**7.1 统计推断** 2019年8月15日09点53分——2020年4月7日10点56分

**例题7.1.1** 电子元器件的使用寿命 指数分布,引出问题

**定义7.1.1 统计模型** 统计模型包括对感兴趣的随机变量的识别(可观察的和仅假设可观察),可观察的随机变量的联合分布或一系列可能联合分布的规范,对这些分布的任意参数的识别.假设这些分布未知并且是假设可观察的,指定未知参数的(联合)分布规范.当我们将未知参数视为随机变量时,以为索引的可观察随机变量的联合分布应理解为给定的可观察随机变量的条件分布.(这段定义翻译不好,需要优化)

在**例题7.1.1**中,可观察到的感兴趣的随机变量来自序列,故障率是假设可观察的.的可能联合分布的族由参数索引.对应的可观察联合分布是具有参数指数分布的i.i.d随机变量.这也是是给定的条件分布,因为我们将视为随机变量,其分布是参数1和2的伽马分布.

**定义7.1.2 统计推断** 统计推断是产生有关统计模型的某些或所有部分的概率陈述的过程.

所谓“概率陈述”,是指一种利用概率论中任何已在本文中进行过讨论或稍后进行讨论的概念的声明.一些典型示例,比如均值,条件均值,分位数,方差,给定一个变量另一个随机变量的条件分布,事件的概率,给定某种东西的事件的条件概率等等.在示例7.1.1中,下面是一些人们可能希望进行的统计推断的示例:

1. 生成一个随机变量(的一个函数)使得.
2. 生成一个随机变量,我们期望它接近.
3. 计算接下来10个生命周期的平均值至少为2的可能性.
4. 在观察之后的概率.

在定义7.1.1中,我们区分了可观察和假设可观察随机变量.对于可观察随机变量,如果我们付出了必要的努力来观察它,我们基本上可以确定它可以观察到.假设可观察随机变量用于需要无限资源进行观察的,例如前个可观察的样本平均值的极限().在本文中,此类假设可观察的随机变量将对应于示例7.1.1中的可观察变量的联合分布参数.由于这些参数在我们将要看到的许多类型的推理问题中占有重要地位,因此有必要对参数的概念进行形式化.

**定义7.1.3 参数/参数空间** 在统计推断问题中,用于决(确)定目标随机变量联合分布的特征或组合特征被称为分布参数.参数所有可能值的集合或参数向量被称为参数空间.

**例题7.1.2** 临床试验 对患者恢复概率的参数空间的解释

**例题7.1.3** 临床试验 回顾**例题2.1.4**和**例题4.7.8**,统计推断的示例说明.

**例题7.1.4** 辐射例子 回顾**例题5.7.8**

**例题7.1.5**蚤甲虫的人体测量学 回顾**例题5.10.2**

**例题7.1.6** 均值区间 回顾**例题5.6.8**

**例题7.1.7** 评委选择中的歧视 回顾**例题5.8.4**

**例题7.1.8** 队列中的服务时间 回顾**例题5.7.3**和**例题5.7.4**

2020年4月13日10点26分

**例题7.1.9** 圆形轴承失效时间 回顾**例题5.6.9**并提出新问题

**定义7.1.4 统计** 假设目标可观测的随机变量是.设是任意个实数变量的实值函数. 则随机变量被称为统计.

**例题7.1.10** 圆形轴承失效时间 **例题7.1.9**的扩展,举例说明什么是统计函数(**公式好像存在书写错误**).

**例题7.1.11** 均值区间 **例题7.1.6**的扩展,举例说明什么是统计.

**例题7.1.12** 参数作为随机变量的极限 因为参数是随便变量的函数,因此参数本身可以作为随机变量,这种观点被称为贝叶斯主义.

本章介绍了许多关于统计推断的概念知识和定义,虽然提供了大量例题,但这些例题仅仅是做统计推断的定义作说明.**有价值的依然是那些大段解释和说明,需要反复阅读理解**.

**7.2 先验和后验分布** 2020年4月20日09点42分

**例题7.2.1** 电子元器件的使用寿命 说明先验分布的概念

**定义7.2.1 先验分布/p.f/p.d.f** 假设存在一个参数为的统计模型.如果是随机的,则在观察其它目标随机变量之前为分配的分布被称为先验分布.如果参数空间是可数的,则先验分布是离散的并且它的p.f.被称为先验p.f.如果先验分布是连续的,则它的p.d.f.被称为先验p.d.f..我们使用符号来标记先验p.f.或先验p.d.f.作为的函数.

当我们将参数视为随机变量时,名称“先验分布”仅仅是参数边际分布的另一个名称.

**例题7.2.2** 公平或两面硬币 举例说明先验分布,需要仔细思考

**例题7.2.3** 残次零件的比例 举例说明先验分布

**例题7.2.4** 荧光灯的使用寿命 举例说明先验分布

**敏感性分析和不当先验** 在第84页的示例2.3.8中,我们看到了一种情况,其中两个非常不同的先验概率集合用于事件收集.但是,在我们观察到数据之后,后验概率非常相似.在第330页的示例5.8.4中,我们为参数使用了大量先验分布,以便查看先验分布对单个重要事件的后验概率有多大影响.通常的做法是比较由几个不同的先验分布产生的后验分布,以了解先验分布对重要问题的答案有多大影响.这种比较称为敏感性分析.

通常情况下,观察到数据后,不同的先验分布不会产生太大差异.如果存在大量数据,或者比较的先前分布非常分散,则尤其如此.这一观察有两个重要含义:首先,如果有大量数据,那么不同的实验者可能不会同意先验分布,因为先验分布已经变得不重要了;其次,如果指定哪个先验分布的关系不大,实验人员可能不太愿意花时间指定一个先验分布.不幸的是,如果没有指定某些先验分布,就无法计算给定数据的参数的条件分布.

**例题7.2.5** 荧光灯的使用寿命 **例题7.2.4**的扩展,引出问题

**定义7.2.2 后验分布/p.f./p.d.f.** 考虑一个参数为的统计推断,并且已经观察到随机变量.在给定条件下的条件分布被称为后验分布.在给定条件下的条件p.f.或p.d.f.被称为的后验p.f.或后验p.d.f.并且被标记为.

当我们将参数视为随机变量时，名称“后验分布”仅仅是在给定数据情况下参数条件分布的另一个名称.

**定理7.2.1** 假设总分布的p.d.f或p.f.为,从中选取个随机变量.同样假设参数值未知且的先验p.d.f.或p.f.为.则的后验p.d.f.或p.f.为

其中是的边际联合p.d.f或p.f.,即

因此,也可以写作为

以上三个公式源于**定理7.2.1**的证明过程*.*

**例题7.2.6** 荧光灯的使用寿命 **例题7.2.5**的**解**,该例题之后的讨论也值得阅读.

**定义7.2.3 似然函数** 当观察到随机样本联合p.d.f.或联合p.f..在给定值条件下作为函数,该函数被称为似然函数.

**例题7.2.7** 残次零件的比例 **例题7.2.3**的解.**该例题值得细品**.

**例题7.2.8** 荧光灯的使用寿命 **例题7.2.6**的进一步扩展,**对似然函数的进一步研究**,**需要结合例题之前的大段文字来理解解题过程**.

**7.3 共轭先验分布** 2019年8月28日09点58分——2020年4月28日09点43分

**例题7.3.1** 临床试验 **例题5.8.5**的回顾 贝塔分布

**定理 7.3.1** 假设从参数为的伯努利分布中选取个随机样本,其中.同样假设先验分布是参数为的贝塔分布.则在给定条件下后验分布是参数为和的贝塔分布.

**更新后验分布** 定理7.3.1的一个含义如下:假设大量物品中有缺陷物品的比例是未知的,的先验分布是具有参数和的贝塔分布,并且从中一次选择个物品.假设物品在给定的条件下是独立的.如果检查的第一物品有缺陷,则的后验分布将是参数为和的贝塔分布.如果第一项无缺陷,则后验分布将是参数为和的贝塔分布.该过程可以按以下方式继续:每次检查一个物品后,将当前的后验贝塔分布更改为新的贝塔分布,其参数或参数的值增加一个单位.每次发现有缺陷的物品,的值增加一个单位,若发现无缺陷的物体,则的值增加一个单位.

**定义 7.3.1 共轭族/超参数** 设,是给定下的条件i.i.d，并且它们具有共同p.f.或p.d.f. .设是参数空间上的可能分布族.假设先验分布不论从怎样选取,以及无论选取多少观察样本,同样,无论观察值是什么,后验分布始终是成员.则被称为来自分布的样本的先验分布共轭族.族也被称为自分布的闭采样(closed under sampling).最终,如果中的分布被其它参数参数化,则于先验分布关联的参数被称为先验超参数,于后验分布关联的参数被称为后验超参数.

**例题7.3.2** 后验贝塔分布的方差

**例题7.3.3** 护士使用手套 举例说明不同的贝塔先验分布对后验分布的影响

**例题7.3.4** 客人到达的数量 研究泊松分布的参数,引出新的问题

**定理7.3.2** 假设从参数为的泊松分布中选取个随机样本,其中未知.同样假设先验分布是参数为的伽马分布.则在给定条件下后验分布是参数为和的伽马分布.(**证明过程需要看懂**)

**例题7.3.5** 客人到达的数量 **例题7.3.4**的解

**例题7.3.6** 后验伽马分布的方差

**例题7.3.7** 汽车排放的污染物 **例题5.6.1**的扩展,研究正态分布的均值,引出新的问题

**定理7.3.3** 假设从均值,方差的正态分布中选取个随机样本,其中未知,而是已知的.同样假设先验分布是均值为,方差为的正态分布.则在给定条件下后验分布是均值为,方差为的正态分布.其中(**证明过程需要理解,不难**)

**注意,在后有一大堆关于均值的讨论,需要补上**.

**例题7.3.8** 汽车排放的污染物 **例题7.3.7**的解.

**例题7.3.9** 后验正态分布的方差

**定理7.3.4** 假设从参数为的指数分布中选取个随机样本,其中未知.同样假设先验分布是参数为的伽马分布.则在给定条件下后验分布是参数为和的伽马分布(**该定理的证明过程域之前三个定理(特别是定理7.3.2)类似,需要掌握这种证明方式**).

**定义7.3.2 不适当先验(Improper Prior)** 设是一个非负函数,其定义域包含统计模型的参数空间.假设.如果我们假装是先验p.d.f.,则我们对于使用了一个不适当先验.

**注意**:**本节内容的笔记部分并不完善, 例题7.3.8后的内容需要抽时间仔细整理**.

7.4 贝叶斯估算器 2019年8月29日10点18分

定义7.4.1 估算器/估值 设是一组观察数据，其联合分布是由从实数子集中取值的参数决定的. 参数的估算器是一个实值函数. 如果被观察到，则被称为的估值.

定义7.4.2 损失函数 损失函数是二元变量的实值函数, 标记为, 其中是一个实数. 该函数的意义是当参数等于且估值等于时的统计损失.

定义7.4.3 贝叶斯估算器/估值 设是一个损失函数. 对于的每一个可能值, 设是一个值使得最小化. 则被称为的贝叶斯估算器. 一旦=被观察到，被称为

贝叶斯估值.

定义7.4.4 平方误差损失函数

推论7.4.1 设是一个实值参数. 假设平方误差损失函数(7.4.4)被使用，并且后验均值是有限的. 则贝叶斯估算器是.

定义7.4.5 绝对误差损失函数

推论7.4.2 当绝对误差损失函数被使用时，实值参数的贝叶斯估算器等于后验分布的中位数.

定义7.4.6 一致估算器 在情况下，一组估算器概率收敛于被估算参数的未知值被称为一致估算序列。

定义7.4.7 估算器/估值 设是一组观察数据，其联合分布是由从维空间子集中取值的参数决定的. 设是从到维空间的函数. 定义. 估算器是一个从维空间取值的函数. 如果被观察到，则被称为的估值.

7.5 最大似然估算器 2019年8月30日10点50分

定义7.5.2 最大似然估算器/估算 对每一个可能的观察向量, 设在情况下使得似然函数最大化, 设为定义在这种方式下的的估算器. 估算器被称为的最大似然估算器. 当被观察时，值被称为的最大似然估值.

7.6 最大似然估算器的属性 2019年9月2日10点43分

定理7.6.1 M.L.E.的不变性 如果是的最大似然估算器并且g是一对一函数，则是的最大似然估算器.

定义7.6.1 M.L.E.函数 设是参数的任意函数, 并且设G是在函数g下的图像. 对于每一个, 定义和

最终，定义的M.L.E，其中

定理7.6.2 设是的M.L.E. 设是关于的函数. 则的M.L,E是.

定义7.6.3 矩方法 假设从k维参数的分布中选取n个随机变量, 并且该分布至少包含k个有限矩. 当 设. 假设函数是一对一函数. 设是其反函数，对于所有的，

通过定义样本矩, 其中 矩估算器方法是.

定理7.6.3 假设是i.i,d, 其分布是由k维参数向量决定的. 假设该分布对于所有前k个矩存在且均为有限值. 同样假设定义在7.6.3中的反函数M是连续的. 则基于矩估算器方法的序列是估算器是连贯序列.

7.7 充分统计量 2019年9月4日10点21分

定义7.7.1 充分统计量 设是随机样本，来自参数为的分布. 设T是一个统计量. 假设对于每一个和T的每一个t值, 在给定T=t条件下的条件联合分布只依赖t而不依赖. 也就是说，对于每一个t值, 在给定T=t条件下的条件分布对于任意值保持不变. 则我们称T是参数的充份统计量.

定理7.7.1 因式分解 设是连续分布或离散分布的随机样本，该分布的p.d.f.或p.f.为, 其中值是未知的并且属于给定的参数空间. 统计量是充分统计量当且仅当的联合p.d.f.或联合p.f. 对于所有的值和所有的值可以被分解称如下公式：

函数和是非负的，可能依赖但是不依赖，依赖于但是依赖通过统计量观察到的值.

推论7.7.1 统计量是充分的当且仅当无论使用哪种先验分布，后验分布只依赖通过T值的观察数据.

7.8 联合充分统计量 2019年9月5日10点12分

定义7.8.1 联合统计量 对于每一个和每一个的可能值, 假设在给定条件下的条件联合分布不依赖值. 则被称为的联合充分统计量.

定理7.8.1 联合充分统计量的因式分解 设是n个实变量函数. 统计量是的联合充分统计量当且仅当联合p.d.f.或联合p.f. 对于所有的值和所有的值可以被分解称如下公式：

定义7.8.2 有序统计量 假设是某个分布中的样本空间. 设是这些随机样本中最小值, 设是第二最小值，是第三最小值，一次类推. 在这种方式中, 表示样本中最大值, 表示下一个最大值. 随机变量被称为样本的有序统计量.

定理7.8.2 有序统计量在随机样本中是充分的 设是某个分布中的样本空间, 该分布p.d.f.或p.f.是. 则有序统计量对于是联合充分的.

定义7.8.3 极小（联合）充分统计量 统计量是极小充分统计量当且是充分的并且是其它充分统计量的函数. 统计向量是极小充分统计量当且是充分的并且是其它充分统计量的函数.

定理7.8.3 M.L.E和充分统计量 设对于是充分统计量. 则的M.L.E. 仅依赖通过统计量的观察数据.

定理7.8.4 贝叶斯估算器和充分统计量 设对于是充分统计量. 则的每一个贝叶斯估算器 仅依赖通过统计量的观察数据.

8 估算的抽样分布

8.1 统计量的抽样分布 2019年9月6日10点08分

定义8.1.1 抽样分布 假设从参数未知的分布中选取一个随机向量. 设是和任意可能的函数. 即. 的分布被称为的抽样分布. 我们使用符号表示从该分布中所计算的的均值.

8.2 卡方分布 2019年9月9日10点05分

定义8.2.1 分布 对于每一个正数, 参数, 的伽马分布被称为自由度为的分布. 其p.d.f.为

当时, .

定理8.2.1 均值和方差 如果随机变量X是自由度为的分布, 则并且.

定理8.2.2 如果随机变量是独立的并且假设是自由度为的分布, 则之和是自由度为的分布.

定理8.2.3 设是标准正态分布. 则随机变量是自由度为1的分布.

推论8.2.1 如果随机变量是符合标准正态分布的i.i.d., 则平方和是自由度为m的分布.

8.3 采样均值和采样方差的联合分布 2019年9月9日10点42分

定理8.3.1 假设从均值为, 方差为的正态分布中选取随机样本. 则样本均值和样本方差是独立随机变量, 是均值为, 方差为的正态分布, 是自由度为n-1的分布.

定理8.3.4 假设随机变量是i.i.d并且每一个都是标准正态分布. 假设A是一个正交矩阵, 且. 则随机变量同样也是i.i.d., 且每一个也都是标准正态分布，且.

8.4 t分布. 2019年9月10日10点29分

定义8.4.1 考虑两个独立随机变量和, 是自由度为m的分布, 是标准正态分布. 假设随机变量被定义为下列公式

则的分布被称为自由度为m的t分布.

定理8.4.1 概率密度函数. 自由度为m的t分布的p.d.f.是

定理8.4.2 假设从均值为, 方差为的正态分布中选取随机样本. 设为样本均值，定义

则是自由度为n-1的t分布.

8.5 置信区间 2019年9月10日11点34分

定义8.5.1 置信区间 设是从某个参数(或参数向量)为的分布中选取的一个样本. 设是的实值函数. 设是两个统计量并且对于所有的值满足下列属性,

则随机区间被称为的系数置信区间或100百分比置信区间.

定理8.5.1 正态分布均值的置信区间 假设从均值为, 方差为的正态分布中选取随机样本. 对于每一个, 闭区间恰好是的系数置信区间：

定义8.5.2 单边置信区间 设是从参数(或参数向量)为的分布中选取的一个样本. 设是的实值函数. 设是一个统计量并且对于所有的值满足下列属性,

则随机区间被称为的单边系数置信区间或单边100百分比置信区间. 同样, 也被称为的系数下界置信区间或100百分比下界置信区间. 类似的，如果B是一个统计量使得

则是的单边系数置信区间或单边100百分比置信区间, 并且B是的系数上界置信区间或100百分比是界置信区间. 如果不等式“”对于所有的满足公式(8.5.5)或公式(8.5.6)，那么相应的置信区间和置信极限被称为恰好.

定理8.5.2 正态分布均值的单边置信区间 假设从均值为, 方差为的正态分布中选取随机样本. 对于每一个, 下列统计量恰好分别是的上界和下界系数置信区间：

定义8.5.3 枢轴 设是从参数(或参数向量)为的分布中选取的一个样本. 设是一个随机变量其分布与一致. 则被称为枢量(或简称为枢轴).

定理8.5.3 来自枢轴的置信区间 设是从参数(或参数向量)为的分布中选取的一个样本. 假设枢轴存在. 设是的c.d.f., 假设G是连续的. 假设公式(8.5.7)中的函数存在，并且假设对于每一个在上是严格递增的. 设使得. 则下列统计量是恰好系数置信区间的端点：

如果对于每一个在上是严格递减的, 则需要交换和的定义.

8.6 正态分布样本的贝叶斯分析

定义8.6.1 正态分布的精度 正态分布的精度被定义为方差的倒数，即.

定理8.6.1 假设从均值为, 未知精度为的正态分布中选取随机样本. 同样假设和的联合先验分布定义如下：在给定的条件下的条件分布是均值为,精度为的正态分布, 并且的边际分布是参数为和的伽马分布. 则在给定条件下，和的联合后验分布定义如下: 在给定条件下的条件分布是均值为,精度为的正态分布，其中

且的边际分布是参数为和的伽马分布, 其中

定义8.6.2 正态伽马分布族 设和是随机变量. 假设在给定条件下的条件分布是均值为, 精度为的正态分布. 同样假设的边际分布是参数为和的伽马分布. 则我们说和的联合分布是混合参数为的正态伽马分布.

定理8.6.2 均值的边际分布. 假设和的先验分布是混合参数为的正态伽马分布. 则的边际分布与t分布有如下方式的关系：

即，的边际分布是自由度为2的t分布.

定理8.6.3 假设和是混合参数为的联合正态伽马分布. 如果, 则. 如果, 则

8.7 无偏移估算器 2019年9月12日10点34分

定义8.7.1 无偏移估算器 估算器是参数函数的无偏移估算器当且对每一个都成立. 其它估算器则被称为偏移估算器. 估算器期望值与之间的差值被称为估算器的偏移量. 也就是说, 作为估算器的偏移等于, 并且是无偏移的当且仅当偏移值对所有的恒等于0.

推论8.7.1 设是有限方差估算器. 则作为的估算器的M.S.E等于其方差加上偏差的平方.

定理8.7.1 一般分布的采样 设是从参数(或参数向量)为的分布中选取的一个样本. 假设分布的方差是有限的. 定义. 下列统计量是方差的无偏移估算器:

8.8 费希尔信息 2019年9月13日09点28分

定义8.8.1 随机变量中的费希尔信息 设X是一个随机变量其分布依赖于参数, 取值于实线开区间. 设X的p.f.或p.d.f.为. 假设使得的的集合对于所有的值都一样, 并且作为的函数是两次可微分的. 则费希尔信息在随机变量X中被定义为:

因此, 如果是一个p.d.f., 则

其中

定理8.8.1 根据定义8.8.1的条件. 同样, 假设对于的二阶导数可以通过反转积分和微分的顺序来计算. 则费希尔信息等于

费希尔信息另一个表达式是

定义8.8.2 随机样本中的费希尔信息 假设是从p.f.或p.d.f.为的分布中选取的一组随机样本, 其中值必须位于实线开区间上. 设是X的联合p.f.或联合p.d.f. 定义

假设使得的集合对于所有的值都是一样, 并且作为的函数是二次可微分的. 则费希尔信息在随机样本中定义为

对于连续分布, 费希尔信息在整个样本中由下列n维积分定义：

对于离散分布, 把n维积分替换称n个累加之和.

额外的, 我们再次假设导数可以在积分下通过, 则我们可以把用下列两种方式表达：

或

定理8.8.2 在定义8.8.1.和8.8.2条件下, 存在

定理8.8.3 克拉美-罗不等式 假设是从p.d.f.为的分布中选取的一组随机样本. 同样假设所有关于的假设都是连续的. 设是一个有限方差的统计量. 设. 假设是关于的可微分函数. 则

公式等号成立当且仅当存在函数和可能依赖但不依赖并且满足关系

推论8.8.1 无偏估算器方差的克拉美-罗下界 在定理8.8.3的假设条件下. 设是的无偏估算器. 则

定义8.8.3 有效估算器 估算器被称为期望的有效估算器当且公式对于所有的存在等号成立.

定理8.8.4 有效估算器的渐进分布. 在定理8.8.3的假设条件下. 设是其均值的有效统计量. 假设从不为0. 则

的渐进分布是标准正态分布.

定理8.8.5 M.L.E的渐进分布 假设在任意问题中M.L.E. 是通过求解等式得到的, 并且额外的二阶导数和三阶导数都存在并满足特定条件. 则的渐进分布是标准正态分布.

定理8.8.6 有效估计 在定理8.8.5中关于似然函数的假设条件下, 假设是估算器的一个序列使得在分布上收敛于某些分布(具体是哪些分布并不重要). 使用作为开始值并朝着发现 M.L.E的方向执行一步牛顿法. 将这一步的结果称为. 则的渐进分布是标准正态分布.

定义8.8.4 向量参数的费希尔信息 假设是从p.d.f.为的分布中选取的一组随机样本, 其中参数的值必须位于维实数空间的开区域中. 设是的联合p.d.f.或联合p.f. 定义

假设使得的集合对于所有的值都是一样, 并且作为的函数是二次可微分的. 则费希尔信息矩阵在随机样本中被定义为矩阵, 元素等于

维实数空间的克拉美-罗不等式：

同样，当公式等号成立当且仅当是下列向量的线性函数:

第九章 假设验证

9.1 假设验证的问题 2019年9月16日10点24分

**定义9.1.1 零假设和备择假设** 考虑一个包含参数的统计问题, 其中未知且位于特定的参数空间. 现假设被分割称两个不同子集和, 统计者想要知道是否位于或. 设表示的假设, 表示的假设. 因为子集和是离散的并且, 因此假设和只能有一个为真. 而统计者必须决定哪一个为真. 这类有两种（可能）决策的问题被称为*假设验证*. 决定选择哪个假设的过程为称为*验证过程*或简称为*验证.* 我们一般将称为零假设, 称为备择假设. 当执行一个验证, 如果我们决定位于, 则我们说拒绝. 如果我们决定位于, 我们说不拒绝.

**定义9.1.2 简单和复合假设** 如果只包含单个值, 则被称为简单假设. 如果包含不止一个值, 则被称为复合假设.（其中）. 例如, 一个简单的零假设的形式必须是

**定义9.1.3 单面和双面假设** 设是一个一维参数. 单面零假设的形式为, 其对应的单面备择假设是. 当零假设是简单的，如公式(9.1.1)所示, 那么备择假设通常是双面的, 即.

一般的，考虑我们想要验证下列假设：

假设统计员在选取假设之前, 她从未知参数为的总体分布中观察到了一个随机样本. 我们设为维随机向量的样本空间. 换句话说, 是所有可能随机样本的集合. 在这种类型的问题中, 统计员可以通过将分割称两个子集来设定测试过程. 子集包含拒绝的值, 另一个子集包含不拒绝的值.

**定义9.1.4 临界域** 上述所定义的集合被称为测试的临界域.

**定义9.1.5 统计测试/拒绝域** 设是一个随机变量, 其分布依赖参数. 设是一个统计量, 设是实数线的子集. 假设对于公式(9.1.3)的测试过程的形式是“如果则拒绝.” 则我们称是统计测试, 称是拒绝域. 定义9.1.4中的集合就是临界域.

**定义9.1.6 幂函数** 设是一个测试过程. 函数被称为测试的幂函数. 如果代表的临界域, 则幂函数为:

如果用统计测试和拒绝域来表示, 那么幂函数为

**定义9.1.7** 我们在做假设检验的时候会犯两种错误：第一，原假设是正确的，而你判断它为错误的；第二，原假设是错误的，而你判断它为正确的。我们分别称这两种错误为第一类错误(Type I error)和第二类错误(Type II error). 其中原假设是指零假设.

用幂函数表示, 如果, 是统计者犯第一类错误的概率. 相似地, 如果, 是犯第二类错误的概率. 如果我们要在多个验证中选择一个. 我们可能会选择一个错误概率最小的验证. 也就是我们要选择一个幂函数对于是较低的, 并且我们希望对于是较大的. 一般情况下, 这两个目标是相对的. 也就是说, 如果我们选择使得对于是较小, 通常对于同样也是较小的. 因此需要折中这两个目标之间的平衡.

在两个目标之间平衡的最流行的方法是选择一个0和1之间的数字使得

**定义9.1.8 等级/大小** 满足的测试被称为等级测试，并且我们称该测试的等级大小为. 除此之外, 测试的大小定义为:

**推论9.1.1** 测试是等级测试当且仅当它的大小至多为. 如果零假设是简单的, 也就是, 则的大小将会是.

定义9.1.9 p值 一般情况下, p值就是最小值使得我们在显著水平上根据观察数据能够拒绝零假设.

从前，你有个朋友，他每月开一辆小货车去市场里采购东西，然后再把所有这些东西转卖出去。月复一月，年复一年。

但是，因为市场行情变动，每一个月他卖东西得到的毛爷爷数目并不一样，差不多有个5000块钱左右吧。你呢能看到他每个月的收入统计。

时间长了，有一次，无聊的你准备戏耍一下这个无辜的小伙伴，你偷偷在他的小货车上装了一小瓶尿。既然他肯定会在集市上卖掉所有东西，那么这一小瓶尿他一定也会卖出去，只不过这瓶尿的价格可能也就是零吧。

这一次他赚了5400块，这可比以往的平均数5000块钱还多。那问题就来了：为什么这个月他赚的比平常多呢？

有两种情况：其一，你那瓶尿很值钱，他自然会多赚；其二，你那瓶尿根本不值钱，他多赚了只是因为这个月市场上行情不错。

那到底是哪种情况呢？这个问题的实质是，你那瓶尿到底值不值钱。如果这瓶尿根本不值钱，那他卖东西和以前卖东西的情况没什么两样，只不过是市场行情影响而已。于是，你翻了他以前卖东西的纪录，算了算所有他的个人月收入超过5400块的概率，大概是0.1。也就是说在你这瓶尿没有任何价值的情况下，只靠市场行情，他只有0.1的概率卖到5400块。

所以你得出的结论是，你的那瓶尿有价值，你这个结论的p值就是0.1。

＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝＝

故事讲完了，来分析一下。

为什么要用p值？换句话说，为什么要做推断统计？我想，p值到底是啥并不重要，我知道你做研究时多半只是负责看看p到底有没有比0.05小的。而这个问题，才是我最想告诉你的。

推断统计，之所以要“推断”，是因为我们没有办法正面验证某种情况。在这个故事中，你以前没干过偷偷把一瓶尿放在车上的事儿，他以前也没卖过你的尿，所以尽管你有的是他的销售数据，但那些旧账本没办法正面告诉你你的尿到底值不值钱。

这就是“没办法正面验证”，在这样的情况下，你就没办法了么？非也，我们可以换个角度考虑这个问题，假设“你那瓶尿毫无价值”是成立的，他卖了这瓶尿跟没卖时候没什么两样，那这一次和以往自然也没什么两样。既然这次卖东西和以往一样，那旧账本中的记录就能帮到你了。你可以算算旧账本，他在没卖过你那瓶尿的情况下，赚到5400块及以上的可能性（概率），这个概率就是“你那瓶尿毫无价值”的概率，这里也就是0.1.

这样，“你那瓶尿有价值”的概率，当然就是这个假设的相反情况，也就是0.9，这也是你的假设成立的可能性。

那这个p值到底显著与否呢？那得看市场行情。这里市场行情是随机的，所以“显著”与否简单点就是在说你那瓶尿能不能跑赢市场。业内经常以两个标准差作为衡量“显著”的标准。绝大多数情况下，作为随机变量的市场行情服从正态分布。而正态分布中，超过两个标准差的概率是0.05，这也就是为什么大家要拿0.05作为“公认”的显著性水平尺度了。

统计推断，核心就是反证法。你那瓶尿没价值的可能性越小，反而越能证明你那瓶尿有价值。

“在原假设成立的情况下抽到的统计量与原假设之间的距离至少等于样本计算值与原假设之间的距离”

这是你的书上写的定义，但是很明显，你把结尾最重要的“的概率”三个字漏掉了。正常情况下，这个定义应该是

“在原假设成立的情况下抽到的统计量与原假设之间的距离至少等于样本计算值与原假设之间的距离的概率”

不信你可以再看看你的书。

但是，怎么可以容忍这么反人类的定义？我们来用这个故事做个转换吧：

“在原假设成立的情况下” －> 在你那瓶尿不值钱情况下

“抽到的统计量与原假设之间的距离” -> 他旧账本里的销售记录

“至少等于” －> 大于或等于

"样本计算值与原假设之间的距离" －> 他这一次的销售记录（5400块）

“的概率” －> 的概率

连起来读读，p值的定义就变成：

“在你那瓶尿不值钱的情况下，他旧账本里的销售记录大于或等于他这一次的销售记录（5400块钱）（这个事件发生）的概率。”

这一次，好懂了点吗？

最后要说明，说得通俗易懂是要承担风险的，因为通俗很可能意味着不严谨，易懂很可能意味着不周全。以上有很多有失严谨之处，希望题主还要多多看书哇。

定理9.1.1 从测试中定义置信集 设是从参数(或参数向量)为的分布中选取的一个样本. 设是一个函数, 并且假设对每一个可能值, 存在一个水平假设测试

对于每一个可能值,定义

设. 则随机集合满足

对所有成立.

定义9.1.10 置信集 如果随即集合对于每一个满足, 对于我们称其为系数置信集.

定理9.1.2 从置信集中定义验证 设是从参数(或参数向量)为的分布中选取的一个样本. 设是一个函数, 并且令是的系数置信集. 对于每一个可能值, 从公式构造下列假设测试: 不拒绝当且仅当. 则是公式水平假设测试.