布.

当和具有连续的联合分布时,定理4.7.2的结果为

定理4.7.1同样也暗示对任意两个随机变量和,

**例题4.7.7** 线性条件期望 假设和b是常量.则.

**定义4.7.3 条件方差** 对每一个给定值,设为给定时的条件分布方差,即

我们称是给定时的条件方差.

**定理4.7.3** 估计量时使得最小化.(**证明过程需要反复细品**)

如果观察到的值并且的估计值为,根据定义4.7.3，则该估计值的M.S.E.为.根据等式(4.7.6),如果使用函数进行估计,则在所有可能的值上取平均值的总M.S.E.将为.(**无法理解,所以稍后待续**)

**定理4.7.4 总概率方差定律** 如果和是任意随机变量，且必要的期望和方差都存在，则.

**例题4.7.8** 临床试验 **例题4.7.4**的解,**目前未知遇到的综合性最强,最复杂例题,需要结合之前的定理和定义逐条理解**.

**本章内容学习并不好,很多证明过程和定理需要手动推导一遍才能加深印象**.

**4.8 效应** 2020年2月29日17点01分

**例题4.8.1** 赌徒的选择 通过示例解释效应理论

**定义4.8.1 效应函数** 个人效应函数是将赋值给每一个可能值,表示个人获取数量的实际价值.

**例题4.8.2** 赌徒的选择 **例题4.8.1**的扩展 效应函数说明,引出下列定义

**定义4.8.2 最大化期望效应** 我们说如果以下条件成立，一个人通过最大化预期效应在博弈之间选择.存在一个效应函数,个人必须在博弈和之间做出选择,如果,他对于更喜欢,如果,则二者之间任选其一.

**例题4.8.3** 线性效应函数

**例题4.8.4** 三次效用函数

**例题4.8.5** 对数效用函数

销售彩票

**例题4.8.6** 二次效用函数

**例题4.8.7** 平方根效用函数

一些统计决策问题

这部分内容对第7-11章关联比较大,在研究统计学内容时需要回顾

**例题4.8.8** 估计随机变量 引出问题

**例题4.8.9** 限制随机变量 对上一节临床试验的扩展

**例题4.8.10** 投资 **例题4.2.2**的效应评估

**本节内容只给出了效应函数的期望计算,但是没有给出如何确定随机变量的效应函数**.