Probablity and Statistics

Jianqi Huang

October 2022

该笔记是基于洪永淼教授所著《概率论与统计学》所学以及日常的积累所做的 lecture note.

1 随机变量基本概念补充

Definition 1.1. 令 $\mathbb{B}_\Omega$ 为从 Ω 生成的包含的 (a,b) 类似形式的最下 $sigma$ 域，其中对于 B 中的任意一个集合 A，都有

$$P(X \in A) = P(Y \in A)$$

则随机变量 X 和 Y 服从同分布.

1.1 CDF 与 P(X)

$CDFF_X(x)$ 是由随机变量的概率函数 $P(\cdot)$ 所决定的，而同样 $CDFF_X(x)$ 可以推出 $P(X)$. 因此这两者包含随机变量相同的概率分布信息.

Definition 1.2. 若两个分布 $F(\cdot)$ 和 $G(\cdot)$ 在实数轴上均满足 $F(x) \leq G(x)$，则称分布 $F(\cdot)$ 一阶随机占优于 $G(\cdot)$.
同样的对任意的 $x \in (\infty, \infty)$ 均满足

$$\int_{-\infty}^{x} F(y)dy \leq \int_{-\infty}^{x} G(y)dy$$

则是二阶随机占优.

1.2 支撑

Definition 1.3. *在实数集 $\mathbb{R}$ 上的概率为正数的点所构成的集合就是 X 的支撑 (Support) 集合*

$$Support(X) = \{x \in \mathbb{R} : f_X(x) > 0\}$$

FOC:

$$Support(X) = \Omega$$

通过定义可以看出，X 的支撑都是取 $P(\cdot)$ 概率为正的点所构成的集合.
对于离散随机变量来说，X 的支撑及概率能够刻画包含的该随机变量的概率分布；而对于连续型的随机变量来说，在其支撑上的集合的点的极小邻域内概率取值都为正，其支撑外的点点极小邻域内的点概率为 0.

$$P_{X\in Support}(x \in (X-\epsilon, X+\epsilon)) > 0$$
$$P_{X\notin Support}(x \in (X-\epsilon, X+\epsilon)) = 0$$

Lemma 1.1. *令 $f_X(x)$ 为随机变量 X 的 PDF，并且 μ 和 $\sigma > 0$，则函数*

$$g_X(x) = \frac{1}{\sigma} f_X(\frac{x-\mu}{\sigma})$$

仍然是 PDF

要满足是一个概率分布函数 (PDF)，需要满足 $g_X(x) \geq 0$ 以及在 $\mathbb{R}$ 上的积分为 1. 因此通过换元法来证明令 $y = \frac{x-\mu}{\sigma}$，则

$$\int_{\mathbb{R}} g_X(x) = \frac{1}{\sigma}\int_{\mathbb{R}} f_X(\frac{x-\mu}{\sigma})dx = \frac{1}{\sigma}\int_{\mathbb{R}} f_Y(y)d\sigma y = 1$$

其中的参数值 (μ, σ) 对应于一个 PDF，由参数 μ 刻画 PDF 的位置，称为位置参数类；由 σ 刻画的函数类 $\frac{1}{\sigma} f_X(\frac{x}{\sigma})$ 称为 PDF $f_X(x)$ 的标准尺度类.

将支撑推广到随机向量上，对对于一个二维随机向量 (X, Y) 的支撑，为所有严格正概率的可能实现值 (x, y) 的集合，即

$$Support(X, Y) = \Omega_{XY} = \{(x, y) \in \mathbb{R}^2 : f_{XY}(x, y) > 0\}$$

1.3 随机变量关系推导

相互独立是一个较为强的命题.

对于一个 $Y = CDF(X)$ 求解 PDF，该方法需要先求出 $Y = g(X)$ 的 CDF $F_Y(y)$，再对求导得到 PDF，若 $g(\cdot)$ 为严格单调函数，若连续随机变量 X 的 PDF 为 $f_X(x)$，函数 g 在上面单调可导. 那么对于随机变量 $Y = g(X)$ 来说，其支撑的密度函数可表示为

$$f_Y(y) = f_X(x)\frac{1}{|g'(x)|}$$

Theorem 1.1. *两个随机变量由原来的联合分布拆解为两个互不影响的函数，只有当 X 和 Y 互相独立:*

$$f_{XY}(x, y) = g(x)h(y)$$

对于所有的 $-\infty < x, -\infty < y$ 均成立.

Theorem 1.2. *进一步来说对于一个随机变量 X 和 Y 也满足相互独立, 则条件 PMF/PDF*

$$f_{Y|X}(y|x) = f_Y(y)$$

对所有的 $(x, y) \in \mathbb{R}^2$ 成立，其中 $f_X(x) > 0$

反过来，若 n 个随机变量的联合 CDF 等于边际 CDF 的乘积

$$F_{\mathbf{X}^n}(\mathbf{x}^n) = \prod_{i=1}^{n} F_{\mathbf{X}_i}(x_i)$$

则其随机变量之间相互独立.

而若将一元转到二元，假设有

$$\begin{cases} U = g_1(X, Y) \\ V = g_2(X, Y) \end{cases}$$

其中的 (X, Y) 是给定联合概率密度. 如何求 (U, V) 的概率密度. 先引入雅可比矩阵

Definition 1.4. 对于一个二元变换来说，

$$\begin{cases} U = g_1(X, Y) \\ V = g_2(X, Y) \end{cases}$$

其中的 $g_1(\cdot,\cdot)$ 和 $g_2(\cdot,\cdot)$ 均连续可导，则 2×2 矩阵

$$J_{UV}(x,y)=\begin{bmatrix}\frac{\partial g_1(x,y)}{\partial x} & \frac{\partial g_2(x,y)}{\partial x}\\ \frac{\partial g_2(x,y)}{\partial x} & \frac{\partial g_2(x,y)}{\partial y}\end{bmatrix}$$

这个称为 (U,V) 的雅可比矩阵，行列式为雅可比.

同时对于一个雅可比矩阵来说，同时还满足的反函数的定义

$$J_{XY}(u,v)=J_{UV}(x,y)^{-1}$$

其中 $x_1=h_1(u,v),x_2=h_2(u,v)$

Theorem 1.3. *假设二元变换连续型随机变量* (X,Y) *的联合 PDF 为* $f_{XY}(x,y)$，*并记* (X,Y) *的支撑为* $\Omega_{XY}=\{(x,y)\in\mathbb{R}^2: f_{XY}(x,y)>0\}$ *其中* $g:\Omega_{XY}\to\mathbb{R}^2$ *为一一映射，在* Ω_{XY} *上连续可导，对于所有的* x，y *都有行列式不为 0，则* (U,V) *的联合 PDF 为*

$$f_{UV}(u,v)=f_{XY}(x,y)\left|det[J_{XY}(u,v)]\right|$$

对所有的 $(u,v)\in\Omega$ *均成立.*

$$\Omega_{UV}=\{(u,v)\in\mathbb{R}^2: u=g_1(x,y),v=g_2(x,y),\forall(x,y)\in\Omega_{XY}\}$$

从 PDF 到 MGF

Theorem 1.4. *假设联合* $MGFM_X(t_1,t_2)=E(e^{t_1X+t_2Y})$ *对在原点的某个邻域内存在，当且仅当对原点的所有点有*

$$M_X(t_1,t_2)=M_X(t_1)M_Y(t_2)$$

成立时，X *和* Y *互相独立.*

2 矩生成函数

随机变量 X 的高阶距可能很难算，因此需要用定义一个所谓的矩生成函数来求各阶距.

Definition 2.1 (矩母生成函数 (Moment Generating Function),MGF)**.** *随机变量* X *的* MGF

可定义为

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum_{x\in\Omega_X} e^{tx} f_X(x), X \\ \int_{-\infty}^{\infty} e^{tx} f_X(x)dx, X \end{cases}$$

MGF 的存在性：若对于 t 在 *0* 的某个领域 $(-\epsilon, \epsilon)$ 上是存在的，则称 $M_X(t)$ 是存在的，若不存在就对于 X 分布来说是不存在的.

因此从 MGF 可以推导到一个各阶距是否是存在的，若对于所有的 t 都存在 MGF，则可推出在所有阶的导数存在.

证明.

$$e^{tx} = \sum_{k=0}^{\infty} \frac{(tx)^k}{k!}$$

则 $M_X(t)$ 的麦克劳林展开为

$$M_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} E(X^k)$$

□

反过来对于某个阶的期望为 ∞，则对于 $t > 0$ 所有的矩母函数都不存在.

Theorem 2.1. 若 *MGF* 的 t 在某个领域 (ϵ, ϵ) 内存在，则 $M_X(0) = 1$

证明. $M_X(t) = \int_{\infty}^{\infty} e^{tx} dF_X(x)$，有 $M_X(0) = \int_{-\infty}^{\infty} dF_X(x) = 1$ □

Theorem 2.2. 若 $MGFM_X(t)$ 对于 t 在 *0* 的某个邻域内存在,则对所有正整数 $k = 1, 2, \cdots$ 有

$$M_X^{(k)}(0) = E(X^k)$$

证明.

$$\begin{aligned} M_X^{(k)}(t) &= \frac{d^k}{dt^k} \int_{-\infty}^{\infty} e^{tx} dF_X(x) \\ &= \int_{-\infty}^{\infty} \frac{d^k}{dt^k}(e^{tx}) dF_X(x) \\ &= \int_{\infty}^{\infty} x^k e^{tx} dF_X(x) \end{aligned}$$

令 t=0，则

$$M_X^{(k)}(0)=\int_{-\infty}^{\infty}x^k dF_X(x)=E(X^k)$$

□

因此若矩母函数存在，可以在原点处对其进行求导 ($k=1,2,\cdots$)，从而可以得到 X 的各阶矩．因此 $M_X(t)$ 也称为矩生成函数．

Example 2.1.

$$M_X^{(1)}(0)=\mu_X$$
$$M_X^{(2)}(0)=E(X^2)\sigma_X^2+\mu_X^2$$

Theorem 2.3. *若 $Y=a+bX$，其中 a 和 b 为两个常数，并且对于在 0 处的某个小邻域内的所有 t，X 的 MGF$M_X(t)$ 存在，则对于 0 在某个小邻域内的所有 t，Y 的 MGF 也存在，且为*

$$M_Y(t)=e^{at}M_X(bt)$$

Theorem 2.4. *假设两个随机变量 X 和 Y 的 MGF $M_X(t)$ 和 $M_Y(t)$ 在 0 的某个邻域内的 $N_\epsilon(0)=\{t\in\mathbb{R}:-\epsilon<t<\epsilon\}$ 存在，则对任意的 $z\in R$ 都有 $F_X(z)=F_Y(z)$，当且仅当所有的 $t\in N_\epsilon(0)$，X 和 Y 都有相同的 MGF $M_X(t)$ 和 $M_Y(t)$.*

因此对于一个猜想分布来说，若能够满足 MGF 的定义，最终的分布就是所猜想的分布．

Example 2.2.

$$M_X(t)=\frac{1}{2}+\frac{1}{4}e^{-t}+\frac{1}{4}e^{t},-\infty<t<\infty$$

其中

$$M_X(t)=E(e^{tX}) \qquad =\sum_x e^{tx}f_X(x)=\frac{1}{2}e^{0\cdot t}+\frac{1}{4}e^{(-1)\cdot t}+\frac{1}{4}e^{1\cdot t}$$

猜想 X 的概率密度为

$$f_X(x)=\begin{cases}\frac{1}{4} & x=-1\\ \frac{1}{2} & x=0\\ \frac{1}{4} & x=1\end{cases}$$

Theorem 2.5. *假设 $\{X_n,n=1,2\cdots\}$ 为随机变量序列，每一个随机变量 X_n 有 MGF 和 CDF，进一步的假设，对于 t 在 0 点某个小领域内的任意值．有*

$$\lim_{n\to\infty}M_n(T)=M_X(t)$$

同时 $F_X(x)$ 也是弱收敛的，即

$$\lim_{n\to\infty} F_n(X) = F_X(x)$$

Theorem 2.6. 假设 $F_X(x)$ 和 $F_Y(y)$ 是有界支撑的 *CDE*，对于所有的整数 $E(X^k) = E(Y^k)$，则任意的 $z \in (-\infty, +\infty)$ 都有 $F_X(x) = F_Y(y)$

证明. 因为 X 和 Y 所属于的是一个有界支撑的集合，因此对于一个充分大的数 M，$P(|X| \leq M) = 1$ 及 $P(|Y| \leq M) = 1$，又因 $E|X| \leq M^k$ 和 $E|Y|^k \leq M^k$ 有

$$\begin{aligned} M_X(t) &= \sum_{k=0}^{\infty} \frac{t^k}{k!} E(X^k) \\ &\leq \sum_{0}^{\infty} \frac{t^k}{k!} E|X|^k \\ &\leq \sum_{k=0}^{\infty} \frac{(tM)^k}{k!} = e^{tM} < \infty \end{aligned}$$

同样的有 $M_Y(t) \leq e^{tM} < \infty$ 根据 MGF 的公式可得到 $M_X(t) = M_Y(t)$，其中 MGF 的唯一性定理则 $F_X(z) = F_Y(z)$ 对于 $\forall z$ □

2.1 联合矩生成函数

从一维到多维，定义同样没有发生较大的变化

$$M_{XY}(t_1, t_2) = E(e^{t_1 X + t_2 Y}), -\infty < t_1, t_2 < \infty$$

其中的期望对于在 $(0, 0)$ 范围内的所有 (t_1, t_2) 均成立．假设所有的 $(0, 0)$ 的某个领域内存在，则对于所有的非负整数 $r, s \geq 0$ 有

$$E(X^r Y^s) = M_{XY}^{(r,s)}(0, 0)$$

同时

$$cov(X^r, Y^s) = M_{XY}^{(r,s)}(0, 0) - M_X^{(r)}(0) M_Y^{(s)}(0)$$

特别的

$$cov(X, Y) = M_{XY}^{(1,1)}(0, 0) - M_X^{(1)}(0) M_Y^{(1)}(0)$$

Theorem 2.7. 假设 $U = g_1(X)$ 和 $V = g_2(Y)$ 是连续可导一一映射的可测函数，则 X 和

Y 相互独立，当且仅当 U 和 V 相互独立.

2.2 独立性和期望

假设 (X, Y) 相互独立，则对于任意的可测函数 $h(X)$ 和 $q(Y)$ 有

$$E[h(X)q(Y)] = E[h(X)]E[q(Y)]$$

或者等等价于协方差为 0.

证明.

$$E[h(X)q(Y)] = E[h(X)]E[q(Y)]$$

□

Lemma 2.1. 假设 X 和 Y 相互独立，并且二者的边际 $MGFM_X(t)$ 和 $M_Y(t)$ 对于 t 在 0 的某个邻域内存在，则 $M_{X+Y}(t)$ 对于 t 在 0 的某个邻域内的所有 t 都存在. 且

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

证明.

$$g(X, Y) = e^{t(X+Y)} = e^{tX}e^{tY} = h(X)q(Y)$$

由上独立性定理可得

$$E[e^{t(X+Y)}] = E(e^{tX})E(e^{tY})$$

即

$$M_{X+Y} = M_X(t)M_Y(t)$$

□

这个的定理可以帮助我们在一些重要分布中得到使用，对于一对 $X \sim N(\mu_1, \sigma_1)$ 和 $Y \sim N(\mu_2, \sigma_2)$ 正态分布来说，若相互独立则

$$X \pm Y \sim N(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2)$$

同样，这个可推及其他分布，柏松分布、卡方分布也有相同的性质.

3 特征函数

对于一些重要分布来说，MGF 并不存在，因此就需要引入一个对任何概率分布都存在的特征函数.

Definition 3.1. 假设随机变量的 *CDF* 为 $F_X(x)$，则其特征函数为

$$\varphi_X(t) = E(e^{itX}) = \int_{\infty}^{\infty} e^{itx} dF_X(x)$$

其中 $i = \sqrt{-1}$，同时

$$e^{itx} = \cos(tx) + i\sin(tx) \tag{1}$$

3.1 特征函数的性质

- 对于任意的概率分布，其特征函数总是存在且有下界的
- $\varphi_X(0) = 1$
- $\varphi_X(0)$ 在 $(-\infty, +\infty)$ 上连续;
- $\varphi_X(-t) = \varphi_X(t)^*$ 其中 $\varphi_X(t)^*$ 表示为复共轭 (complex conjugate)
- 假设 $Y = a + bX$ 其中 a 和 b 为任意实常数，则

$$\varphi_Y(t) = e^{iat}\varphi_X(bt)$$

- 若 MGF$M_X(t)$ 在 0 的某个邻域内的所有 t 都存在，则对于所有的 t，都有

$$\varphi_X(t) = M_X(it)$$

Example 3.1. 假设随机变量 X 服从柯西 $(0,1)$ 分布，其 *PDF* 为

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}, -\infty < x < \infty$$

则其特征函数存在且为

$$\varphi_X(t) = e^{-|t|}$$

Theorem 3.1. 假设 X 的 k 阶矩存在，则 $\varphi_X(t)$ 对 $t \in (-\infty, +\infty)$ 是 k 阶可导的，且

$$\varphi_X(0)^{(k)} = i^k E(X^k)$$

Theorem 3.2. 假设两个随机变量 X 和 Y 的特征函数为 $\varphi_X(t)$ 和 $\varphi_Y(t)$，则 X 和 Y 具有同分布，当且仅当对所有的 $t \in (-\infty, +\infty)$ 有 $\varphi_X(t) = \varphi_Y(t)$

证明. 根据定义，特征函数为 CDF 的傅立叶变换

$$\varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx} dF_X(x)$$

因此可以看到对于每一个给定的分布 $F_X(x)$ 都有唯一的特征函数.

□

Theorem 3.3. 对于随机向量序列 $\{X_n\}$ 的 *CDF* 和特征函数分别为 $F_n(x)$ 和 $\phi_n(t)$，又假设随机变量 X 的 *CDF* 和特征函数分别为 $F_X(x)$ 和 $\phi_X(t)$，令 $n \to 0$ 若对 $F_X(x)$ 对所有连续点 x，有 $F_n(X) \to F_X(x)$，则对任意的 $t \in (-\infty, \infty)$ 都有 $\phi_n(t) \to \varphi_X(t)$ 若对于任意的，则对于 $F_X(x)$ 的连续点 x，有 $F_n(x) \to F_X(x)$

一些情况下使用特征函数比分布函数更为简便.

4 统计抽样理论

Definition 4.1. 一个随机样本是由 n 个随机变量 $X_1, \cdots, X_n$ 构成的序列，记作 $\boldsymbol{X}^n = (X_1, \cdots, X_n)$，而随机样本的一个实现值就是从随机样本 $\boldsymbol{X}^n$ 中抽取的一个数据集 $\boldsymbol{x}^n$

当抽取一个样本时候获得样本的实现值 $\mathbf{x}^n$，我们想要获取信息，就是要通过在原有的样本空间下生成一个映射 $T(\cdot)$（一个标量），那么这个 $T(\mathbf{X}^n)$ 就是一个统计量.

随机样本的对数似然函数令 $\mathbf{X}^n = (X_1, \cdots, X_n)$ 来自于总体的 IID 随机样本，其中 θ 是未知参数 $\mathbf{X}^n$ 的联合 PMF/PDF 的对数为

$$\hat{L}(\theta | X^n) = \ln \prod_{i=1}^{n} f(X_i, \theta) = \sum_{i=1}^{n} \ln f(X_i, \theta)$$

Definition 4.2. 统计量 $T(\hat{\boldsymbol{X}}^n)$ 的概率分布称为 $T(\boldsymbol{X}^n)$ 的抽样分布. 实际运用中并非这么称呼，一般叫“样本”“样本方差”等等.

Definition 4.3 (充分统计量)**.** 令 $\boldsymbol{X}^n$ 为来自以 θ 的某个总体的随机样本，给定统计量

$T(\boldsymbol{X}^n)$ 的值，若 $\boldsymbol{X}^n=\boldsymbol{x}^n$ 的条件分布并不依赖于 θ 的值，那么对于所有可能的 θ 值，有

$$f_{\boldsymbol{X}^n|T(\boldsymbol{x}^n)}[\boldsymbol{X}^n|T(\boldsymbol{X}^n),\theta]=h(\boldsymbol{x}^n)$$

充分统计量概括了样本 $\mathbf{X}^n$ 的关于 θ 的所有信息，数据集 $\mathbf{x}^n$ 就没有提供更多的信息. 可帮助计算参数时候只需要关注于低维的统计量.
上述的定理中，等式左边为给定的 $T(\mathbf{X}^n)=T(\mathbf{x}^n)$ 时，$\mathbf{X}^n=\mathbf{x}^n$ 依赖于 θ，而等式右边并不依赖于 θ. 也就可以理解为给定 θ 和统计量 $T(\mathbf{X}^n)$ 下，f 分布只与样本有关.

Theorem 4.1 (因子分解定理 (Factorization Theorem))**.** 令 $f_{\boldsymbol{X}^n}(\boldsymbol{x}^n,\theta)$ 为随机样本 $\boldsymbol{X}^n$ 的联合 *PDF/PMF*，当且仅当存在 $g(t,\theta)$ 和 $h(\boldsymbol{x}^n)$ 时候，满足对于任意样本点的 $\boldsymbol{x}^n$ 使得对任意参数值有

$$f_{\boldsymbol{X}^n}(\boldsymbol{x}^n,\theta)=g[T(\boldsymbol{x}^n),\theta]h(\boldsymbol{x}^n)$$

证明略.

Theorem 4.2 (不变性原理 (Invariance Principle))**.** 假设 $T(\boldsymbol{X}^n)$ 是 θ 的充分统计量，则对于任意一一对应的函数 $R(\boldsymbol{X}^n)=r[T(\boldsymbol{X}^n)]$ 也是 θ 的充分统计量. 也是变换参数 $r(\theta)$ 的充分统计量.

Definition 4.4. 概率分布族称为指数分布族，若有总体 *PMF/PDF* 可表示为

$$f(x;\theta)=h(x)c(\theta)e^{\sum_{j=1}kw_j(\theta)t_j(x)}$$

随机样本 $\mathbf{X}^n$ 下关于参数 θ 的信息可能无法由单个统计量概括，需要多个统计量描述. 这里的充分统计量是一个向量. 在另一个角度，随机样本本身就是 θ 的一个充分统计量. 其分解为

$$f_{X^n}(\boldsymbol{x}^n,\theta)=g[T(\boldsymbol{x}^n),\theta]h(\boldsymbol{x}^n)$$

一般来说，同一个参数 θ 存在多个充分统计量，而如何对这个统计量内部再进行刻画：引入最小充分统计量

Definition 4.5 (最小充分统计量)**.** 对于任何的其他充分统计量 $R(\boldsymbol{X})^n$ 来说，$T(\boldsymbol{X})^n$ 总是存在一个 $r(\cdot)$ 使得

$$T(\boldsymbol{X}^n)=r[R(\boldsymbol{X}^n)]$$

则称充分统计量 $T(\boldsymbol{X}^n)$ 是参数 θ 的最小充分统计量.

4.1 $\mathcal{F}$ 分布

Definition 4.6. U 和 V 是自由度分别为 p 和 q 的两独立卡方分布随机变量，则

$$F=\frac{U/p}{V/q}\sim\mathcal{F}_{p,q}$$

服从自由度为 p 和 q 的 $\mathcal{F}$ 分布.

$\mathcal{F}$ 分布的性质

- if $X\sim\mathcal{F}_{p,q}$，则 $X\sim\mathcal{F}_{q,p}$
- if $X\sim t_q$，$X^2\sim\mathcal{F}_{1,q}$
- if $q\to\infty$，$p\mathcal{F}_{p,q}\to\chi_p^2$

Example 4.1. 总体方差相等的假设检验：令 $\boldsymbol{X}=(X_1,\cdots,X_n)$，为来自总体正态分布 $N(\mu_X,\sigma_X^2)$，样本容量为 n 的 *IID* 随机样本，$Y^m=(Y_1,\cdots,Y_n)$ 为来自 $N(\mu_Y,\sigma_Y^2)$，样本容量为 m 的 *IID* 随机样本. 假设两个随机变量相互独立. 考虑一个 $\mathbb{H}_0:\sigma_X^2=\sigma_Y^2$，实际我们需要检验：

$$\frac{S_X^2}{S_Y^2}$$

在 $\mathbb{H}_0:\sigma_X^2=\sigma_Y^2$ 的假设下，有

$$\frac{S_X^2}{S_Y^2}=\frac{S_X^2/\sigma_X^2}{\sigma_Y^2/\sigma_Y^2}\sim\frac{\chi_{n-1}^2/(n-1)}{\chi_{m-1}^2/(m-1)}\sim\mathcal{F}_{n-1,m-1}\tag{2}$$

t 分布和 F 分布在数据生成过程中来自正态总体的抽样问题发挥重要作用.

5 收敛和极限定理

Definition 5.1. 一个足够大的数 M，存在一个有限整数 $N(M)$ 使得对所有的 $n\geq N(M)$，有 $n^{-\lambda b_n}<M$，序列 $\{b_n,n=1,2\cdots\}$ 的最高数量级为 n^λ 记作 $b_n=O(n^\lambda)$ 用 M 来控制 $b_n=O(n^\lambda)$ 使得

$$\lim_{n\to\infty}\frac{b_n}{n^\lambda}=C<\infty$$

数量级刻画的是一个数增长的速度，而进一步对于统计中，我们需要找到一些样本，而样本本身的个数是有限的，自然需要利用数量级的概念来刻画统计量的性质.

因此,在一个 n 维的有限样本中,n 可以无穷大,X^n 的实现值是一个 n 维的向量 $x^n(x_1, x_2, \cdots, x_n)$.因为 X^n 是一个随机变量的序列，可用 $\mathbf{X}^n$ 的联合概率来刻画随机样本.

$$f_{\mathbf{X}^n}(\mathbf{x}^n) = \prod_{i=1}^{n} f_{X_i|X^{i-1}}(x_i|\mathbf{x}^{i-1})$$

统计分析的重要目的在于用给定的随机样本生成的数据集，估计未知的参数.

5.1 依二次方均值和 L-p 收敛和依概率收敛

Definition 5.2. 令 $\{Z_n, n = 1, 2, \cdots\}$ 为一个随机变量序列，Z 为随机变量. 若当 $n \to \infty$ 时

$$E(Z_n - Z)^2 \to 0$$

或等价的

$$\lim_{n\to\infty} E(Z_n - Z)^2 = 0$$

则称随机序列依二次方均值收敛于 Z.

Definition 5.3. 假设 $0 < p < \infty$，$\{Z_n, n = 1, 2, \cdots\}$ 为满足 $E[Z_n]^p < \infty$ 的一个随机变量序列，Z 为满足条件的一个随机变量. 若

$$\lim_{n\to\infty} E|Z_n - Z|^p = 0$$

则随机变量序列 $\{Z_n\}$ 依 L_p 收敛到 Z.

Definition 5.4. 利用 $\epsilon - \delta$ 语言和概率理论对收敛的刻画:

$$P[|Z_n - Z| > \epsilon] \to 0$$

则称随机变量序列 $\{Z_n, n = 1, 2, \cdots\}$ 依概率收敛.

依概率收敛也称为弱收敛. 在偏差上理解：在 $|Z_n - Z| > \epsilon$ 中，可认为 $|Z_n - Z|$ 为大偏差，若 $|Z_n - Z| \leq \epsilon$ 则为小偏差，显然 epsilon 越小，n 越大，来保证收敛.

一个特例：当 $Z_n \to b$ 时候，其中 b 为常数，则称 Z_n 为 b 的一致估计量. 且 b 为 Z_n 的概率极限，记为 $b = p\lim_{n\to\infty} Z_n$

Definition 5.5. 对任意的常数 $\delta > 0$，存在常数 $M = M(\delta)$ 和有限整数 $N = N(\delta)$，使得对所有 $n \le N$，有 $P(|Z_n| > M) < \delta$. 则称 Z_n 是有界的.

在收敛性中，我们极限中引入概率对收敛性进行了一个刻画，同样对于统计量的刻画，我们也可以引入概率空间对其进行刻画.

Definition 5.6. 假设 $\boldsymbol{X}^n = (X_1, \cdots, X_n)$ 为来自均值 $E(X_i) = \mu$ 和方差 $var(X_i) = \sigma^2 < \infty$ 的总体分布的 *IID* 随机样本. 当 $n \to \infty$ 时，有

$$P[|\overline{X}_n - \mu| \le \epsilon] \to 1$$

Definition 5.7 (几乎处处收敛)**.** 若对于每一个给定的常数 $\epsilon > 0$ 有

$$P(\lim_{n\to\infty} |Z_n - Z|) = 0$$

其中 S 为样本空间，称随机 *1* 变量序列几乎处处收敛于随机变量 Z.

5.2 依分布收敛

当我们抽取的样本是一个非正态分布时候，$\overline{X_n}$ 的抽样分布会变得复杂，一个可行的方式是考虑容量 $n \to \infty$ 时，$\overline{X_n}$ 的一个极限分布，也称为渐近分布. 现实中的抽样，我们的总体分布是未知的，此时可以用依分布收敛概念对统计量进行近似分析.

Definition 5.8. 令 $\{Z_n, n = 1, 2, 3 \cdots\}$ 为随机变量序列，其对应的 *CDF* 序列为 $\{F_n(z) : n = 1, 2 \cdots\}$，令随机变量 Z 的 *CDF* 为 $F(z)$，若在每一个连续点都收敛，即在每个 $F(z)$ 连续的 z 点处都有:

$$\lim_{n\to\infty} F_{Z_n}(z) = F(z)$$

称 $n \to \infty$ 时候，Z_n 依分布收敛于 Z. 其中的 $F(z)$ 为随机变量序列的渐近分布.

Lemma 5.1. 当 $n \to \infty$ 时，$Y_n - Z_n \xrightarrow{p} 0$ 且 $Z_n \xrightarrow{d} Z$，则当 $n \to \infty$，$Y_n \xrightarrow{d} Z$

Definition 5.9. 对于某个点 c，若 $P(Z = c) = 1$，则称随机变量 Z 服从在 c 点的退化分布.

5.3 中心极限

Theorem 5.1. 令 $\boldsymbol{X}^n = (X_1, X_2, \cdots, X_n)$ 为来自均值 μ，方差为 σ^2 的总体分布的 *IID* 随机样本. 定义样本均值 $\bar{X_n} = n^{-1}\sum_{i=1}^{n} X_i$. 当 $n \to \infty$ 时，标准化样本均值为

$$
\begin{aligned}
Z_n &= \frac{\bar{X_n} - E(\bar{X_n})}{\sqrt{var(\bar{X_n})}} \\
&= \frac{\bar{X_n} - \mu}{\sigma/\sqrt{n}} \\
&= \frac{\sqrt{n}(\bar{X_n} - \mu)}{\sigma} \to N(0,1)
\end{aligned}
$$

Theorem 5.2 (独立随机变量的中心极限定理)**.** 假设随机变量 $X_1, \cdots, X_n$ 联合独立，且对 $i = 1, 2, \cdots, n$ 有 $E|X_i - \mu_i|^3 < \infty$，其中 $E(X_i) = \mu_i$，同时假定

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} E|X_i - \mu_i|^3}{(\sum_{i=1}^{n} \sigma_i^2)^{3/2}}$$

则当 $n \to \infty$ 标准化随机变量

$$Z_n = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i}{(\sum_{i=1}^{n} \sigma_i^2)^{1/2}} \to N(0,1)$$

当随机变量 $X_1, X_2, \cdots, X_n$ 之间存在一定程度上的相互依赖性时，CLT 仍然成立，生成的随机样本一般会出现这种相互依赖性.

Theorem 5.3 (Slutsky 定理)**.** 假设当 $n \to \infty$ 时，$X_n \xrightarrow{d} X$ 和 $C_n \xrightarrow{p} c$，其中 c 为常数. 则当 $n \to \infty$ 时，有

$$
\begin{aligned}
&(1) X_n + C_n \xrightarrow{d} c + X; \\
&(2) X_n - C_n \xrightarrow{d} X - c; \\
&(3) X_n C_n \xrightarrow{d} cX; \\
&(4)\ c \neq 0, \frac{X_n}{C_n} \xrightarrow{d} \frac{X}{c}
\end{aligned}
$$

Lemma 5.2 (Delta 方法)**.** 假设 $n \to \infty$，$\sqrt{n}(\overline{X_n} - \mu)/\sigma \xrightarrow{d} N(0,1)$ ，同时 $g(\cdot)$ 为连续可导函数且 $g'(\mu) \neq 0$. 则当 $n \to \infty$ 时候，有

$$\sqrt{n}[g(\bar{X_n}) - g(\mu)] \xrightarrow{d} N(0, \sigma^2[g'(\mu)]^2)$$

Lemma 5.3. 假设 $n \to \infty$，有 $\sqrt{n}(\overline{X_n} - \mu)/\sigma \xrightarrow{d} N(0,1)$，同时 $g(\cdot)$ 为二次可导函数，并满足 $g'(\mu) = 0, g''(\mu) \neq 0$，则当 $n \to \infty$ 时候：

$$\sqrt{n}[g(\bar{X}_n) - g(\mu)]/\sigma^2 \xrightarrow{d} \frac{g''(\mu)}{2}\chi_1^2$$

6 参数估计

假设概率分布族 $\mathbb{F}$ 包含了生成观测数据 $\mathbf{x}^n$ 的所有未知总体分布．即

$$f_X(x) = f(x, \theta_0)$$

对几乎所有的 $x \in \Omega$ 都成立．则称 $\mathbb{F}$ 是一个正确设定．

Example 6.1. *Cox* 在 *1970* 年提出生存模型，考虑时间建模上，人们失业重新找到工作耗费的时间、癌症患者的生存时间、价格波动的时长间隔等．在久期分析中，人们常常因那些尚未结束的事情仍持续多久感兴趣．假设随机变量 T_i 是一个已经发生的经济事件的持续时间，其概率密度为 $f(t)$，概率分布函数为 $F(t)$，则生存函数为

$$S(t) = P(T_i > t) = 1 - F(t)$$

风险率为

$$\lambda = \lim_{\delta \to o^+} \frac{P(t < T_i \leq t + \delta)}{\delta} \tag{3}$$

直觉上，风险率 $\lambda(t)$ 是指事件持续了 t 时期并将在时间点 t 结束的瞬时概率．同样我们对于个体异质性可进行描述

$$\lambda_i(t) = exp\{X_i'\theta\lambda_o(t)\}$$

其中，λ_0 为基准风险函数．在给定 X_i 时 T_i 的条件概率密度函数

$$f_i(t) = \lambda_i(t)S_i(t)$$

其生存函数

$$S_i(t) = e^{-\int_0^t \lambda_i(s)ds}$$

当然进一步会去思考如何得到正确的参数估计．极大似然是一个常用的方法，是由 R.A.Fisher 提出．

Definition 6.1. 一个似然函数是由给定的数据集，随机样本 $\boldsymbol{X}^n$ 的联合 *PMF/PDF* 作为 θ 的函数，

$$\hat{L}(\theta|\boldsymbol{x}^n) = f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \theta)$$

称随机样本 $\boldsymbol{X}^n$ 在取值为观测数据 $\boldsymbol{x}^n$ 下的似然函数. $\ln \hat{L}(\theta|\boldsymbol{x})$ 为对数似然函数 *(log-likehood function)*

Definition 6.2 (极大似然估计量)**.** 令 Θ 为有限维参数空间，假设统计量 $\hat{\theta} = \hat{\theta}_n(\boldsymbol{X}^n)$ 在 $\theta \in \Theta$ 上最大化 $\hat{L}(\theta|\boldsymbol{X}^n)$ 即

$$\hat{\theta} \equiv \hat{\theta}_n(\boldsymbol{X}^n) = \arg\max_{\theta\in\Theta} \hat{L}(\theta|\boldsymbol{X}^n)$$

一般不同的数据集会有不同的估计.

Theorem 6.1 (MLE 存在性)**.** 假设 $\hat{L}(\theta|\boldsymbol{X}^n)$ 为 $\theta \in \Theta$ 的连续函数的概率为 *1*，参数空间是紧集. 则存在以下的全局最优解 $\hat{\theta}$ 即

$$\hat{\theta} \equiv \hat{\theta}_n(\boldsymbol{X})^n = \arg\max_{\theta\in\Theta} \hat{L}(\theta|\boldsymbol{X}^n)$$

FOC 仅为最大化的必要条件而非充分条件. 仅仅为 MLE 提供了可能的候选参数，也就是在局部是可能最大的，全局最小，但全局是不一定最大. 需要进一步检验二阶条件:

$$\hat{H}(\theta) = \frac{\partial^2 \ln \hat{L}(\theta|\mathbf{X}^n)}{\partial\theta\partial\theta'}$$

对所有的 $\theta \in \Theta$ 是负定的，则 $\hat{\theta}$ 是全局最大解. 许多情况下无法推出，因此需要转变为推 $\hat{H}(\theta)$ 为负定相对容易.

Theorem 6.2 (极大似然估计的不变性)**.** 假设 $\hat{\theta}$ 为 $\theta \in \Theta$ 的 *MLE* 估计量，$g(\cdot)$ 为参数空间上的一一映射. 则 $g(\hat{\theta})$ 是 $g(\theta)$ 的 *MLE* 估计.

Theorem 6.3. 假设随机样本 $\boldsymbol{X}^n$ 的似然函数的 $f_{\boldsymbol{X}^n}(\boldsymbol{X}^n, \theta)$，且 $T(\boldsymbol{X}^n)$ 是 θ 的充分统计量. 则最大化随机样本的似然函数的 *MLE* 估计量 $\hat{\theta}$ 也是最大充分统计量的似然函数的 *MLE* 估计量.

6.1 极大似然的渐近性质

- $\mathbf{X}^n = (X_1, X_2, \cdots, X_n)$ 是来自于未知分布的 IID 随机样本的总体分布 $f(x)$
- 参数空间 Θ 为有界闭集，或等价于 Θ 为紧集.

- 参数值 θ_0 是 $E[\ln f(X_i,\theta)]$ 的唯一最优解.
- θ_0 是参数空间的内点.
- 对于每一个内点来说，关于 θ 的二阶可导.

为便于分析假设参数为标量.

Theorem 6.4. *若假设成立，且 $\hat{\theta} = \arg\max_{\theta\in\Theta} \sum_{i=1}^{n} \ln f(X_i,\theta)$，$n \to \infty$ 时，则几乎处处都有：*

$$\hat{\theta} \to \theta_0$$

Theorem 6.5. *假设 PDF 模型 $f(x;\theta)$ 对关于 $\theta \in \Theta$ 二次连续可导，其中 θ 是参数空间的内点.*

$$I(\theta) = \int_{-infty}^{+\infty} [\frac{\partial \ln f(x;\theta)}{\partial \theta}]^2 f(x,\theta)dx$$
$$H(\theta) = \int_{-infty}^{+\infty} [\frac{\partial^2 \ln f(x;\theta)}{\partial \theta^2}] f(x,\theta)dx$$

则对于所有 Θ 内点 θ

$$I(\theta) + H(\theta) = 0$$

同样 PMF 模型也有类似的结论.

6.2 矩估计与广义矩方法

矩估计 (MME) 是统计学中较为古老的参数估计方法. 通过对总体分布的若干阶矩与其相对应的样本矩进行匹配，获得一定数量的匹配方程以求得总体分布的未知参数值.

首先，假设 $f(x,\theta)$ 为未知总体分布 $f_X(x)$ 的 PMF/PDF 模型

$$\hat{m} = \hat{m}_n(\mathbf{X}^n)$$

其数学期望为

$$\begin{aligned} M(\theta) &= E_\theta \\ &= \int_{\mathbb{R}^n} \hat{m}_n(\mathbf{x}^n) f_{\mathbf{X}^n}(\mathbf{x}^n,\theta) d\mathbf{x}^n \end{aligned}$$

$M_k(\theta)$ 的更一般形式为

$$\begin{cases} \int_{-\infty}^{\infty} x^k f(x,\theta)dx, \text{X is continuity} \\ \sum_{x\in\Omega_X} x^k f(x,\theta), \text{X is discrete} \end{cases}$$

其次，求解方程组

$$\hat{m} = M(\hat{\theta}) = n^{-1}\sum_{i=1}^{n} X_i^k, k = 1, 2, \cdots$$

展开形式为

$$\begin{cases} \hat{m}_1 = M_1(\hat{\theta}) \\ \hat{m}_2 = M_2(\hat{\theta}) \\ \qquad \cdots \\ \hat{m}_p = M_p(\hat{\theta}) \end{cases}$$

即选择参数值 $\hat{\theta}$ 对样本和总体矩的估计，求得的 $\hat{\theta} = \hat{\theta_n}(\mathbf{X}^n)$ 称为真实参数值 θ_0 的矩估计量 MME.

6.2.1 广义矩估计方法

在实际中，我们很难真正能够掌握所有的矩，因为总体分布未知. 经济学往往会使用一组矩条件刻画经济理论或经济假说，用有经济理论的矩条件来估计真实的模型的参数值.

$$E[m(\mathbf{X}, \theta_0)] = 0$$

其中 $E(\cdot)$ 是关于随机变量 X 的概率分布的数学期望，这些矩可能来自于经济理论，且 $q \geq p$

给定 q 个总体矩条件:

$$E[m(X_i, \theta_0)] = 0$$

其中 $m(\mathbf{X}_i, \theta)$ 是 $q \times 1$ 维的随机向量.

Theorem 6.6 (GMM 估计量的存在性)**.** *假设二次型 $\hat{m}(\theta)'\hat{W}^{-1}\hat{m}(\theta)$ 在 $\theta \in \Theta$ 上连续概率为 1，且参数空间 Θ 是一个紧集，则存在满足一下的最小化问题的最优解.*

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \hat{m}(\theta)'\hat{W}^{-1}\hat{m}(\theta)$$

6.3 最优无偏估计

对于一个参数来说我们想要去得到最好的估计，而证明去衡量“最好”这件事？这里就定义了一个 OLS 常用的均方误准则

Definition 6.3.

$$MSE_\theta(\hat{\theta}) = E_\theta(\hat{\theta} - \theta)^2$$

其度量的是一个未知参数 θ 与估计 $\hat{\theta}$ 量之间的总体偏离程度．MSE 越小估计越好，在均方误下的 θ 最优估计就是 MSE 最小的那个．

Definition 6.4.

$$Bias_\theta(\hat{\theta}) = E_\theta(\hat{\theta}) - \theta$$

若 $Bias$=0，称估计量 $\hat{\theta}$ 为 θ 的无偏估计量．

一个无偏估计量从统计意义上给出了一个正确估计．即不存在任何的向上和向下偏差．

Theorem 6.7.

$$E(\hat{\theta} - \theta)^2 = var(\hat{\theta}) + [Bias(\hat{\theta})]^2$$

证明.

$$E(\hat{\theta}-\theta)^2 = E[\hat{\theta}-E(\hat{\theta})+E(\hat{\theta})-\theta]^2 = E[\hat{\theta}-E(\hat{\theta})]^2+[E(\hat{\theta})-\theta]^2+2E\{[\hat{\theta}-E(\hat{\theta})][E(\hat{\theta})-\theta]\}$$

其中交叉项：

$$E\{[\hat{\theta} - E(\hat{\theta})][E(\hat{\theta}) - \theta]\} = 0$$

故成立． □

由此 MSE 的两部分，一部分由 var 解释，是因抽样变化所导致的波动性 (variability)，Bias 测度了关于估计方法的估计精度 (accuracy)．对于任何的无偏估计量，$MSE = E(\hat{\theta}-\theta)^2 = var(\theta)$，因此**最优**无偏估计量是方差最小的无偏估计量．（无偏性是由偏差所控制了，在这里就是一个前提）

Definition 6.5 (相对有效性)**.**

$$MSE(\hat{\theta}) \leq MSE(\widetilde{\theta})$$

Definition 6.6 (Generalized Unbiased Estimation)**.** *$\hat{\tau} = \hat{\tau}_n(\boldsymbol{X}^n)$ 是 $\tau(\theta)$ 的无偏估计量，若*

$$E_\theta(\hat{\tau}) = \tau(\theta), \text{对所有的 } \theta \in \Theta$$

Definition 6.7. 一致最优无偏估计 Γ 为参数定义 $\tau(\theta)$ 的一类无偏估计量的集合，其中 $\theta \in \Theta$. 若估计量满足对于所有的 $\theta \in \Theta, E_\theta(\hat{\tau}^*) = \tau(\theta)$; 对 Γ 中对 $\tau(\theta)$ 中的任意估计量都有 $var(\hat{\tau}^*) \leq var(\hat{\tau})$ 成立，也就是最小的. 则认为估计量 $\hat{\tau^*}$ 是一致最优估计无偏.

6.4 Cramer-Rao 下界

Theorem 6.8 (Gramer-Rao 下界)**.** 令 X^n 为一个随机样本, 联合 PMF/PDF 为 $f_{X^n}(x^n, \theta)$, 令 $\hat{\tau} = \hat{\tau_n} X^n$ 为参数 $\tau(\theta)$ 的任意估计量，且 $E_\theta \hat{\tau}$ 是 θ 的可导函数. 期望 $E_\theta(\cdot)$ 是定义在随机样本 X^n 的联合分布. 则对满足 $E_\theta(\cdot) < \infty$ 的任意函数 $h : \mathbb{R}^n \to \mathbb{R}$, 若以下的条件成立，则对于所有的 $n > 0$ 和所有的 $\theta < \Theta$,

7 假设检验

Definition 7.1. 假设检验是一种统计决策，它设定了对于什么样的样本值，拒绝 H_0 接受 H_A 为真.

Definition 7.2. 因此需要考虑如何构建一个临界值或拒绝域: 随机样本 $\boldsymbol{X}^n$ 的样本空间中的那些将拒绝 $\mathbb{H}_0$ 的样本点集合称为拒绝域或临界域. 拒绝域的补集称为接受域. 接受域表示为:

$$\mathbb{A}_n(c) = \{x^n : T(x^n) \leq c\}$$

拒绝域为:

$$\mathbb{C}_n(c) = \{x^n : T(x^n) > c\}$$

c 为分界点，表示临界值.

Definition 7.3. 检验功效: 若 $\mathbb{C}$ 是原假设 $\mathbb{H}_0 : \theta \in \Theta$ 检验的拒绝域. 则函数 $\pi(\theta) = P_0(\theta)(\boldsymbol{X}^n \in C)$ 称为拒绝原假设 $\mathbb{H}_0$ 的检验功效.

检验功效是关于拒绝 $\mathbb{H}_0$ 的概率. 比如在正态分布检验时候，对于正态分布统计量的检验 $T(\mathbf{X}^n) = \sqrt{n}(\overline{X}_n - \mu_0)/\sigma$ 的功效可表示为

$$\pi(\mu) = P(\left|\frac{\overline{X}_n - \mu_0}{\sigma\sqrt{n}}\right| > z_{\alpha/2})$$

对于假设检验来说，是一个犯错误的概率来评价与比较. 而两类错误通常是负相关的，通常情况下以固定第一类错误的概率来获得在第二类错误发生概率的降低. 进一步的，我们可得到若能够满足这样的统计量，我们称为一致的统计检验有最大功效.

Definition 7.4. 令 $\mathbb{T}$ 为一族关于 $\mathbb{H}_0 : \theta \in \Theta$ 和 $\mathbb{H}_A : \theta \in \Theta$ 的检验集合.

7.1 Neyman-Pearson 引理

Theorem 7.1. 考虑一个简单的原假设 $\mathbb{H}_0 : \theta = \theta_0$ 和一个简单的备择假设 $H_A : \theta = \theta_1$. 其中对应的 $\theta_i(i = 1, 2)$ 的随机样本 $\boldsymbol{X}^n$ 的 PMF/PDF 为 $f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \theta_i)$ 给定某个参数 $c \geq 0$ 则定义一个检验的拒绝域 $\mathbb{C}_n(c)$ 和接受域 $\mathbb{A}_n(c)$ 分别为

$$\mathbb{C}_n(c) = \{\boldsymbol{x}^n : \frac{f_{\boldsymbol{X}^n}(\boldsymbol{X}^n, \theta_1)}{f_{\boldsymbol{X}^n}(\boldsymbol{X}^n, \theta_0)} > c\}$$

$$\mathbb{A}_n(c) = \{\boldsymbol{x}^n : \frac{f_{\boldsymbol{X}^n}(\boldsymbol{X}^n, \theta_1)}{f_{\boldsymbol{X}^n}(\boldsymbol{X}^n, \theta_0)} \leq c\}$$

且:

$$P[\boldsymbol{X}^n \in \mathbb{C}_n(c) | \mathbb{H}_0] = \alpha$$

充分性：满足上述条件的任意检验为 α 的一致最大功效检验. 必要性：若存在一个检验当 $c > 0$ 时候满足上述检验，这对于每一个水平的 α 的一致最大功效检验.

Lemma 7.1. 假设 $T(\boldsymbol{X}^n)$ 是 θ 的充分统计量，$g(t, \theta_i)$ 为对应于 θ_i 的 PMF/PDF.

7.2 Wald 检验

$g(\theta)$ 为一个线性向量值函数

$$g(\theta) = R\theta - r$$

其中 $R = J \times p$ 为已知常数矩阵，r 是 $J \times 1$ 是已知常数向量. 原假设

$$\mathbb{H}_0 : R\theta = r$$

之后的概率模型是建立在已知概率正确的基础上.

一组正则条件:

- $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V)$，其中 V 为 $p \times p$ 对称有界非奇异矩阵. 真实参数为内点.
- 当 $n \to \infty$ 时，$\hat{V} \xrightarrow{d} V$
- $g : \mathbb{R}^p \to \mathbb{R}^J$，是 $\theta \in \Theta$ 的连续可导函数. 其中 $J \leq p$