概率论与数理统计知识整理

注：这是作者在大二上学期的课，与线代合并成一门。概统部分仅用一个月学完，比较粗糙，很多内容没有讲，但是对非数学专业而言，这些知识已经够用。

第一章 概率论

一、概率基本概念

1、基本事件：一次随机试验中的可能结果。

2、样本空间Ω：所有基本事件的集合。

3、事件：包含一个或多个基本事件的集合（或者空集）。事件是样本空间的子集。

4、频率：N 次试验中事件 A 发生的次数为 n，则定义试验中 A 的频率 $F_N(\mathrm{A})=\dfrac{n}{N}$.

☆频率的性质

（1）非负性：$F_N(\mathrm{A})\geq 0$.（同时有 $F_N(\mathrm{A})\leq 1$）

（2）规范性：$F_N(\Omega)=\dfrac{N}{N}=1$.

（3）可加性：若 A, B⊆Ω且二者不同时发生，则 $F_N(\mathrm{A}+\mathrm{B})=F_N(\mathrm{A})+F_N(\mathrm{B})$.

（4）稳定性：必存在常数 f 使得 $\lim\limits_{N\to\infty}F_N(\mathrm{A})=f$.

5、样本点：样本空间中的元素。其分布分为离散与连续。

（1）离散：有限个或可列无限个。（可列：可以与自然数集合建立一一映射关系，例如全体偶数、全体有理数等）

（2）连续：必要条件是无限且不可列。（如全体实数）

（3）集合 A 的元素个数一般记为|A|或 card(A)。对事件而言相当于样本点个数。

二、古典概型

1、古典概型的特征：样本空间Ω元素个数有限；所有基本事件发生概率相同（等可能性）。

2、古典概型中事件 A 的概率 $P(\mathrm{A})=\dfrac{|\mathrm{A}|}{|\Omega|}$.

3、处理古典概型的常见方法

（1）最直接的方法：将Ω与 A 的所有元素列出，数个数。（写集合、画树状图；交并关系复杂时使用文氏图）

（2）排列组合，一切随缘；除此之外，暂时无可奉告。

三、几何概型

1、基本特征：样本点无限多（一般用坐标系中的一个区域表示）；等可能性——若 A, B⊆Ω且 $V(\mathrm{A})=V(\mathrm{B})$（见下面定义），则 $P(\mathrm{A})=P(\mathrm{B})$。

2、测度：A 的测度 $V(\mathrm{A})$（Volume）表示长度、面积、体积（包括更高维度的对应量）中与 A 维度相同的一个。例如，一维区间的测度为长度，二维区域的测度为面积，三维区域的测度为体积。

3、几何概型中事件 A 的概率 $P(\mathrm{A})=\dfrac{|\mathrm{A}|}{|\Omega|}=\dfrac{V(\mathrm{A})}{V(\Omega)}$.

4、几何概型问题的关键：将纯文字的题干转化为图像。需要熟悉解析几何知识，

并且学会用不等式表示相应的区域。

5、子集簇：样本空间中所有事件的集合（事件本身是集合，子集簇就是集合的集合），也叫σ域、事件域。记为$\mathcal{F} = \{A|A\subseteq\Omega\}$。

☆子集簇的性质：

（1）$\Omega\in\mathcal{F}$；

（2）任意 $A\in\mathcal{F}$，有 $\overline{A}=\Omega\backslash A\in\mathcal{F}$；

（3）任意 $A_1, A_2,...\in\mathcal{F}$（有限个或可列无限个），有 $\bigcup\limits_{i=1}^{\infty}A_i\in\mathcal{F}$。

四、事件的运算

1、$A\cap B$ 或 AB：表示 A 与 B 同时发生的事件。

若 $A\cap B=\varnothing$，则说明二者不会同时发生，称 A 与 B <u>互斥</u>。

2、$A\subseteq B$：A 包含于 B，指 A 发生时 B 一定发生。A 的样本点全部属于 B。（小的发生时大的一定发生，请勿记反）

若 $A\subseteq B$ 且 $B\subseteq A$，则 $A=B$。

3、$A\cup B$：表示 A 发生或 B 发生。（至少发生一个）

若 A 与 B 互斥，也记 $A\cup B$ 为 $A+B$。

若 $A\cup B=\Omega$，且 $A\cap B=\varnothing$，说明 A 与 B 不同时发生，也不会都不发生（必有一个发生），称 A 与 B 互逆，B 是 A 的对立事件，记为 $B=\overline{A}$。

4、A\B：$\{x|x\in A$ 且 $x\notin B\}$。称为 A 与 B 的差事件。表示 A 发生但 B 不发生。

5、事件运算律

（1）交换律：$A\cup B=B\cup A$（或者 $AB=BA$）。

（2）结合律：$(A\cup B)\cup C=A\cup(B\cup C)$（或$(AB)C=A(BC)$）。

（3）分配律：$(A\cup B)\cap C=(A\cap C)\cup(B\cap C)$。

（4）De Morgan 律：$\overline{A\cup B}=\overline{A}\cap\overline{B}, \overline{A\cap B}=\overline{A}\cup\overline{B}$。

e.g. 已知三个事件 A, B, C。

①A 与 B 发生，C 不发生：$AB\overline{C}$或$AB\setminus C$或$AB\setminus(ABC)$。

②三件事都发生：ABC。

③三件事至少发生一个：$A\cup B\cup C$ 或 $\overline{\overline{A}\,\overline{B}\,\overline{C}}$。（算概率一般用后一个）

④三件事有且只有两个发生：$\overline{A}\cup\overline{B}\cup\overline{C}$ 或 $\overline{A}BC+\overline{B}AC+\overline{C}AB$。

（注意∪何时能写成+，何时不能）

五、概率公理化定义与概率空间

1、概率的三要素：样本空间Ω，子集簇$\mathcal{F}$，$\mathcal{F}$上的概率映射 $P(A)$。

2、概率的性质

（1）非负性：任意 $A\in\mathcal{F}$，$P(A)\geq 0$.

（2）规范性：$P(\Omega)=1$.

（3）可列可加性：若 $A_1, A_2,...\in\mathcal{F}$ 且互斥，则 $P(\bigcup\limits_{i=1}^{\infty}A_i)=\sum\limits_{i=1}^{\infty}P(A_i)$.

以上三条是最基本的性质；广义上，任何满足这三条性质的函数 $P(A)$都可以叫做概率；称$(\Omega,\mathcal{F},P)$为概率空间。

（4）$P(\varnothing)=0$.

（5）有限事件可加性：若 $A_1, A_2,..., A_n\in\mathcal{F}$ 且互斥，$P(\bigcup\limits_{i=1}^{n}A_i)=\sum\limits_{i=1}^{n}P(A_i)$.

（6）互逆事件：$P(\mathrm{A})+P(\overline{\mathrm{A}})=1.$
（7）若 B⊆A，则 $P(\mathrm{A}\backslash\mathrm{B}) = P(\mathrm{A}) - P(\mathrm{B}).$
（8）$P(\mathrm{A}\cup\mathrm{B}) = P(\mathrm{A}) + P(\mathrm{B}) - P(\mathrm{AB}).$
（9）一般情况下，$P(\mathrm{A}\backslash\mathrm{B}) = P(\mathrm{A}) - P(\mathrm{AB}).$
结合公式 7，若 B⊆A，则 $P(\mathrm{AB}) = P(\mathrm{B}).$
（10）容斥原理（多还少补定理）

$$P(\bigcup_{i=1}^{n}\mathrm{A}_i)=\sum_{i=1}^{n}P(\mathrm{A}_i)-\sum_{1\le i<j\le n}P(\mathrm{A}_i\mathrm{A}_j)+\sum_{1\le i<j<k\le n}P(\mathrm{A}_i\mathrm{A}_j\mathrm{A}_k)-...+(-1)^{n+1}P(\mathrm{A}_1\mathrm{A}_2...\mathrm{A}_n).$$

①内涵：n 个集合并集元素个数的计算。把每个集合的元素加起来，那么任意两个集合的交集都被算了两次，所以要减掉；减的过程中，任意三个集合的交集都被减了两次，所以要加回来；以此类推……

②转化为不等式：$P(\bigcup_{i=1}^{n}\mathrm{A}_i)\le\sum_{i=1}^{n}P(\mathrm{A}_i).$（次可加性）

（11）连续性：

若 $\mathrm{A}_1\supseteq\mathrm{A}_2\supseteq...$，记 $\lim_{n\to\infty}\mathrm{A}_n=\bigcap_{i=1}^{\infty}\mathrm{A}_i$，则 $\lim_{n\to\infty}P(\mathrm{A}_n)=P(\lim_{n\to\infty}\mathrm{A}_n).$

若 $\mathrm{A}_1\subseteq\mathrm{A}_2\subseteq...$，记 $\lim_{n\to\infty}\mathrm{A}_n=\bigcup_{i=1}^{\infty}\mathrm{A}_i$，则 $\lim_{n\to\infty}P(\mathrm{A}_n)=P(\lim_{n\to\infty}\mathrm{A}_n).$

（在事件序列层层包含时，事件概率的极限等于事件极限的概率）

六、条件概率与两个重要概率公式
1、条件概率的定义：在已知 B 发生的情况下 A 发生的概率被称为 A 关于 B 的条件概率，记为 $P(\mathrm{A}|\mathrm{B})$。

2、条件概率公式：若 $P(\mathrm{B})\neq0$，则 $P(\mathrm{A}\,|\,\mathrm{B})=\dfrac{P(\mathrm{AB})}{P(\mathrm{B})}.$

3、理解条件概率的意义：条件概率 $P(\mathrm{A}|\mathrm{B})$与 $P(\mathrm{AB})$是不一样的，虽然看上去都是 A 与 B 同时发生。在考察交集事件 AB 时，我们不默认 A 或者 B 一定会发生；在考察条件概率 $P(\mathrm{A}|\mathrm{B})$时，却是默认 B 一定会发生。这为认知者带来更多信息，因此事件 A 的元素个数有所减少，某些情况被排除，从而 $P(\mathrm{A}|\mathrm{B})\geqslant P(\mathrm{AB})$。我们对 A 的把握会更加准确。（这也说明“概率”本身依赖于我们知道的信息与理解句子的方式）
e.g. 家中有两个孩子。
（1）求一男（m）一女（f）的概率：
Ω = {(m,m),(m,f),(f,m),(f,f)}，A = {(m,f),(f,m)}，故 $P(\mathrm{A}) = 1/2.$
（不添加任何已知信息，默认什么都不知道）
（2）已知有一个是女孩，求一男一女的概率。
【法一】直接修改样本空间。Ω’ = {(m,f),(f,m),(f,f)}，A = {(m,f),(f,m)}，故 $P(\mathrm{A}|\Omega')$ = 2/3.（Ω中(m,m)被排除，分母减小）
【法二】套用条件概率公式。$P(\mathrm{A}|\Omega') = P(\mathrm{A}\Omega')/P(\Omega') = (1/2)/(3/4) = 2/3.$
4、条件概率的改写：$P(\mathrm{AB}) = P(\mathrm{B})P(\mathrm{A}|\mathrm{B}) = P(\mathrm{A})P(\mathrm{B}|\mathrm{A}).$
推广到 n 个事件——多事件乘法公式：
$P(\mathrm{A}_1\mathrm{A}_2...\mathrm{A}_n) = P(\mathrm{A}_1)P(\mathrm{A}_2|\mathrm{A}_1)P(\mathrm{A}_3|\mathrm{A}_1\mathrm{A}_2)...P(\mathrm{A}_n|\mathrm{A}_1\mathrm{A}_2...\mathrm{A}_{n-1}).$

内涵：$a_1a_2...a_n = a_1\dfrac{a_1a_2}{a_1}\dfrac{a_1a_2a_3}{a_1a_2}...\dfrac{a_1a_2...a_n}{a_1a_2...a_{n-1}}$.（只是形式上类似，上述概率不一定可以分解成乘积！）

4、全概率公式

若事件 $A_1, A_2,...$（有限或可列无限）与 B 满足

（1）$A_1, A_2,...$互斥，且概率均大于 0。

（2）$B\subseteq\sum_{i=1}^{\infty}A_i$ 。

则 $P(B)=\sum_{i=1}^{\infty}P(A_i)P(B\mid A_i)$.

原理：$B=B\cap(A_1\cup A_2\cup...)=(BA_1)\cup(BA_2)\cup...=\sum_{i=1}^{\infty}BA_i$.（最初的等号是因为第二个条件）

*简化形式：$P(B)=P(A)P(B\mid A)+P(\overline{A})P(B\mid\overline{A})$.（B 一定包含于 $A\cup\overline{A}$ ）

e.g. 决策问题：排异使得当前选择胜算下降

已知 n（$n\geqslant2$）个袋子 $A_1, A_2,..., A_n$ 中只有一个含有奖品。抽奖者选中 A_1，随后得知 $A_2, A_3,..., A_{n-1}$ 都没有奖品。他获得一次机会，可以决定是否换选 A_n。

设一开始猜 A_1 正好猜中为事件 A，最终获得奖品为事件 B。

对于事件 A：$P(A)=\dfrac{1}{n}$，$P(\overline{A})=1-\dfrac{1}{n}$.

如果得知信息后决定换：$P(B\mid A)=0$，$P(B\mid\overline{A})=1$.（若 A_1 本身是对的，换了就没有奖；否则，换了反而没有奖。）

如果不换：$P(B\mid A)=1$，$P(B\mid\overline{A})=0$.（理由与上面类似）

如果换，$P(B)=P(A)P(B\mid A)+P(\overline{A})P(B\mid\overline{A})=1-\dfrac{1}{n}$.

如果不换，$P(B)=P(A)P(B\mid A)+P(\overline{A})P(B\mid\overline{A})=\dfrac{1}{n}$.

只要 $n\geqslant2$，则换的胜算远大于不换。

【理解 1】最开始，奖品在每个袋子中的概率都是 $1/n$；得知信息后，后 $n-1$ 个袋子的可能性全部集中到 A_n，因此选 A_n 胜算更大；不能认为知道信息之后中奖的概率是 1/2，因为选 A_1 在先。只有先知道信息再选 A_1 或 A_n 才是 1/2 的概率。

【理解 2】对于事件 A，猜不中的可能性比猜得中要大。这时，只剩两个选项，猜 A_1 猜不中就等价于选 A_n 是对的，因此 A_n 对的可能性比 A_1 大，所以当然会换成 A_n。同样，在先排除 $A_2, A_3,..., A_{n-1}$ 后再选 A_1 或 A_n，这两个选项才是平等的，这和一开始就只给两个选项是一样的。

e.g. 抽签问题：n 人抽 n 张签，只有一张有奖。

（1）不知道任何信息时，得奖概率和抽签顺序无关。因为第 k 人得奖的概率为

$$P(\overline{A}_1...\overline{A}_{k-1}A_k)=P(\overline{A}_1)P(\overline{A}_2\mid\overline{A}_1)P(\overline{A}_3\mid\overline{A}_1\overline{A}_2)...P(A_k\mid\overline{A}_1...\overline{A}_{k-1})$$

$=\dfrac{n-1}{n}\dfrac{n-2}{n-1}\dfrac{n-3}{n-2}...\dfrac{1}{n-k+1}=\dfrac{1}{n}$，与 k 无关。

（2）在已经知道前面 $k-1$ 人未得奖时，第 k 人得奖的概率为

$P(A_k\mid\overline{A}_1...\overline{A}_{k-1})=\dfrac{1}{n-k+1}$.这时越后抽奖越有利。

5、贝叶斯公式：在全概率公式的两个条件下，有

$$P(\mathrm{A}_i\,|\,\mathrm{B})=\frac{P(\mathrm{A}_i)P(\mathrm{B}\,|\,\mathrm{A}_i)}{P(\mathrm{B})}=\frac{P(\mathrm{A}_i)P(\mathrm{B}\,|\,\mathrm{A}_i)}{\sum_{k=1}^{\infty}P(\mathrm{A}_k)P(\mathrm{B}\,|\,\mathrm{A}_k)}.$$

（1）内涵：全概率公式是由各个因素贡献之和求出结果，贝叶斯公式是由结果回推各个因素的贡献。（二者等号左边待求的事件不一样）

（2）由贝叶斯公式得出条件概率的关系：$P(\mathrm{A})P(\mathrm{B}\,|\,\mathrm{A})=P(\mathrm{B})P(\mathrm{A}\,|\,\mathrm{B})$.（事实上这是显然的，两者都等于 $P(\mathrm{AB})$）

6、事件的独立性

（1）统计独立：若两个事件是否发生不会互相影响，则称二者（统计）独立。

（2）独立事件 A、B 的性质

① $P(\mathrm{A}\,|\,\mathrm{B})=P(\mathrm{A})$.（B 发不发生，不影响 A 的概率）

② $P(\mathrm{AB})=P(\mathrm{A})P(\mathrm{B})$.

以上两条是事件独立的**充要条件**，可用于判断两个事件是否独立。

③任何事件均与$\varnothing$ 独立，也必然与样本空间Ω独立。

（3）多个事件的独立：n 个事件 $\mathrm{A}_1, \mathrm{A}_2,..., \mathrm{A}_n$，等价于对一切可能组合 $1\leqslant i<j<k<...<n$，有
$$\begin{cases}P(\mathrm{A}_i\mathrm{A}_j)=P(\mathrm{A}_i)P(\mathrm{A}_j)\\P(\mathrm{A}_i\mathrm{A}_j\mathrm{A}_k)=P(\mathrm{A}_i)P(\mathrm{A}_j)P(\mathrm{A}_k)\\...\\P(\mathrm{A}_1\mathrm{A}_2...\mathrm{A}_n)=P(\mathrm{A}_1)P(\mathrm{A}_2)...P(\mathrm{A}_n)\end{cases}.$$

（4）非独立的一个充分条件：若事件 A、B 互斥，且 $P(\mathrm{A})P(\mathrm{B})\neq 0$，则 A、B 不独立。

【理解 1】代数上，$P(\mathrm{AB})=0\neq P(\mathrm{A})P(\mathrm{B})$，因此不满足独立的条件。

【理解 2】本质意义上，两个互斥事件，一个发生必然导致另一个不发生，这便是两件事之间的影响；如果两件事发生的概率都不是 0，那么二者就有互相影响的可能性，所以不独立。

（5）若 A 与 B 独立，则 A与$\overline{\mathrm{B}}$，$\overline{\mathrm{A}}$与B，$\overline{\mathrm{A}}$与$\overline{\mathrm{B}}$ 也独立。

推广的结论：以上四组中，有任意一组独立，则另外三组也独立。

七、一维随机变量

（一）一维离散型随机变量

1、离散型随机变量的定义：取值集合的元素个数是有限个或可列无限个。

2、离散型随机变量ξ的分布常用概率分布表来表示，第一行为ξ的可能取值，第二行为各个取值对应的概率值。

☆☆☆概率分布表最基本的特征：归一性。所有概率值之和为 1，这是求解参数时几乎不得不使用但是很容易被忘记的一个结论！

3、Bernoulli 试验：多次重复、独立的试验，每次试验只有两种互斥的结果。例如抛硬币。若干次 Bernoulli 试验对应的模型称为 Bernoulli 概型。0−1 分布、二项分布、几何分布均与 Bernoulli 试验有关。

4、常见的一维离散型分布

（1）0−1 分布：一次 Bernoulli 试验中，记成功为$\xi=1$，失败为$\xi=0$。一般记成功的概率为 p，失败的概率为 $q=1-p$。

（2）二项分布：Bernoulli 试验总次数为 n，记其中成功的次数为ξ，则ξ满足二项分布。对于 0≤k≤n，有 $P(\xi=k)=C_n^k p^k q^{n-k}$. p 和 q 的定义同上。这一关系记为$\xi\sim \mathrm{B}(n,p)$，而 $P(\xi=k)$记为$b(k;n,p)$.

（3）几何分布：Bernoulli 试验总次数为 n，记第一次成功时的试验次数为ξ，则ξ满足几何分布。

①对于 1≤k≤n，有 $P(\xi=k)=pq^{k-1}$.

②无记忆性：若从第一次试验开始ξ服从几何分布，则对于任意正整数 m，若前 m 次试验均失败，则从第 m 次试验开始计算，第一次成功时的试验次数$\eta=\xi-m$ 也服从几何分布。

$$P(\eta=k\mid 前m次失败)=\frac{P(\eta=k且前m次失败)}{P(前m次失败)}=\frac{pq^{m+k-1}}{q^m}=pq^{k-1}=P(\xi=k).$$

（实例：多次抽奖，前面抽不到不会使后面中奖的概率变大。）

（4）Poisson 分布

①定义：对于取值集合为自然数集的随机变量ξ，若 $P(\xi=k)=\dfrac{\lambda^k}{k!}e^{-\lambda}$，则称$\xi$服从 Poisson 分布，记作$\xi\sim \mathrm{P}(\lambda)$。

②Poisson 定理：在某些特殊的 Bernoulli 试验中，每次成功的概率与次数 n 有关，记作 p_n；若 $\lim\limits_{n\to\infty} np_n=\lambda$，则 $\lim\limits_{n\to\infty}P(\xi=k)=\lim\limits_{n\to\infty}C_n^k p_n^k(1-p_n)^{n-k}=\dfrac{\lambda^k}{k!}e^{-\lambda}$.（二项分布在试验次数充分大时近似为 Poisson 分布）

③*满足以下条件的变量服从 Poisson 分布（也是 Poisson 分布的性质）：

（i）平稳性：落在区间$[t_0,t_0+\Delta t]$内的次数与Δt 有关，与 t_0 无关。

（ii）独立增量性：落在区间$[t_0,t_0+\Delta t]$内的次数与 t_0 之前的事件无关。

（iii）普通性：在任意无限小区间内至多只有一次呼叫（不会出现奇点）。

（二）一维连续型随机变量

1、分布函数

（1）对于随机变量ξ，定义其分布函数 $F_\xi(x)=P(\xi\le x)$.（实质：从负无穷大开始的概率值的累积。**这个函数对离散和连续变量都可以定义。**）

（2）分布函数的性质

注：离散变量的分布函数多数是不连续的（阶梯状），要注意书写函数解析式时分段点的函数值。与此同时，分布函数都是定义在$(-\infty,+\infty)$上的，在ξ所能取的最小以下，函数值为 0；在ξ所能取的最大值以上，函数值为 1。

① $P(\xi<a)=\lim\limits_{b\to a^-}P(\xi\le b)=F(a^-)$.

② $P(\xi=a)=P(\xi\le a)-P(\xi<a)=F(a)-F(a^-)$.

③ $P(\xi>a)=1-P(\xi\le a)=1-F(a)$.

④ $P(a<\xi<b)=P(\xi<b)-P(\xi\le a)=F(b^-)-F(a)$.

⑤单调性：若 $a<b$，则 $F(a)\leqslant F(b)$.

⑥规范性：$\lim\limits_{x\to-\infty}F(x)=0,\lim\limits_{x\to+\infty}F(x)=1$.（$x\leqslant-\infty$不可能发生，$x\leqslant+\infty$一定发生）

☆☆☆这是求解参数时几乎不得不使用但是很容易被忘记的一个结论！

⑦右连续性：$F(x^+)=F(x)$.

对比：若定义 $G_\xi(x)=P(\xi<x)$，则它应当是左连续的。

2、概率密度函数（<u>仅对连续型变量定义</u>）

（1）连续变量不存在点概率：连续变量相当于有无限不可列个取值。因此，如果要计算$P(\xi=a)$，它总是等于有限的事件样本除以无限的基本事件个数，总是趋于 0（具体体现为分布函数的连续性，即$P(\xi=a)=F(a)-F(a^-)=0$）。在连续变量的讨论中，概率为 0 不一定是不可能事件，概率为 1 不一定是必然事件。因此，不能用概率为 0 或者 1 来定义不可能与必然。

*典型例子：在一条线段 AB 上等可能地取点。若 C 在线段 AB 上，可以计算点落在 AC 段的概率为$\frac{|\mathrm{AC}|}{|\mathrm{AB}|}$（是几何概型），但是由于 C 点本身没有长度，因此这个点落在 C 上的概率为$\frac{0}{|\mathrm{AB}|}=0$，落在 C 点以外位置的概率为$\frac{|\mathrm{AC}|+|\mathrm{CB}|}{|\mathrm{AB}|}$=1.

但是这个点显然不是不可能落在 C 点，也不是必定落在 C 以外的位置。

<u>**结论：连续型随机变量不能够定义取值在某个点的概率，只能定义在一段区间内取值的概率。为了体现出不依赖于区间长度的性质，定义单位区间长度上的概率，即为概率密度。**</u>

（2）概率密度的规范定义：若随机变量ξ可以取得在某个区间内的所有实数，并且存在<u>**非负可积函数**</u>$\varphi(x)$使得$F_\xi(x)=\int_{-\infty}^{x}\varphi(t)\mathrm{d}t$，则称$\varphi(x)$为$\xi$的概率密度函数。

①离散变量的分布函数总是存在不可导、不连续点，因此不能定义概率密度；连续变量与离散变量的区别就好像刚体与质点组的区别，质点的质量全部集中在一个点，是不可能对它谈论密度的。

②只有存在概率密度的函数才是连续性随机变量，这是它的一个判定依据。

（3）密度函数的性质

①$P(a<\xi\le b)=F(b)-F(a)=\int_a^b\varphi(x)\mathrm{d}x.$（质量＝密度×体积，概率＝概率密度×区间长度。如果不是均匀分布，上述式子改写成积分。）

②连续性：$P(\xi=a)=F(a)-F(a^-)=\lim\limits_{\varepsilon\to0^+}\int_{a-\varepsilon}^{a}\varphi(x)\mathrm{d}x=0.$

③对一般点（不考虑可能出现的端点），$\varphi(x)=\dfrac{\mathrm{d}F(x)}{\mathrm{d}x}.$

④非负性：$\varphi(x)\geqslant0$.

⑤归一性：$\int_{-\infty}^{+\infty}\varphi(x)\mathrm{d}x=1.$

☆☆☆概率密度最基本的特征：归一性。这是求解参数时几乎不得不使用但是很容易被忘记的一个结论！

（任意满足$\int_{-\infty}^{+\infty}\varphi(x)\mathrm{d}x=1$的非负函数$\varphi(x)$均可视为某个变量的密度函数。）

3、常见的连续分布

（1）区间$[a,b]$上的均匀分布 $\mathrm{U}[a,b]$：

$$\varphi(x)=\begin{cases}\dfrac{1}{b-a}，a\le x<b,\\0，其他.\end{cases}\qquad F(x)=\begin{cases}0，x<a,\\\dfrac{x-a}{b-a}，a\le x<b,\\1，x\ge b.\end{cases}$$

（只在区间$[a,b]$上有分布，各个地方的取值可能性一样。相当于前述的线段取点对应的几何概型。）

（2）指数分布 E(λ)：

$$\varphi(x)=\begin{cases}\lambda e^{-\lambda x}，x\ge 0,\\ 0，x<0.\end{cases}\qquad F(x)=\begin{cases}1-e^{-\lambda x}，x\ge 0,\\ 0，x\le 0.\end{cases}$$

指数分布同样有无记忆性：$P(\xi>s+t\mid\xi>s)=P(\xi>t)$.

（3）☆正态分布 N(μ,σ^2)：注意括号里面第二个数是σ^2而不是σ。（σ＞0）

$$\varphi(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}.\qquad F(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(t-\mu)^2}{2\sigma^2}}\mathrm{d}t.$$

①熟记积分：$\int_{-\infty}^{+\infty}e^{-x^2}\mathrm{d}x=\sqrt{\pi}$.（可以用二重积分或者$\Gamma$函数证明）

②正态分布的标准化：若$\xi\sim$ N(μ,σ^2)，则$\dfrac{\xi-\mu}{\sigma}\sim$ N(0,1). 称 N(0, 1)为标准正态分布，所有正态分布都可以归结为标准正态分布。

③若$\xi\sim$ N(μ,σ^2)，则 $a\xi+b\sim$ N($a\mu+b, a^2\sigma^2$)

④正态分布的分布函数一般记作$\Phi_\xi(x)$，标准正态分布记为$\Phi_0(x)$。有以下性质：

（i）$\Phi_\xi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(t-\mu)^2}{2\sigma^2}}\mathrm{d}t\overset{u=\frac{t-\mu}{\sigma}}{=}\int_{-\infty}^{\frac{x-\mu}{\sigma}}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\mathrm{d}u=\Phi_0(u)=\Phi_0(\frac{x-\mu}{\sigma})$.

（ii）$\Phi_0(x)+\Phi_0(-x)=1$.

（iii）$P(a<\xi\le b)=\Phi_\xi(b)-\Phi_\xi(a)=\Phi_0(\frac{b-\mu}{\sigma})-\Phi_0(\frac{a-\mu}{\sigma})$.

（iv）统计学 3σ原则：$P(|\xi-\mu|<3\sigma)=\Phi_0(3)-\Phi_0(-3)=0.9973$. 正态分布的变量几乎全都落在区间($\mu-3\sigma,\mu+3\sigma$)中。可以用于检验$\mu$和$\sigma$测量值是否准确。

八、二维随机变量

1、高维随机变量的定义：在样本空间Ω中定义 n 个一维随机变量$\xi_1,\xi_2,\dots,\xi_n$，它们可能独立也可能有牵连。它们构成的 n 维向量($\xi_1,\xi_2,\dots,\xi_n$)称为 n 维随机变量。

2、二维随机变量(ξ,η)联合分布的表示

（1）离散型随机变量：联合分布表。

（2）离散/连续型：联合分布函数。$F_{\xi,\eta}(x,y)=P(\xi\le x,\eta\le y)$.

① $P(a_1\le\xi\le b_1,a_2\le\eta\le b_2)=F(a_1,a_2)+F(b_1,b_2)-F(a_1,b_2)-F(b_1,a_2)$.（($\xi,\eta$)落在矩形区域的概率是右斜对角线减去左斜对角线）

推论：$F(a_1,a_2)+F(b_1,b_2)-F(a_1,b_2)-F(b_1,a_2)\ge 0$.（$a_1<b_1$，$a_2<b_2$）

②值域：0≤$F_{\xi,\eta}(x,y)$≤1.

③$F_{\xi,\eta}(x,y)$对 x 与 y 均单调递增且右连续。

④任意实数 x 与 y，$F(x,-\infty)=F(-\infty,y)=0$，$F(+\infty,+\infty)=1$.

☆☆☆这是求解参数时几乎不得不使用但是很容易被忘记的结论！

（3）连续型：联合密度函数。（二维变量落在一个点处的概率同样趋于 0，只能定义落在一个二维区域内的概率。为了体现不依赖于区域面积的性质，定义单位面积上的概率，即为二维概率密度。）

①定义：若存在非负可积二元函数$\varphi(x,y)$使得$F_{\xi,\eta}(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}\varphi(s,t)\mathrm{d}s\mathrm{d}t$，则称$\varphi(x,y)$为($\xi,\eta$)的联合概率密度函数。（这也是判定二维连续变量的一个依据）

②$\frac{\partial F}{\partial x}=\int_{-\infty}^{y}\varphi(x,t)\mathrm{d}t$，$\frac{\partial F}{\partial y}=\int_{-\infty}^{x}\varphi(s,y)\mathrm{d}s$；$\frac{\partial^2 F}{\partial x\partial y}=\varphi(x,y)$.

3、边缘分布

（1）定义：由二维（或更高维）联合分布生成的一维变量的分布律。

（2）离散型(ξ,η)的边缘分布：ξ,η的取值均是有限个或可列无限个。

$$P(\xi=x_i)=\sum_{j=1}^{\infty}P(\xi=x_i,\eta=y_j)=\sum_{j=1}^{\infty}p_{ij}\equiv p_{i*}.\ (i=1,2,...)$$

$$P(\eta=y_j)=\sum_{i=1}^{\infty}P(\xi=x_i,\eta=y_j)=\sum_{i=1}^{\infty}p_{ij}\equiv p_{*j}.\ (j=1,2,...)$$

（3）连续型(ξ,η)的边缘分布：

$$\varphi_\xi(x)=\int_{-\infty}^{+\infty}\varphi(x,t)\mathrm{d}t=\left.\frac{\partial F}{\partial x}\right|_{y=+\infty}，\varphi_\eta(y)=\int_{-\infty}^{+\infty}\varphi(s,y)\mathrm{d}s=\left.\frac{\partial F}{\partial y}\right|_{x=+\infty}.$$

（4）用分布函数得到边缘分布（离散/连续）：

$F_\xi(x)=F_{\xi,\eta}(x,+\infty)$，$F_\eta(y)=F_{\xi,\eta}(+\infty,y)$.

（5）边缘分布推导过程的实质：用全概率公式消元。

4、常见二维连续分布

（1）二维区域 D（面积为 S_D）上的均匀分布：$\varphi(x,y)=\begin{cases}\frac{1}{S_D}，(x,y)\in D\\ 0，\text{其他}\end{cases}$

（2）二维正态分布（函数解析式不必记住）：$\mathrm{N}(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$

$$\varphi(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}]}.$$

①边缘分布：不论ρ取值如何，均有$\xi\sim\mathrm{N}(\mu_1,\sigma_1^2)$，$\eta\sim\mathrm{N}(\mu_2,\sigma_2^2)$。

②对于 n 维随机变量$\boldsymbol{\xi}=\begin{pmatrix}\xi_1\\ ...\\ \xi_n\end{pmatrix}$，设其对应自变量$\boldsymbol{x}=\begin{pmatrix}x_1\\ ...\\ x_n\end{pmatrix}$，$\boldsymbol{\mu}=\begin{pmatrix}\mu_1\\ ...\\ \mu_n\end{pmatrix}$，则有 n 维正态分布函数$\varphi(\boldsymbol{x})=\frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|\boldsymbol{B}|}}e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$.

其中协方差矩阵$\boldsymbol{B}=\begin{pmatrix}\sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & ... & \rho_{1n}\sigma_1\sigma_n\\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & & \\ ... & & ... & \\ \rho_{1n}\sigma_1\sigma_n & & & \sigma_n^2\end{pmatrix}$.

③若$\xi_i\sim\mathrm{N}(\mu_i,\sigma_i^2)(1\le i\le n)$，则$\sum_{i=1}^{n}a_i\xi_i\sim\mathrm{N}(\sum_{i=1}^{n}a_i\mu_i,\sum_{i=1}^{n}a_i^2\sigma_i^2)$.

九、随机变量的独立性与相关性

1、ξ与η统计独立$\Leftrightarrow F_{\xi,\eta}(x,y)=F_\xi(x)F_\eta(y)$.

如果是连续变量：等价于$\varphi_{\xi,\eta}(x,y)=\varphi_\xi(x)\varphi_\eta(y)$.

如果是离散变量：等价于 $p_{ij}=p_{i*}p_{*j}$.

2、二维正态分布 $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$中，$\xi$与$\eta$统计独立$\Leftrightarrow \rho=0.$

十、随机变量函数

1、已知ξ是随机变量，则ξ的函数$\eta=f(\xi)$也是随机变量。

2、离散随机变量函数的概率分布：将η的值和ξ的值对应起来，得到各个事件的概率。

3、连续性随机变量函数的概率分布：$\varphi_\eta(z)=\varphi_\xi(x(z))\left|\dfrac{\mathrm{d}x}{\mathrm{d}z}\right|.$

【理解 1】若$f(\xi)$随ξ单调递增，则$\eta\le z\Leftrightarrow \xi\le x(z).$ 故$F_\eta(z)=F_\xi(x(z)).$

从而$\varphi_\eta(z)=\dfrac{\mathrm{d}F_\eta(z)}{\mathrm{d}z}=\dfrac{\mathrm{d}F_\xi(x(z))}{\mathrm{d}z}=\dfrac{\mathrm{d}F_\xi(x(z))}{\mathrm{d}x}\dfrac{\mathrm{d}x}{\mathrm{d}z}=\varphi_\xi(x(z))\left|\dfrac{\mathrm{d}x}{\mathrm{d}z}\right|.$（复合函数求导）

若$f(\xi)$随ξ单调递减，则$\eta\le z\Leftrightarrow \xi\ge x(z).$ 故$F_\eta(z)=1-F_\xi(x(z)).$

从而$\varphi_\eta(z)=-\dfrac{\mathrm{d}F_\eta(z)}{\mathrm{d}z}=-\dfrac{\mathrm{d}F_\xi(x(z))}{\mathrm{d}z}=-\dfrac{\mathrm{d}F_\xi(x(z))}{\mathrm{d}x}\dfrac{\mathrm{d}x}{\mathrm{d}z}=\varphi_\xi(x(z))\left|\dfrac{\mathrm{d}x}{\mathrm{d}z}\right|.$

【理解 2】当ξ落在小区间$(x,x+\Delta x)$时，η落在一个对应的小区间$(z,z+\Delta z)$，这两个事件等价，从而概率相同：$\varphi_\eta(z)\Delta z=\varphi_\xi(x)\Delta x.$ 因此

$\varphi_\eta(z)=\varphi_\xi(x(z))\dfrac{\Delta x}{\Delta z}=\varphi_\xi(x(z))\left|\dfrac{\mathrm{d}x}{\mathrm{d}z}\right|.$

（1）若$\eta=a\xi+b$，则$\varphi_\eta(x)=\dfrac{1}{|a|}\varphi_\xi\left(\dfrac{x-b}{a}\right).$

（2）若二维随机变量(ξ,η)的联合概率密度函数为$\varphi(x,y)$，$\zeta=\xi+\eta$，则

$$F_\zeta(z)=P(\xi+\eta\le z)=\iint\limits_{x+y\le z}\varphi(x,y)\mathrm{d}x\mathrm{d}y=\int_{-\infty}^{+\infty}\mathrm{d}x\int_{-\infty}^{z-x}\varphi(x,y)\mathrm{d}y$$

$$=\int_{-\infty}^{+\infty}\mathrm{d}x\int_{-\infty}^{z}\varphi(x,t-x)\mathrm{d}(t-x)=\int_{-\infty}^{z}\mathrm{d}t\int_{-\infty}^{+\infty}\varphi(x,t-x)\mathrm{d}x.$$

因此$\varphi_\zeta(z)=\dfrac{\mathrm{d}}{\mathrm{d}z}\displaystyle\int_{-\infty}^{z}\mathrm{d}t\int_{-\infty}^{+\infty}\varphi(x,t-x)\mathrm{d}x=\int_{-\infty}^{+\infty}\varphi(x,z-x)\mathrm{d}x.$

（3）最值函数：若ξ与η统计独立，$\zeta=\max\{\xi,\eta\}$，$\kappa=\min\{\xi,\eta\}$，则

$F_\zeta(z)=P(\xi\le z,\eta\le z)=F_\xi(z)F_\eta(z).$

$F_\kappa(z)=1-P(\xi>z,\eta>z)=1-[1-F_\xi(z)][1-F_\eta(z)].$

然后可以求得相应的密度函数。

十一、随机变量的数字特征

1、数学期望的概念：数学期望 $E\xi$是随机变量ξ在概率意义下的均值，即多次试验中，ξ各个取值的频率完全和概率一致时对应的平均值。

2、数学期望的计算

（1）单个离散变量：$E\xi=\displaystyle\sum_{i=1}^{\infty}x_ip_i.$（有限/可列无限；要求这个级数绝对收敛，否则期望不存在）

（2）单个连续变量：$E\xi=\displaystyle\int_{-\infty}^{+\infty}x\varphi(x)\mathrm{d}x.$（要求此积分绝对收敛，否则期望不存在）

（3）离散的随机变量函数：$\mathrm{E}f(\xi)=\sum_{i=1}^{\infty}f(x_i)p_i.$

（4）连续的随机变量函数：$\mathrm{E}f(\xi)=\int_{-\infty}^{+\infty}f(x)\varphi(x)\mathrm{d}x.$

（5）数学期望的性质

①若$\xi=c$为常数，则$\mathrm{E}\xi=c$.

②线性：$\mathrm{E}\sum_{i=1}^{n}c_i\xi_i=\sum_{i=1}^{n}c_i\mathrm{E}\xi_i$.（$\mathrm{E}(a\xi+b)=a\mathrm{E}\xi+b.$）

③若$\xi_1,\xi_2,\ldots,\xi_n$**统计独立**且期望均存在，则$\mathrm{E}\prod_{i=1}^{n}\xi_i=\prod_{i=1}^{n}\mathrm{E}\xi_i$.

④$\mathrm{E}(\xi-\mathrm{E}\xi)=\mathrm{E}\xi-\mathrm{E}\xi=0$.

3、方差的概念：方差$\mathrm{D}\xi$（有时也写成$\mathrm{Var}\xi$，$\mathrm{V}\xi$）是随机变量ξ离散程度的体现。平方是为了让分布在平均值左边和右边的值对方差都是正的贡献，而不会抵消。

4、方差的计算

（1）定义式：$\mathrm{D}\xi=\mathrm{E}(\xi-\mathrm{E}\xi)^2$.（可以看出$\mathrm{D}\xi\geqslant 0$。定义$\sqrt{\mathrm{D}\xi}$为标准差。）

（2）与数学期望的关系（常用）：$\mathrm{D}\xi=\mathrm{E}(\xi^2)-(\mathrm{E}\xi)^2$.（平方的平均值未必等于平均值的平方，并且是$\mathrm{E}(\xi^2)\geqslant(\mathrm{E}\xi)^2$）

（3）离散变量：$\mathrm{D}\xi=\sum_{i=1}^{\infty}(x_i-\mathrm{E}\xi)^2p_i=\sum_{i=1}^{\infty}x_i^{\,2}p_i-(\sum_{i=1}^{\infty}x_ip_i)^2.$

（4）连续变量：$\mathrm{D}\xi=\int_{-\infty}^{+\infty}(x-\mathrm{E}\xi)^2\varphi(x)\mathrm{d}x=\int_{-\infty}^{+\infty}x^2\varphi(x)\mathrm{d}x-\left(\int_{-\infty}^{+\infty}x\varphi(x)\mathrm{d}x\right)^2.$

（5）方差的性质

①若$\xi=c$为常数，则$\mathrm{D}\xi=0$.

②$\mathrm{D}(a\xi+b)=a^2\mathrm{D}\xi$.

③若$\xi_1,\xi_2,\ldots,\xi_n$**统计独立**且方差均存在，则$\mathrm{D}\sum_{i=1}^{n}c_i\xi_i=\sum_{i=1}^{n}c_i^{\,2}\mathrm{D}\xi_i$.

④若$\xi_1,\xi_2,\ldots,\xi_n$**统计独立**且服从相同分布（如各次 Bernoulli 试验的结果），则有
$\mathrm{E}\bar{\xi}=\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)=\mathrm{E}\xi_1$；$\mathrm{D}\bar{\xi}=\mathrm{D}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)=\frac{1}{n}\mathrm{D}\xi_1$.

5、常见离散分布的期望与方差

分布律	概率公式	期望	方差
0−1 分布	$P(\xi=1)=p$ $P(\xi=0)=q$	p	pq
二项分布	$P(\xi=k)=C_n^kp^kq^{n-k}$	np	npq
Poisson 分布	$P(\xi=k)=\frac{\lambda^k}{k!}e^{-\lambda}$	λ	λ
几何分布	$P(\xi=k)=pq^{k-1}$	$\frac{1}{p}$	$\frac{q}{p^2}$

6、常见连续分布的期望与方差

分布律	期望	方差
均匀分布 U[a, b]	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
指数分布 E(λ)	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
正态分布 N(μ, σ^2)	μ	σ^2

7、协方差与相关系数

（1）协方差：$\mathrm{Cov}(\xi, \eta) = \mathrm{E}[(\xi - \mathrm{E}\xi)(\eta - \mathrm{E}\eta)]$.（注意：如果$\xi, \eta$不独立，它就不能拆成 $\mathrm{E}(\xi - \mathrm{E}\xi)\cdot\mathrm{E}(\eta - \mathrm{E}\eta)$，因此它不一定是 0）

（2）协方差矩阵：$\xi_1, \xi_2,..., \xi_n$ 的协方差矩阵为$(\mathrm{Cov}(\xi_i, \xi_j))_{n\times n}$.

（3）协方差的两个计算式：

①$\mathrm{Cov}(\xi, \eta) = \mathrm{E}(\xi\eta) - \mathrm{E}\xi\mathrm{E}\eta$.

②$\mathrm{Cov}(a\xi, b\eta) = ab\mathrm{Cov}(\xi, \eta)$.

（4）相关系数：$\rho_{\xi,\eta} = \dfrac{\mathrm{Cov}(\xi,\eta)}{\sqrt{\mathrm{D}\xi\cdot\mathrm{D}\eta}}$.

①取值范围：$-1\le\rho_{\xi,\eta}\le 1$.

②若 $\rho_{\xi,\eta}=0$，则称ξ, η不相关。

③独立是不相关的充分不必要条件。

这是因为：由 $\varphi_{\xi,\eta}(x,y)=\varphi_{\xi}(x)\varphi_{\eta}(y)$. 可以得出

$$\mathrm{E}(\xi\eta)=\iint\limits_{\mathbf{R}^2}xy\varphi_{\xi,\eta}(x,y)\mathrm{d}x\mathrm{d}y=\int_{-\infty}^{+\infty}x\varphi_{\xi}(x)\mathrm{d}x\int_{-\infty}^{+\infty}y\varphi_{\eta}(y)\mathrm{d}y=\mathrm{E}\xi\mathrm{E}\eta,$$

但是已知 $\iint\limits_{\mathbf{R}^2}xy\varphi_{\xi,\eta}(x,y)\mathrm{d}x\mathrm{d}y=\int_{-\infty}^{+\infty}x\varphi_{\xi}(x)\mathrm{d}x\int_{-\infty}^{+\infty}y\varphi_{\eta}(y)\mathrm{d}y=\iint\limits_{\mathbf{R}^2}xy\varphi_{\xi}(x)\varphi_{\eta}(y)\mathrm{d}x\mathrm{d}y$ 并不能直接得到 $\varphi_{\xi,\eta}(x,y)=\varphi_{\xi}(x)\varphi_{\eta}(y)$.（两个期望可能只是恰好相等，就好像由 $\int_a^b f(x)\mathrm{d}x=\int_a^b g(x)\mathrm{d}x$ 并不能得到 $f(x)=g(x)$。充分不必要性的来由就是：积分只能体现一个区域上的总体性质而不能体现每个点的性质，而每个点明确之后却可以得到总体性质）

④如果两个随机变量服从正态分布，那么相关与独立是等价的。

原因：$\mathrm{Cov}(\xi, \eta)=\rho\sigma_1\sigma_2$，$\rho_{\xi,\eta}=\rho$，故 ξ,η独立 $\Leftrightarrow \rho=0 \Leftrightarrow \rho_{\xi,\eta}=0 \Leftrightarrow \xi,\eta$不相关。

8、分位数：若 $F(x_\alpha)=P(\xi\le x_\alpha)=\alpha$，则称 x_α 为随机变量的下侧α分位数。同理可以定义上侧α分位数 x'_α：$1-F(x'_\alpha)=P(\xi\ge x'_\alpha)=\alpha$. 当$\alpha = 0.5$ 时，对应的分位数称为中位数。

十二、大数定理与中心极限定理（不是重点）

1、Bernoulli 大数定律：对于 n 次满足 0−1 分布的 Bernoulli 试验$\{\xi_1, \xi_2,..., \xi_n\}$，若每次成功概率为 p，则任意ε＞0，$\lim\limits_{n\to\infty}P\left(\left|\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i-p\right|\le\varepsilon\right)=1$. 平均值必然趋向于数学期望。

2、大数定律的一般形式：已知随机变量序列$\{\xi_n\}$，若存在实数ξ，使得任意$\varepsilon>0$，$\lim\limits_{n\to\infty}P\left(\left|\xi_n-\xi\right|\le\varepsilon\right)=1$，则称$\{\xi_n\}$服从大数定律，也称$\{\xi_n\}$依概率收敛于$\xi$，记作$\{\xi_n\}\xrightarrow{P}\xi$.

3、切比雪夫不等式：若随机变量ξ存在方差，则对任意$\varepsilon>0$，$P\left(\left|\xi-\mathrm{E}\xi\right|\ge\varepsilon\right)\le\dfrac{\mathrm{D}\xi}{\varepsilon^2}$.

意义：说明随机变量大部分还是分布在期望附近。由此可以将正态分布的3σ原则推广至任意分布——随机变量ξ大部分落在区间($\mathrm{E}\xi-3\sqrt{\mathrm{D}\xi},\mathrm{E}\xi+3\sqrt{\mathrm{D}\xi}$)中。

4、大数定律的三种表述

设$\{\xi_n\}$是彼此独立的随机变量序列，期望都存在。如果：

（1）（切比雪夫）存在公共的实数L使得$\mathrm{D}\xi_i\le L$；

（2）（马可夫）$\lim\limits_{n\to\infty}\dfrac{1}{n^2}\sum\limits_{i=1}^{n}\mathrm{D}\xi_i=0$；

（3）（辛钦）$\{\xi_n\}$服从同分布。

三个条件满足任意一个，则$\dfrac{1}{n}\sum\limits_{i=1}^{n}\xi_i\xrightarrow{P}\dfrac{1}{n}\sum\limits_{i=1}^{n}\mathrm{E}\xi_i$.

①以上三个条件的设计都是基于以下理由：

$$\mathrm{D}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i-\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\xi_i\right)=\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{D}\xi_i\Rightarrow P(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i-\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\xi_i\right|\ge\varepsilon)\le\frac{1}{n^2\varepsilon^2}\sum_{i=1}^{n}\mathrm{D}\xi_i.$$

三个条件满足任意一个，则$\lim\limits_{n\to\infty}\dfrac{1}{n^2\varepsilon^2}\sum\limits_{i=1}^{n}\mathrm{D}\xi_i=0.$由夹逼性可以得到结论。

②意义：数据量很大时，频率终将趋向概率；可以用频率逼近从而求得概率，也可以用概率来解释频率。概率理论是有实际价值的，而不是理想化的空中楼阁。

5、中心极限定理的几种表述

（1）（Laplace）对于n次满足 0−1 分布的 Bernoulli 试验$\{\xi_1,\xi_2,...,\xi_n\}$，成功次数$S_n=\sum\limits_{i=1}^{n}\xi_i$，则在$n$充分大时，$\dfrac{S_n-np}{\sqrt{np(1-p)}}$近似服从标准正态分布 N(0,1)。数学形式为$\lim\limits_{n\to\infty}P\left(\dfrac{S_n-np}{\sqrt{np(1-p)}}\le x\right)=\int_{-\infty}^{x}\dfrac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\mathrm{d}t.$

（2）（定义形式）已知随机变量序列$\{\xi_n\}$，若存在数列$\{A_n\}$与$\{B_n\}$使得n充分大时$\dfrac{1}{B_n}\sum\limits_{i=1}^{n}\xi_i-A_n$逐渐趋于 N(0,1)，则称$\{\xi_n\}$满足中心极限定理。

（3）（Lindeberg−Levy）设$\{\xi_n\}$是彼此独立且同分布的随机变量序列，它们具有相同的期望μ与方差σ^2，则$\lim\limits_{n\to\infty}P\left(\dfrac{1}{\sqrt{n}\sigma}\sum\limits_{i=1}^{n}(\xi_i-\mu)\le x\right)=\int_{-\infty}^{x}\dfrac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\mathrm{d}t.$

（4）中心极限定理说明，很多概率分布的标准化形式$\dfrac{1}{\sqrt{\mathrm{D}\xi}}\sum\limits_{i=1}^{n}(\xi_i-\mathrm{E}\xi)$的最终归宿都是标准正态分布。数理统计中将重点研究正态分布和与之相关的分布。

第二章 数理统计概论

一、抽样分布的基本理论

1、数理统计基本概念

（1）总体：研究对象的全体。（相当于奖池）

（2）个体：总体中的每个元素（相当于奖池中的每张奖券）

（3）抽样：随机抽出总体中的某些个体。（抽奖）

（4）样本：抽样得到的若干个体所组成的集合。（抽到的奖券）其元素个数称为样本容量。

（5）观察值：也叫样本值，是具体观察到的样本的值。

（6）统计量：不含未知参数的关于样本的函数。代入样本观察值之后得到的函数值称为统计量的观察值。

*常见的统计量：均值 $\overline{X}$（有时也记为$\overline{X}_n$）$=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$；方差 $S_n^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$；样本均方差（标准差）$S_n=\sqrt{\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}$；$k$ 阶原点矩 $A_k=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^k$；$k$ 阶中心矩 $M_k=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^k$；顺序统计量 $X_{(1)}, X_{(2)},..., X_{(n)}$（把 $X_1, X_2,..., X_n$ 从小到大排列生成的数组）

☆☆☆在数理统计中的方差，如果没有特殊规定，都是以 $n-1$ 为分母；概率论中随机变量的方差 $\mathrm{D}\xi$ 如果写成平均数的形式，是以 n 为分母。前者的理由将在后面给出。

2、数理统计中的三个重要分布（<u>**重点在于定义！**</u>）

（1）χ^2 分布

①定义：若$\{X_1, X_2,..., X_n\}$相互独立且均服从标准正态分布 N(0,1)，则称 $\chi^2=\sum\limits_{i=1}^{n}X_i^2$ 服从 n 个自由度的χ^2 分布，记为$\chi^2(n)$.

②密度函数（不要求掌握）：
$$\varphi(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, & x>0\\ 0, & x\le 0\end{cases}$$

其中 $\Gamma(s)=\int_0^{+\infty}x^{s-1}e^{-x}\mathrm{d}x.$（$s>0$）

③ $\mathrm{E}\chi^2=n$，$D\chi^2=2n.$

④常用数值：χ^2 分布的上侧α分位数 $\chi^2_\alpha(n)$，满足 $P(\chi^2\ge\chi^2_\alpha(n))=\alpha.$

（2）t 分布

①定义：若随机变量 X 与 Y 相互独立，$X\sim$ N(0,1)，$Y\sim\chi^2(n)$，则称 $T=\dfrac{X}{\sqrt{Y/n}}$ 服从 n 个自由度的 t 分布，记为 t(n).

②密度函数（不要求掌握）：
$$\varphi(x)=\frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\dfrac{n}{2}\right)}\left(1+\frac{x}{n^2}\right)^{-\frac{n+1}{2}}.$$

③$n\geqslant 2$ 时，存在 E$T=0$；$n\geqslant 3$ 时，存在 D$T=\dfrac{n}{n-2}$.

④常用数值：t 分布的上侧α分位数$\mathrm{t}_\alpha(n)$，满足$P(T\ge \mathrm{t}_\alpha(n))=\alpha$.

性质：$\mathrm{t}_{1-\alpha}(n)=-\mathrm{t}_\alpha(n)$.

（3）F 分布

①若随机变量X与Y相互独立，$X\sim\chi^2(n_1)$，$Y\sim\chi^2(n_2)$，则称$F=\dfrac{X/n_1}{Y/n_2}$服从第一个自由度为n_1，第二个自由度为n_2的 F 分布，记为$\mathrm{F}(n_1, n_2)$.

②密度函数（不要求掌握）：$\varphi(x)=\begin{cases}\dfrac{\Gamma\left(\dfrac{n_1+n_2}{2}\right)}{\Gamma\left(\dfrac{n_1}{2}\right)\Gamma\left(\dfrac{n_2}{2}\right)}n_1^{\frac{n_1}{2}}n_2^{\frac{n_2}{2}}\dfrac{x^{\frac{n_1}{2}-1}}{(n_1x+n_2)^{\frac{n_1+n_2}{2}}}，& x>0\\ 0，x\le 0\end{cases}$

③n≥3 时，存在$\mathrm{E}F=\dfrac{n}{n-2}$；$n$≥5 时，存在$\mathrm{D}F=\dfrac{2n_2^2}{n_1(n_2-2)^2(n_2-4)}$.

④常用数值：F 分布的上侧α分位数$\mathrm{F}_\alpha(n_1,n_2)$，满足$P(F\ge \mathrm{F}_\alpha(n_1,n_2))=\alpha$.

性质：$\mathrm{F}_\alpha(n_1,n_2)=\dfrac{1}{\mathrm{F}_{1-\alpha}(n_2,n_1)}$.

☆形式上简记几种分布的关系：

$\chi^2(n)=\sum\limits_{i=1}^{n}[\mathrm{N}(0,1)]^2$，$\mathrm{t}(n)=\dfrac{\mathrm{N}(0,1)}{\sqrt{\chi^2(n)/n}}$，$\mathrm{F}(n_1,n_2)=\dfrac{\chi^2(n_1)/n_1}{\chi^2(n_2)/n_2}$.

在记忆这些关系时，需明确两个问题：**每一种分布由什么分布生成？生成的表达式是怎样的？**

在推导一个量满足什么分布时，一般不会通过分布函数来证明（不要求掌握，并且十分麻烦），而是要努力配凑三个分布定义式的形式。

3、正态分布统计量满足的分布（不要把这些结论用在其他分布）

（1）总体：$X\sim\mathrm{N}(\mu,\sigma^2)$；$\dfrac{X-\mu}{\sigma}\sim\mathrm{N}(0,1)$.

（2）均值：$\overline{X}\sim\mathrm{N}(\mu,\dfrac{\sigma^2}{n})$; $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim\mathrm{N}(0,1)$.

（3）方差S^2与均值独立，且$\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$.

（4）$\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim\mathrm{t}(n-1)$.

（5）若还有一个独立的正态分布量Y，满足$X\sim\mathrm{N}(\mu_1,\sigma_1^2)$，$Y\sim\mathrm{N}(\mu_2,\sigma_2^2)$，则有

$\dfrac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_{\mathrm{W}}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim\mathrm{t}(n_1+n_2-2)$，$\dfrac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2}\sim\mathrm{F}(n_1-1,n_2-1)$.

其中$S_{\mathrm{W}}=\sqrt{\dfrac{(n_1-1)S_X^2+(n_2-1)S_Y^2}{n_1+n_2-2}}$.（方差的加权平均）

二、参数估计

1、参数估计的目的：在有数据但是未知分布中的参数（如正态分布的μ, σ^2）时，需要通过统计数据估计参数的值。
2、参数估计的方法：得到一些样本观察值，用含有这些观察值及相关的统计量（如均值，方差，矩）的代数式来表示参数。
例如，在正态分布中，如果数据量够大，可认为样本均值近似为期望，样本方差$\frac{n-1}{n}S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$（此为旧的形式）近似为方差$\sigma^2$。于是有估计值：

$$\begin{cases} \hat{\mu} = \overline{X}, \\ \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{n-1}{n}S^2. \end{cases}$$

3、点估计（用一些数据点进行的参数估计）

（1）矩估计法：设待求参数有 k 个，为$\theta_1, \theta_2, ..., \theta_k$；利用原点矩$\begin{cases} \mathrm{E}X \approx \overline{X} \\ \mathrm{E}X^2 \approx \overline{X^2} \\ ... \\ \mathrm{E}X^k \approx \overline{X^k} \end{cases}$，解出这些参数的估计值。

（2）极大似然估计法：取得 n 个样本 $X_i = x_i\,(1 \le i \le n)$后，认为这种取值取得的概率是可能的极大值，即$P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)$关于各个参数应取极大值。
①形象例子：若 A 箱中有 99 个白球、1 个黑球，B 箱中有 1 个白球、99 个黑球。在不知道哪个是 A、哪个是 B 的情况下，任取一个箱子，随机取球，取出的是白球，这时倾向于认为这个箱子像是 A，即“似然”。也即：我们倾向于认为样本的概率分布是使得 $X_i = x_i\,(1 \le i \le n)$发生概率达到极大值的情况，如果该概率分布没有尽其所能做到“极大”，那么更可能发生的应是其他的事件。
②离散变量的最大似然估计：最大似然函数$L(\theta_1, ..., \theta_k) = P(X_1 = x_1)...P(X_n = x_n)$.
它应当取极值，所以$\frac{\partial L}{\partial \theta_j} = 0.\ (1 \le j \le k)$（在这里不讨论条件极值，一般来说这些参数都是独立的）
③连续变量的最大似然估计：$P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)$对连续变量而言趋于 0，没有意义，因此用极限来代替：$\lim\limits_{\varepsilon \to 0^+} P(|X_1 - x_1| \le \varepsilon, ..., |X_n - x_n| \le \varepsilon)$为极大。也就是这些变量落在长度相同的微小区间内的概率应该取得极大值，因此对应的概率密度应为最大值。对应的最大似然函数为：

$$L(\theta_1, ..., \theta_k) = \lim_{\varepsilon \to 0^+} \frac{P(|X_1 - x_1| \le \varepsilon, ..., |X_n - x_n| \le \varepsilon)}{V(|X_1 - x_1| \le \varepsilon, ..., |X_n - x_n| \le \varepsilon)} = \lim_{\varepsilon \to 0^+} \frac{\prod_{i=1}^{n} \int_{x_i - \varepsilon}^{x_i + \varepsilon} \varphi(t)\,\mathrm{d}t}{(2\varepsilon)^n}$$

$$= \lim_{\varepsilon \to 0^+} \prod_{i=1}^{n} \frac{\varphi(\xi_i) \cdot 2\varepsilon}{2\varepsilon} = \varphi(x_1)...\varphi(x_n);\text{ 其中，}\ \xi_i \in (x_i - \varepsilon, x_i + \varepsilon).$$

这里运用了积分中值定理。然后，$\frac{\partial L}{\partial \theta_j} = 0.\ (1 \le j \le k)$

（3）点估计的效果评价
①无偏性：若$\mathrm{E}\hat{\theta} = \theta$，则称$\hat{\theta}$是无偏的，否则称其为有偏。（不负期望）

②有效性：$\mathrm{D}\hat{\theta}$越小，则称$\hat{\theta}$越有效。

③一致性：若对任意ε>0，$\lim\limits_{n\to\infty}P\left(\left|\hat{\theta}-\theta\right|<\varepsilon\right)=1$，则称$\hat{\theta}$是$\theta$的一致估计。

④为什么方差分母是 $n-1$：

方差的旧定义是$\widetilde{S}^2=\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$.将它直接当做正态分布$\sigma^2$的估计值，则

$$\mathrm{E}\hat{\sigma}^2=\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right]=\frac{1}{n}\mathrm{E}\left[\sum_{i=1}^{n}(X_i^2-2X_i\overline{X}+\overline{X}^2)\right]$$

$$=\frac{1}{n}\mathrm{E}\left[n\overline{X^2}-2n\overline{X}\cdot\overline{X}+n\overline{X}^2\right]=\mathrm{E}\overline{X^2}-\mathrm{E}(\overline{X}^2)=\mathrm{E}(X^2)-\mathrm{E}(\overline{X}^2)$$

$$=\mathrm{D}X+(\mathrm{E}X)^2-\mathrm{D}\overline{X}-(\mathrm{E}\overline{X})^2=\mathrm{D}X-\mathrm{D}\overline{X}=\left(1-\frac{1}{n}\right)\sigma^2\neq\sigma^2.$$

也就是说，旧的方差定义使得σ^2的估计值有偏离。从推导看，这偏离来自于$\mathrm{D}\overline{X}$，也即平均值也是有波动的。所以，修改方差定义，$S^2=\frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2$，将它作为$\sigma^2$的估计值，则$\mathrm{E}\hat{\sigma}^2=\sigma^2$. 鉴于正态分布较为常用，对一般的统计数据的方差也都使用新的定义。

4、区间估计

（1）基本思想：考察参数落在哪个区间中的概率比较大。

（2）置信区间：若$P(\theta_1<\theta<\theta_2)=1-\alpha$，则称$(\theta_1,\theta_2)$是$\theta$的置信度为 $1-\alpha$的置信区间。

（3）☆正态分布的区间估计（其中 u 指标准正态分布 N(0,1)；截图来自朱胜林老师概率统计课件，此课件完整版不外传）

待估计参数	已知	使用分布公式	区间
μ	σ^2	$U=\frac{\overline{X}-\mu}{\sqrt{\sigma^2/n}}$	$\lvert U\rvert<u_{\alpha/2}$
μ		$T=\frac{(\overline{X}-\mu)\sqrt{n}}{S}\sim t(n-1)$	$\lvert T\rvert<t_{\alpha/2}(n-1)$
σ^2	μ	$\chi=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\sim\chi^2(n)$	$\chi^2_{1-\alpha/2}(n)<\chi<\chi^2_{\alpha/2}(n)$
σ^2		$\chi=\frac{n-1}{\sigma^2}S^2\sim\chi^2(n-1)$	$\chi^2_{1-\alpha/2}(n-1)<\chi<\chi^2_{\alpha/2}(n-1)$
$\mu_1-\mu_2$	σ_1^2,σ_2^2	$U=\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$	$\lvert U\rvert<u_{\alpha/2}$
$\mu_1-\mu_2$		$T=\frac{\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}}{\sqrt{\frac{(n_1-1)S_X^2+(n_2-1)S_Y^2}{n_1+n_2-2}}}\sim t(n_1+n_2-2)$	$\lvert T\rvert<t_{\alpha/2}(n_1+n_2-2)$
σ_1^2/σ_2^2		$F=\frac{\sigma_2^2}{\sigma_1^2}\frac{S_X^2}{S_Y^2}\sim F(n_1-1,n_2-2)$	$F_{1-\alpha/2}(n_1-1,n_2-2)<F<F_{\alpha/2}(n_1-1,n_2-2)$

（最后一行：n_2-2 均改为 n_2-1）

三、假设检验

1、假设检验是区间估计的逆过程，主要是在对参数进行了假设之后看样本所处的区间是否可以接受。

2、假设检验的基本概念

（1）零假设 H_0：对参数的初始假设，可以是等式或不等式。

（2）备假设 H_1：与 H_0 对立的假设。

（3）弃真错误：H_0 正确，但是检验认为 H_0 错误。
（4）取伪错误：H_0 错误，但是检验认为 H_0 正确。
（5）双边检验：$H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$.
（6）左边检验：$H_0 : \theta = \theta_0$ vs $H_1 : \theta < \theta_0$.（H_0 也可以是$\theta \geqslant \theta_0$）
（7）右边检验：$H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$.（H_0 也可以是$\theta \leqslant \theta_0$）
3、假设检验的基本思想：确定 H_0 之后，选择一个关于样本均值的置信区间（在这里称为接受域，其补集称为拒绝域）。在置信度较大的情况下，若有数据落入拒绝域，说明小概率的事件发生了，此时一般拒绝假设 H_0；反之，则接受 H_0。
4、☆正态分布的假设检验（截图来自朱胜林老师概率统计课件，此课件完整版不外传）

名称	H_0	公式	拒绝域	说明
单一样本 u-检验	$\mu=\mu_0$	$u=\sqrt{n}\frac{\bar{x}-\mu_0}{\sigma}$	$\lvert u\rvert>u_{\alpha/2}$	$n>30$，σ 已知
双样本 u-检验	$\mu_1=\mu_2$	$u=\frac{(\overline{x_1}-\overline{x_2})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$\lvert u\rvert>u_{\alpha/2}$	σ_1,σ_2 已知
单一样本 t-检验	$\mu=\mu_0$	$t_{(n-1)}=\frac{\bar{x}-\mu_0}{s/\sqrt{n}}$	$\lvert t_{(n-1)}\rvert>t_{\alpha/2}$	$n>30$，σ 未知
双样本 t-检验	$\mu_1=\mu_2$	$t_{(n_1+n_2-2)}=\frac{(\overline{x_1}-\overline{x_2})-(\mu_1-\mu_2)}{s_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$，其中：$s_w^2=\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$	$\lvert t_{(n_1+n_2-2)}\rvert>t_{\alpha/2}$	$n_1+n_2>40$ 且 $\sigma_1=\sigma_2$ 未知
双样本方差的F-检验	$\sigma_1^2=\sigma_2^2$	$F_{(n_1-1,n_2-1)}=\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$	$F_{(n_1-1,n_2-1)}>F_{\alpha/2}$或 $F_{(n_1-1,n_2-1)}<F_{1-\alpha/2}$	
方差的 χ^2-检验	$\sigma^2=\sigma_0^2$	$\chi^2_{(n-1)}=(n-1)\frac{s^2}{\sigma_0^2}$	$\chi^2_{(n-1)}>\chi^2_{\alpha/2}$或 $\chi^2_{(n-1)}<\chi^2_{1-\alpha/2}$	
单一样本 u-检验	$\mu\le\mu_0$	$u=\sqrt{n}\frac{\bar{x}-\mu_0}{\sigma}$	$u>u_\alpha$	$n>30$，σ 已知
其他的单边检验可如上一条相似得到，不一一列举。				

5、总体分布的假设检验：如果连分布类型也不知道，则需要根据数据选择合适的模型进行拟合，拟合结果有待假设检验。

在此只介绍总体分布的χ^2 拟合检验法。

（1）总体分布函数 $F(x)$未知。初始假设为 H_0：$F(x) = F_0(x)$.
（2）若 $F_0(x)$不带任何参数，取实数$-\infty = t_0 < t_1 < t_2 <\ldots< t_{k-1} < t_k = +\infty$，记区间 $A_i = (t_{i-1}, t_i]$；设 n 个样本观察值中落在 A_i 内的个数为 n_i，$p_i = F_0(t_i) - F_0(t_{i-1})$。如果假设正确，则概率与频率应接近，即$\left|p_i-\frac{n_i}{n}\right|$很小。令$\chi^2=\sum_{i=1}^{k}\frac{n}{p_i}\left(p_i-\frac{n_i}{n}\right)^2$；假设正确的充要条件是$\chi^2 \sim \chi^2(k-1)$，拒绝域为 $\chi^2>\chi^2_\alpha(k-1)$.
（3）若 $F_0(x)$带有 r 个参数，将它们全部用最大似然估计值代入，然后按照上述步骤进行。假设正确的充要条件是$\chi^2 \sim \chi^2(k-r-1)$，拒绝域为 $\chi^2>\chi^2_\alpha(k-r-1)$.

[参考资料]
[1]复旦大学数学科学学院朱胜林老师《线性代数与概率统计》课程
[2]金路, 童裕孙, 於崇华, 张万国. 高等数学（第四版，下册）.高等教育出版社, 第 11-12 章
[3]茆诗松, 程依明, 濮晓龙. 概率论与数理统计教程（第二版）.高等教育出版社