高等数理统计复习

第一章

- 样本空间及相关概念
    - **随机试验**：受偶然性因素影响，导致结果不正确的试验；
    - **个体**：试验的结果；
    - **样本空间**：所有个体组成的集合；
    - **样本**：多次（n次）试验的结果构成的一个向量；
        - 样本有两重性，在抽样之前是随机的，抽样之后表现为一组数。
    - **简单随机样本**：试验在相同的条件下独立进行得到的样本；
    - **样本量**：试验重复的次数n；
    - **事件**：样本空间的子集；
    - **总体分布**：称随机抽样得到的一个样本点的分布为总体分布（简称为总体），这是感兴趣的对象。
    - **样本分布**：样本的概率分布称为样本分布；简单随机样本的分布由总体分布及样本量确定。
- 测度与积分
    - **定义1.1**：设全集为$\mathcal{X}$, $\mathcal{A}$为其一些子集构成的集合，称为$\sigma-$域，若：
        1. $\mathcal{X} \in \mathcal{A}$;
        2. $A \in \mathcal{A}$, 则 $A^c \in \mathcal{A}$（关于补运算封闭）；
        3. 对至多可数集列$\{A_n\} \subset \mathcal{A}$，有$\cup_n A_n \in \mathcal{A}$（关于可列并封闭）。

  此时称二元组$(\mathcal{A}, \mathcal{X})$为**可测空间**，$\mathcal{A}$中的元素称为**可测集**。
    - **定义1.2**：对可测空间$(\mathcal{A}, \mathcal{X})$，定义在$\mathcal{A}$上的非负函数称为**测度**，若对任意可数两两不交集列$\{A_n\} \subset \mathcal{A}$, $\mu(\cup_n A_n) = \sum_n \mu(A_n)$（称为$\sigma-$可加性)。此时三元组$(\mathcal{A}, \mathcal{X}, \mu)$称为测度空间。如果$\mathcal{X}$能被可数个$\sigma-$测度有限的事件$A_n$所覆盖，则称$\mu$为$\sigma-$有限的。
    - **定义1.3**：给定一个测度空间$(\mathcal{X}, \mathcal{A}, \mu)$，定义于$\mathcal{X}$取值为$\mathbb{R}$上的函数$f$称为可测函数，若直线上的任意$\mathrm{Borel}$集$B$在$f$之下的原象是$\mathcal{A}$可测的：

$$\{x \mid f(x) \in B\} =: f^{-1}(B) \in \mathcal{A}.$$

    - **定义1.5**：设υ和μ是可测空间$(\mathcal{X}, \mathcal{A})$上的两个测度，称$\upsilon$关于$\mu$绝对连续（也称$\upsilon$受控于$\mu$），如果对$\forall A \in \mathcal{A}$, $\mu(A) = 0$蕴含着$\upsilon(A) = 0$，此时记$\upsilon \ll \mu$。
    - **定理1.1**：（Radon-Nikodym定理）设μ 和 υ是两个σ有限测度，则$\upsilon \ll \mu$当且仅当存在一个处处有限非负$\mathcal{A}$可测函数f，使得

$$\upsilon(A) = \int_A f(\omega) d\mu(\omega), \ \forall A \in \mathcal{A}.$$

如果不计在一个μ零测集上的差异，则f由上式唯一确定，称其为$R-N$导数，记为$dv/d\mu$。

- **定义1.8**：统计模型$(\Omega,\mathcal{A},F_\theta),\theta\in\Theta$称为参数可识别的，若$F_{\theta_1}=F_{\theta_2}\Rightarrow\theta_1=\theta_2$。（参数若不可识别，对参数空间进行约束来使得可识别。）
- 给定一个统计模型$(\Omega,\mathcal{A},F_\theta),\theta\in\Theta$，若$T$为样本$X_1,\ldots,X_n$的取值空间$\Omega$到其值域空间（通常为欧式空间）上的不依赖于$\theta$的映射，则称$T$为统计量（这里的意思是这个$T$的分布可以是依赖$\theta$的）。
- **定义1.9**：统计量$T=T(X)$称为对θ是辅助统计量，若其分布与θ无关。
- 指数族与群族
  - **定义1.10**：称$\mathcal{P}$为一个k参数指数族，若从该总体抽得的i,i.d.样本$X_1,\ldots,X_n$的联合密度（相对于测度μ）有形式

    $$f_\theta(x)=h(x)\exp\left[\sum_{i=1}^{k}c_i(\theta)T_i(x)-d(\theta)\right]$$

    其中x为$n\times1$向量，$h(x)$为非负可测函数，$c_i(\theta)$和$d(\theta)$均为标量，$k\geq1$。
  - **定义1.11**：对于概率空间$(\Omega,\mathcal{A},P)$，集合$\{x\mid\forall A\in\mathcal{A}\text{为包含x的开矩形},P(A)>0\}$称为概率测度$P$的支撑。
  - **定义1.12**：对一族概率测度$\mathcal{P}=\{P_\theta,\theta\in\Theta\}$，若$P_\theta$的公共支撑$A_\theta$不依赖于$\theta$（记公共值为$A$），则称$\mathcal{P}$有公共的支撑$A$。
    - 注 设点x为随机变量X的一个支撑点，则X取值于x的任何一个领域的概率大于0:$\forall\epsilon>0,P(x-\epsilon<X<x+\epsilon)>0$。因此若$X$有逐段连续的密度函数$f(x)$，可证明满足$f(x)>0$的$x$都是支撑点，且支撑为$\{x:f(x)>0\}$的闭包。
    - 对于定义1.10中的指数族，注意到

      $$\{x\mid f_\theta(x)>0\}=\{x:h(x)>0\}$$

      与θ无关，故有注知指数族有公共的支撑。
    - 注 有公共支撑的分布族也不一定是指数族。
  - 为方便起见，常将指数族分布重新参数化：$\omega_i=c_i(\theta),i=1,\ldots,k$，得到密度的点则化形式：

    $$f_\omega(x)=h(x)\exp\left[\sum_{i=1}^{k}\omega_iT_i(x)-d(w)\right].$$

    典则化形式并不唯一，事实上，对ω进行一一变换后所得密度函数也有典则化形式。

  - **定义1.13**：注意到$f_\omega(x)$为密度当且仅当

$$0<\int h(x)\exp\left[\sum_{i=1}^k \omega_i T_i(x)\right]d\mu(x)<\infty.$$

满足这种情形的点ω的全体$\mathcal{W}$称为自然参数空间，ω称为自然参数。
  - 对指数族分布有

$$ET_i=\frac{\partial d(\omega)}{\partial \omega_i},\operatorname{cov}(T_i,T_j)=\frac{\partial^2 d(\omega)}{\partial\omega_i\partial\omega_j}.$$

- **定义1.14**：统计量$T=T(X)$对θ称为充分统计量，若条件分布$X\mid T$与θ无关。
  - 充分统计量不唯一。设T为充分统计量，g为任一已知的一一映射，则$g(T)$ 也是一个充分统计量；
  - 设T是一个充分统计量，则对任何统计量$S=S(X)$，条件分布$S\mid T=t$也与θ无关。
- **定理1.3**：（因子分解定理）设分布族$P_\theta,\theta\in\Theta$受控于μ，记$f_\theta(x)=\frac{dP_\theta}{d\mu}(x)$，则$T=T(X)$对$\theta$是充分统计量$\Leftrightarrow$ $f_\theta(x)=g_\theta[T(X)]h(x),a.s.\ \mu$，其中$h(x)$为非负可测函数。（证明见pptP38或者书上找）
- **定义1.15**：一个充分统计量T称为极小的（minimal）若对任一充分统计量S，存在g使得$T=g(S),a.s.$。
  - 极小充分统计量在所有充分统计量中最大限度的简化了数据。
- **定理1.4**：设$\{F_\theta,\theta\in\Theta\}$是一个具有公共支撑的分布族,$\{F_\theta,\theta^{'},\theta^{'}\in\Theta\}$为子族，即$\Theta^{'}\subset\Theta$。若统计量$T$关于$F_{\theta^{'}}$是极小充分的同时关于$F_\theta$是充分的，则关于$F_\theta$也是极小充分的。
- **定理1.5**：设$\mathcal{F}$是有限个具有密度$f_{\theta_i}(i=0,1,\ldots,k)$的分布构成的族，其具有公共支撑。
  - 统计量$S=S(X)$是充分统计量$\Leftrightarrow$对任何固定的θ_0和θ，$\frac{f_\theta(X)}{f_{\theta_0}(X)}$是$S(X)$的函数；
  - 统计量$T(X)=\left(\frac{f_{\theta_1}(X)}{f_{\theta_0}(X)},\frac{f_{\theta_2}(X)}{f_{\theta_0}(X)},\cdots,\frac{f_{\theta_k}(X)}{f_{\theta_0}(X)}\right)^T$是极小充分统计量。
- **推论1.1**：设X的分布具有指数形式

$$f_W(x)=h(x)\exp\left[\sum_{i=1}^k\omega_iT_i(x)-d(\omega)\right].$$

则当满足下列条件之一时，$T=T(X)=(T_1,T_2,\cdots,T_k)^{'}$是极小充分统计量。
  - $T_1,T_2,\cdots,T_k$线性无关（也等价于自然参数空间$\mathcal{W}$有内点）；
  - 参数空间含有$k+1$个点$\omega^{(0)},\omega^{(1)},\cdots,\omega^{(k)}$，它们张成欧氏空间$\mathbb{R}^k$。
- **推论1.6**：设分布族$\{P_\theta,\theta\in\Theta\}$受控于$\mu$，记

$$f_\theta(x)=\frac{dP_\theta}{d\mu}(x),$$

若

$$\frac{f_\theta(x)}{f_\theta(y)}\text{与}\theta\text{无关}\Leftrightarrow T(x)=T(y),$$

则$T=T(x)$是极小充分统计量。

- **定义1.16**：给定一个统计模型$\{(\Omega,\mathcal{A},P_\theta),\theta\in\Theta\}$。$T=T(X)$称为完全统计量，若对任何满足条件$E_\theta g(T)=0,\forall\theta\in\Theta$ 的 g 必有 $P_\theta(g(T)=0)=1,\forall\theta\in\Theta$（或者$g(T)=0,a.s.P_\theta,\forall\theta\in\Theta$）。
    - 若上述命题对任何有界g成立，则称T为有界完全统计量。
- **定理1.7**：（Basu）若T是关于分布族$\{P_\theta,\theta\in\Theta\}$的完全充分统计量，则任一辅助统计量$S$都与$T$独立。（证明在ppt P50）
- **定理1.8**：若极小充分统计量存在，则一个完全充分统计量必定是极小的。（证明在ppt P51）
    - 其逆定理不存在，即便是极小充分统计量存在，它不一定是完全统计量。
- **定理1.9**：设X的分布具有指数族形式

$$f_\omega(x)=h(x)\exp\left[\sum_{i=1}^{k}\omega_iT_i(x)-d(\omega)\right].$$

且自然参数空间$\mathcal{W}$有内点，则$T=(T_1,T_2,\cdots,T_k)^T$是完全的。

- 随机变量序列的收敛性
    - 若随机变量序列$\{X_n\}$**依概率有界**，则记为$X_n=O_p(1)$；
    - **依概率有界**：对$\forall\varepsilon>0,\exists M_\varepsilon$使得$P(|X_n|\geq M_\varepsilon)\leq\varepsilon$。
    - 若随机变量序列$\{X_n\}$**依概率收敛**，则记为$X_n=o_p(1)$；
    - **依概率收敛**：$\forall\varepsilon>0,\lim_{n\to\infty}P(|X_n-X|\geq\varepsilon)=0$。
    - $X_n=O_p(Y_n)$当且仅当$X_n/Y_n=O_p(1)$；
    - $X_n=o_p(Y_n)$当且仅当$X_n/Y_n=o_p(1)$；
    - $X_n=o(Y_n)$当且仅当$X_n/Y_n=o(1)$。
    - 若$X_n\xrightarrow{d.f.}X$，则$X_n=O_p(1)$。
    - $O_p(1)\times o_p(1)=o_p(1)$。
- **定理1.10**：（Slutsky's Theorem）设随机变量序列$\{X_n\},\{Y_n\}$满足$X_n\xrightarrow{d.f.}X,Y_n\xrightarrow{Pr.}c$(常数)，则
    - $X_n+Y_n\xrightarrow{d.f.}X+c$;

- $X_nY_n \xrightarrow{d.f.} cX$；
- $\frac{X_n}{Y_n} \xrightarrow{d.f.} \frac{X}{c}, (c \neq 0)$。

- 设随机变量序列$X_n \xrightarrow{Pr.} c$，函数g在c处连续，则$g(X_n) \xrightarrow{Pr.} g(c)$。
- Wald统计决策理论简介（P57-P69）
  - 可容许性：设两个决策函数δ_0, δ满足

$$R_\theta(\delta_0) \leq R_\theta(\delta), \forall \theta \in \Theta,$$

    则称δ_0一致的优于δ。若至少对某个θ成立严格不等式，则称δ_0严格一致的优于δ，此时称δ是**不可容许的**。若不存在这样的δ_0，则称δ是**可容许的**。
- 秩序统计量及有关分布 (P70)
  - $X_{(i)}$的分布和密度

$$F_i(x) = \frac{n!}{(i-1)!(n-i)!}\int_0^{F(x)} t^{i-1}(1-t)^{n-i}dt,$$

    密度为

$$f_i(x) = \frac{n!}{(i-1)!(n-i)!}[F(x)]^{i-1}[1-F(x)]^{n-i}f(x).$$

  - $(X_{(i)}, X_{(j)})$的联合密度$(1 \leq i < j \leq n)$当$u < v$时

$$\begin{aligned} f_{ij}(u,v) = & \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(u)]^{i-1} \\ & [F(v)-F(u)]^{j-i-1}[1-F(v)]^{n-j}f(u)f(v) \end{aligned}.$$

    当$u \geq v$时，$f_{ij}(u,v) = 0$。
  - **定理1.11**：设$0 < p_1 < p_2 < \cdots < p_k < 1, X_1, X_2, \cdots, X_n \sim F(x)$为i.i.d样本，密度$f(X)$在$p_i$处有连续的非零导数，则若$n_i = np_i + o(n^{1/2})$,$1 \leq i \leq k$，有

$$\sqrt{n}(X_{(n_1)} - \lambda_{p_1}, \cdots, X_{(n_k)} - \lambda_{p_k})^T \xrightarrow{d.f.} N(0, \Lambda),$$

    其中Λ为k阶对称方阵，其(i,j)元素为

$$\frac{p_i(1-p_j)}{f(\lambda_{p_i})f(\lambda_{p_j})}.$$

第二章

- **定义2.1**：估计$\hat{\theta}$的偏倚（bias）定义为

$$\mathrm{Bias}_\theta(\hat{\theta}) = E_\theta\hat{\theta} - \theta.$$

  一个估计称为无偏的（unbiased）若

$$\mathrm{Bias}_\theta(\hat{\theta}) = 0 \text{ 即} E_\theta\hat{\theta} = \theta.$$

  无偏估计不具有变换不变性，即$\hat{\theta}$为θ的无偏估计，通常$g(\hat{\theta})$不是$g(\theta)$的无偏估计，除非g是线性的。
- **定义2.2**：估计$\hat{\theta}$的平均绝对误差（mean absolute error，MAE）定义为

$$\mathrm{MAE}_\theta(\hat{\theta}) = E_\theta|\hat{\theta} - \theta|;$$

  均方误差（mean square error，MSE）定义为

$$\mathrm{MSE}_\theta(\hat{\theta}) = E_\theta(\hat{\theta} - \theta)^2.$$

  - $$\mathrm{MSE}_\theta(\hat{\theta}) = \mathrm{var}_\theta(\hat{\theta}) + [\mathrm{Bias}_\theta(\hat{\theta})]^2.$$
- **定义2.3**：估计序列$\{\hat{\theta}_n\}$称为θ的（弱）相合估计（consistent estimator），如果$\hat{\theta}_n$依概率收敛到θ，即

$$\lim_{n\to\infty} P_\theta(|\hat{\theta}_n - \theta| > \varepsilon) = 0, \quad \forall\varepsilon > 0, \theta \in \Theta.$$

  若$P_\theta(\lim_{n\to\infty}\hat{\theta}_n = \theta) = 1$，则称$\hat{\theta}_n$是$\theta$的强相合估计。
  相合性的意义：只要样本量充分大，就可以使估计误差在一定意义下充分地小；如果一个估计量没有相合性，则无论用多少样本估计都会产生系统误差。
- **定义2.4**：设$\hat{\theta}_n$为θ的估计，若存在$\sigma_n^2(\theta)$满足$\sigma_n^{-1}(\theta)(\hat{\theta}_n - \theta) \xrightarrow{d.f.} N(0,1)$，则称估计$\hat{\theta}_n$具有渐近正态性，记为

$$\hat{\theta}_n \sim \mathrm{AN}(\theta, \sigma_n^2(\theta)),$$

  $\sigma_n^2(\theta)$称为渐近方差。（若某个无偏估计的方差达到此下界，则此估计必为UMVUE）
- **替代原理**（P12-P25）

- 怎样估计F?
  - $\hat{F}_n(x)=\frac{1}{n}\sum_{i=1}^n I(X_i\le x)$,经验分布函数（P17）。
  - $\hat{\theta}=T(\hat{F}_n)=\frac{1}{n}\sum_{i=1}^n h(X_i)$。
  - 核密度估计
- 替代原理能否得到好的估计$\hat{\theta}$?
- 矩估计方法

- **一致最小方差无偏估计**
  - **定义2.5**：设T是参数$g(\theta)$的无偏估计，称T为一个UMVUE，如果对$g(\theta)$的任意无偏估计S满足如下不等式：

$$\mathrm{var}_\theta(T)\le \mathrm{var}_\theta(S),\forall\theta,$$

  - $g(\theta)$的无偏估计可能不存在
  - 例：设$X\sim B(n,p),0<p<1$，则 $g(p)=p^{-1}$的无偏估计不存在（证明在P27）。
  - 一些时候无偏估计存在但是没有任何意义（P28）
  - **定理2.1**：（Rao-Blackwell定理）设$T=T(X)$ 是 θ的充分统计量，令$S=S(X)$ 为 $g(\theta)$的无偏估计且对$\forall\theta\in\Theta$ 有 $\mathrm{var}_\theta(S)<\infty$，则$S^*=E(S\mid T)$ 是 $g(\theta)$的无偏估计，且对$\forall\theta\in\Theta$ 有 $\mathrm{var}_\theta(S^*)\le\mathrm{var}_\theta(S)$，“=” 成立 $\Leftrightarrow P_\theta(S^*=S)=1$。（证明P29）
    - 即说明只用在充分统计量中寻找UMVUE。
    - $$E_\theta S^*=E_\theta[E(S\mid T)]=E_\theta S=g(\theta),\forall\theta\in\Theta,$$
    - $$\begin{aligned}\mathrm{var}_\theta(S)=E_\theta(S-E_\theta S)^2&=E_\theta(S-E(S\mid T)+E(S\mid T)-E_\theta S)^2\\&=E_\theta(S-E(S\mid T))^2+E_\theta(E(S\mid T)-E_\theta S)^2\\&+2E_\theta[(S-E(S\mid T))(E(S\mid T)-E_\theta S)]\end{aligned}$$
    - $$\begin{aligned}&E_\theta[(S-E(S\mid T))(E(S\mid T)-E_\theta S)]\\=&E_\theta\left\{E_\theta(S-E(S\mid T))(E(S\mid T)-E_\theta S)\mid T\right\}\\=&E_\theta\left\{(E(S\mid T)-E_\theta S)(E_\theta((S-E(S\mid T))\mid T))\right\}\\=&0\end{aligned}$$

- **推论2.1**：设S 为 $g(\theta)$的无偏估计且对$\forall\theta\in\Theta$ 有 $\mathrm{var}_\theta(S)<\infty$，$T_1$和$T_2$都是充分统计量且$T_2=h(T_1)$，令$S_1^*=E(S\mid T_1),S_2^*=E(S\mid T_2)$，则对$\forall\theta\in\Theta$有$\mathrm{var}_\theta(S_1^*)\ge\mathrm{var}_\theta(S_2^*)$。
  - 若极小充分统计量存在，则θ的UMVUE必有形式$E(S\mid T)$。
  - 不同的无偏估计对应的条件期望不一定都是UMVUE。（P31）
- **定理2.2**：（Lehmann-Scheffe定理）设T是充分完全统计量，S为$g(\theta)$的无偏估计且对$\forall\theta\in\Theta$有$\mathrm{var}_\theta(S)<\infty$。令$S^*=E(S\mid T)$，则$S^*$为UMVUE。$S^*$ 和 T^*为UMVUE$\Rightarrow P_\theta(S^*=T^*)=1,\theta\in\Theta$，即UMVUE是以概率1地唯一的。（证明P32）
  - 解方程（找函数h使得$E_\theta h(T)=g(\theta)$）

- 条件化（先找一个无偏估计S）

- **定理2.3**：设$T=T(X)$为$g(\theta)$的无偏估计且对$\forall\theta\in\Theta$有$\mathrm{var}_\theta(T)<\infty$。令$\mathcal{U}$为所有方差有限的0的无偏估计，则$T$为UMVUE的充分必要条件是对$\forall U\in\mathcal{U}$有$E_\theta[T(X)U(X)]=0,\forall\theta\in\Theta$。（证明P35）
- 对$\forall U$如$EU^2<\infty$，则

$$\mathrm{var}_\theta(S)\geq\frac{[\mathrm{cov}_\theta(S,U)]^2}{\mathrm{var}_\theta(U)}.$$

- **定理2.4**：（Cramer-Rao下界）设$X\sim f(x,\theta)$，统计量T为$g(\theta)$的无偏估计，对$\forall\theta\in\Theta$有$\mathrm{var}_\theta(T)<\infty,g(\theta)$可导。若满足如下两个条件：
  - 分布支撑$A=\{x\mid f(x,\theta)>0\}$不依赖于$\theta$；
  - 对$\forall x\in A,f(x,\theta)$ 关于 θ可导，且导数

$$\frac{\partial}{\partial\theta}\int f(x,\theta)dx \text{ 和 } \frac{\partial}{\partial\theta}\int T(x)f(x,\theta)dx$$

    中求导与积分可交换顺序。

  则$\mathrm{var}_\theta(T)\geq I^{-1}(\theta)[g^{'}(\theta)]^2$，其中$I(\theta)=E_\theta(\frac{\partial}{\partial\theta}\log f(X,\theta))^2$称为**Fisher信息量**。（证明P38）（注意这里不是对样本，而是对单一的X）
- **定理2.5**：（证明P39）能取到C-R下界当且仅当分布属于单参数指数族。（T为其中充分统计量）。
- 随机样本X，若其方差$\mathrm{var}(X)=\sigma^2$存在，则称σ为该总体的标准差；而一个估计量$\hat{\theta}$的标准误。
- δ法（delta method）：设随机变量序列X_n满足

$$\sqrt{n}(X_n-\theta)\xrightarrow{d.f.}N(0,\sigma^2)$$

  则对任意函数g，如满足$g'(\theta)$存在且g'在θ处连续，则有

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

- 极大似然估计
  - 若$T=T(X)$是θ的充分统计量且θ的MLE$\hat{\theta}$存在唯一，则$\hat{\theta}$为T的函数；
  - MLE具有某种不变性。若$\phi=g(\theta)$为Θ上的严格增函数（推广后任意函数皆可），则若$\hat{\theta}$为θ的MLE，$g(\hat{\theta})$必为ϕ的MLE。
  - **求解MLE的方法**
    - 直接极大法
    - 似然方程法：$s(\theta)=\frac{\partial l(\theta)}{\partial\theta}$称为得分函数。似然方程即$s(\theta)=0$。
  - Jensen不等式：g严格凸函数，EX存在，则$Eg(X)\geq g(EX)$。

- 以下总假定$X_1, X_2, \cdots, X_n \sim f(x,\theta), l(\theta)=\sum_{i=1}^n \log f(X_i,\theta)$为对数似然，假定条件
  - Θ为开集；
  - $A=\{x: f(x,\theta)>0\}$不依赖于θ；
  - 对固定$x\in A, f(x,\theta)$关于θ有三阶连续导数；
  - $\forall\theta\in\Theta$有$E_\theta[\frac{\partial}{\partial\theta}\log f(X_i,\theta)]=0$，记

    $$I(\theta)=\operatorname{var}_\theta\left(\frac{\partial}{\partial\theta}\log f(X_i,\theta)\right)$$

    假设$0<I(\theta)<\infty$；
  - $\forall\theta\in\Theta$，记$J(\theta)=-E_\theta[\frac{\partial^2}{\partial\theta^2}\log f(X_i,\theta)]$，假设$0<J(\theta)<\infty$；
  - 对任意给定θ和$\delta>0, \exists M(x)$，使得当$|t-\theta|<\delta$时$|\frac{\partial^3}{\partial\theta^3}\log f(x,\theta)|_{\theta=t}|\le M(x)$且$E_\theta M(X_i)<\infty$。
  - 若假定$\int f(x,\theta)dx$关于θ的偏导数可在积分下二次求导，则$I(\theta)=J(\theta)$。
  - **定理2.5（P62）**：设条件A1，A2满足，$f(x,\theta)$关于θ可导，则$n\to\infty$时，似然方程$\sum_{i=1}^n\frac{\partial\log f(X_i,\theta)}{\partial\theta}=0$以概率1的存在一个解$\hat\theta_n$，此解$\hat\theta_n$为$\theta$的相合估计。此外若条件A1-A6成立，则似然方程任一相合解序列$\hat\theta_n$满足$\sqrt{n}(\hat\theta_n-\theta)\xrightarrow{d.f.}N(0,\frac{I(\theta)}{J^2(\theta)})$。
- MLE的标准差
  - 若$I(\theta)$有显式表达式，则用$\hat\theta_n$代替θ，这样得到$I(\theta)$的估计为$I(\hat\theta_n)$，从而$\hat\theta_n$的标准差估计为

    $$\widehat{\mathrm{SE}}(\hat\theta_n)=[nI(\hat\theta_n)]^{-\frac{1}{2}}.$$

    称$nI(\hat\theta_n)$为θ的期望Fisher信息。
  - 由于

    $$I(\theta)=-E_\theta\frac{\partial^2\log f(X_i,\theta)}{\partial\theta^2},$$

    故可用

    $$\hat I(\hat\theta_n)=-\frac{1}{n}\sum_{i=1}^n\frac{\partial^2\log f(X_i,\theta)}{\partial\theta^2}\bigg|_{\theta=\hat\theta_n}=-\frac{1}{n}\frac{\partial^2 l(\theta)}{\partial\theta^2}\bigg|_{\theta=\hat\theta n}$$

    来估计$I(\theta)$。从而$\hat\theta_n$的标准差估计为

$$\widehat{\mathrm{SE}}(\hat{\theta}_n) = \left(-\frac{\partial^2 l(\theta)}{\partial \theta^2}\bigg|_{\theta=\hat{\theta}n} \right)^{-\frac{1}{2}}$$

称$n\hat{I}(\theta_n) = -\frac{\partial^2 l(\theta)}{\partial \theta^2}\Big|_{\theta=\hat{\theta}n}$为$\theta$的观测Fisher信息。

- 多维情形类似（P68-P75）
- **MLEs数值计算**
  - **Newton-Raphson算法**
    - 记负对数似然函数的Hessian矩阵为$H(\theta) = -\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta'} = -\frac{\partial s(\theta)}{\partial \theta'}$。
    - 第0步：令k=0，选初始值$\hat{\theta}^{(k)}$;
    - 第1步：$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + \left[H(\hat{\theta}^{(k)})\right]^{-1} s(\hat{\theta}^{(k)})$;
    - 第2步：若$|s(\hat{\theta}^{(k)})|$或$|\hat{\theta}^{(k+1)} - \hat{\theta}^{(k)}|$充分小，则停止；否则k=k+1，转到第1步。（证明在P77，以及一步迭代法）
  - **Fisher得分算法**
    - 将Newton-Raphson算法中$H(\theta) = -\frac{\partial^2 l(\theta)}{\partial \theta \partial \theta'}$换成$H^*(\theta) = E_\theta H(\theta) = -E_\theta \frac{\partial^2 l(\theta)}{\partial \theta \partial \theta'}$就得到Fisher得分算法，即

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + \left[H^*(\hat{\theta}^{(k)})\right]^{-1} s(\hat{\theta}^{(k)}).$$

      即在迭代算法中，Newton-Raphson算法用的是观测Fisher信息矩阵而在Fisher得分算法中用的是期望Fisher信息矩阵。
  - 利用Newton-Raphson算法需要指定θ的初始值$\hat{\theta}^{(0)}$。初始值$\hat{\theta}^{(0)}$的选取对算法收敛性有较大影响。
  - 可以通过粗略方法例如矩方法得到估计值的初始值$\hat{\theta}^{(0)}$。
  - **两个算法的比较**（P81）
  - （P98 **定理2.7**）若$\tilde{\theta}$是θ的$\sqrt{n}$-相合，则Newton-Raphson和Fisher得分的一步迭代是渐近有效的。
- **EM算法**（P83）
- **渐近相对效率和渐近有效性**
  - **定义2.7**：设$\hat{\theta}_n, \tilde{\theta}_n$是基于样本$X_1, X_2, \cdots, X_n$的$\theta$的两个估计，且$\hat{\theta}_n \sim \mathrm{AN}(\theta, \sigma_{1n}^2(\theta)), \tilde{\theta}_n \sim \mathrm{AN}(\theta, \sigma_{2n}^2(\theta))$，则$\hat{\theta}_n$相比于$\tilde{\theta}_n$的渐近相对效率定义为

$$\mathrm{ARE}_\theta(\hat{\theta}_n, \tilde{\theta}_n) = \lim_{n\to\infty} \frac{\sigma_{2n}^2(\theta)}{\sigma_{1n}^2(\theta)}$$

（若极限存在）

  - **定义2.8**：若估计序列$\hat{\theta}_n$满足$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d.f.} N(0, [I(\theta)]^{-1})$，则称$\hat{\theta}_n$是渐近有效的。
  - 若对某个$\theta_0 \in \Theta$，渐近正态的方差$\sigma^2(\theta_0) < [I(\theta_0)]^{-1}$，则称$\hat{\theta}_n$为超有效估计。
  - **定义2.9**：估计序列$\hat{\theta}_n$称为θ的$\sqrt{n}$-相合估计，若$\sqrt{n}(\hat{\theta}_n - \theta) = O_p(1)$。
- **Stein现象**（P104，在平方损失的意义下最小风险，UMVUE不可容许）
- **不变估计**
  - 设统计模型$(\Omega, \mathcal{B}, P_\theta), \theta \in \Theta$。变换$g : \Omega \to \Omega$为1-1，且对每个$\theta \in \Theta$，当$X \sim P_\theta$时，$gX \sim P_{\theta^*}$，若$\theta$取遍空间$\Theta$时，$\theta^*$也取遍空间$\Theta$，则称模型在变换$g$下是不变的（invariant）。若对变换群$G$中任一变换$g$都是不变的，则称模型在变换群$G$下是不变的。
  - 由g可以诱导出一个映射$g^* : \Theta \to \Theta$， $g^*\theta = \theta^*$。诱导变换g^*是Θ上的1-1变换；
  - $G^* = \{g^* : g \in G\}$形成一个群。
  - g是针对X的映射。
  - 对损失函数$L(a, \theta), a \in A$为行动空间，若对任意的$g \in G$，存在唯一的$a^* \in A$使得$L(a, \theta) = L(a^*, g^*\theta), \forall \theta \in \Theta$则称损失函数在变换群$G$下是不变的。
  - 由g可以诱导出一个映射$g^{**} : A \to A$， $g^{**}a = a^{**}$。
  - 一个估计量$\hat{\theta}$满足$\hat{\theta}(gX) = g^{**}\hat{\theta}(X), \quad \forall g \in G$，则称$\hat{\theta}$在变换群$G$下是不变估计。
  - **定义**若$\hat{\theta}$为一个不变估计，它在所有不变估计中有最小的risk，则$\hat{\theta}$称为MRIE（minimum risk invariant estimator）。
  - **位置参数的不变估计**：
    - P110
    - **Pitman估计**
    - $$\hat{\theta}^* = \frac{\int_\infty^\infty tf(X_1 - t, X_2 - t, \cdots, X_n - t)dt}{\int_\infty^\infty f(X_1 - t, X_2 - t, \cdots, X_n - t)dt}.$$
    - 例子在P114，多看作业解答。
  - **刻度参数的不变估计**：
    - P115
    - **Pitman估计**
    - $$\hat{\theta}^* = \frac{\int_\infty^\infty t^{n+h-1} f(tX_1, tX_2, \cdots, tX_n)dt}{\int_\infty^\infty t^{n+2h-1} f(tX_1, tX_2, \cdots, tX_n)dt}.$$
    - 例子在P121，多看作业解答。

第三章

- H_0为原假设，H_1为备择假设。一个假设称为简单的，若只包含一个点，否则称为复合的。

- 可测函数ϕ称为**检验函数**，若$\theta \in \Theta_0$，则$\phi(X)=0$，若$\theta \in \Theta_1$，则$\phi(X)=1$。ϕ通过某个统计量$T=T(X)$而依赖于X，称T为检验统计量。检验函数ϕ把样本空间分成两个不交的集合$\Omega_A=\{X \mid \phi(X)=0\}, \Omega_R=\{X \mid \phi(X)=1\}$，分别称为**接受域**和**拒绝域**。
- **（拒真）** 若$\theta \in \Theta_0$，但$\phi(X)=1$，称为检验ϕ的第I类错误（type I error），其发生的概率为$P_\theta(\phi(X)=1)=E_\theta\phi(X), \theta \in \Theta_0$;
- **（取伪）** 若$\theta \in \Theta_1$，但$\phi(X)=0$称为检验ϕ的第II类错误（type II error）其发生的概率为$P_\theta(\phi(X)=0)=1-E_\theta\phi(X), \theta \in \Theta_1$。
- **例3.2**：设两个检验函数ϕ_1, ϕ_2，记$R_i=\{X \mid \phi_i(X)=1\}$。设$R_1 \subset R_2$，这意味着$E_\theta\phi_1(X) \le E_\theta\phi_2(X), \quad \forall\theta \in \Theta$。因此$1-E_\theta\phi_1(X) \ge 1-E_\theta\phi_2(X), \forall\theta \in \Theta$。这表明，当$\theta \in \Theta_0$时，欲减少犯第I类错误的概率，在$\theta \in \Theta_1$时就会增加犯第II类错误的概率。
- 经典做法给定$\alpha \in [0,1]$，使检验ϕ满足$E_\theta\phi(X) \le \alpha, \forall\theta \in \Theta_0$，即控制第I类错误的概率$\le \alpha$，$\alpha$称为检验$\phi$的显著性水平。然后在此基础上减少犯第II类错误的概率。
- **功效函数**：对给定的检验函数，令$\beta(\theta)=E_\theta\phi(X)$，作为$\theta \in \Theta$傻姑娘的函数，称为检验$\phi$的功效函数（power function）。
  - 若$\theta \in \Theta_1$，称$\beta(\theta)$为检验ϕ在θ处的功效（power）。
  - 水平为α的检验，即要求$\beta(\theta) \le \alpha, \forall\theta \in \Theta_0$。
- **一致最优检验（UMP检验）**
  - **定义3.1**：水平为α的检验ϕ^*称为一致最优检验（uniformly most powerful test）若对任一水平为α的检验$\phi, \forall\theta \in \Theta_1, \beta_{\phi^*}(\theta) \ge \beta_\phi(\theta)$。（即第II类错误概率最小的检验）
  - **引理3.1**：设$T=T(X)$为θ的充分统计量，则对任意检验函数$\phi(X)$，存在只依赖于T的检验函数与ϕ有相同的功效函数。（所以找UMP只要在充分统计量构成的检验中去寻找即可）
  - 对于H_0 和 H_1都为简单模型，**定理3.1**给出UMP检验存在性以及唯一性。
    - $\forall\alpha \in [0,1]$，存在水平为α的UMP检验ϕ^*，满足

$$\phi^*(X)=\begin{cases}1 & f_1(X)>kf_0(X)\\ \gamma & f_1(X)=kf_0(X)\\ 0 & f_1(X)<kf_0(X)\end{cases}$$

      这里$\gamma \in [0,1], k \ge 0$为常数（可为$\infty$）且使得$E_{\theta_0}\phi^*(X)=\alpha$。
  - **引理3.2**：设水平为α的检验ϕ^*，若对$\forall\theta_1 \in \Theta_1$的检验问题$H_0: \theta \in \Theta_0 \text{ v.s. } H_1: \theta=\theta_1$都是UMP检验，则$\phi^*$为检验问题$H_0: \theta \in \Theta_0 \text{ v.s. } H_1: \theta \in \Theta_1$的UMP检验。
  - **定义3.2**：分布族$F(x,\theta), \theta \in \Theta \subset R$（密度为$f(x,\theta)$）称为单调似然比分布族（MLR族），若存在实值统计量$T=T(x)$使得对$\forall\theta_1<\theta_2, \frac{f(x,\theta_2)}{f(x,\theta_1)}$是$T(x)$的非降函数。（$T(x)$是$\theta$的充分统计量）
    - 设X为单参数指数分布族$f(x,\theta)=\exp[c(\theta)T(x)-b(\theta)+s(x)]$,$x \in A$，假设$c(\theta)$为$\theta$的严格增函数，则其为MLR族。

- **引理3.3**：设$X \sim f(x,\theta)$对统计量$T=T(x)$具有MLR，若ϕ为T的非降函数，则$E_\theta\phi(T)$为θ的非降函数。
- **定理3.2**：设$X \sim f(x,\theta)$关于统计量$T=T(x)$为MLR，考察检验问题$H_0:\theta \le \theta_0 \text{ v.s. } H_1:\theta>\theta_0$，
  - 令

$$\phi^*(X)=\begin{cases}1 & T(X)>k\\ \gamma & T(X)=k\\ 0 & T(X)<k\end{cases}$$

    此处k,γ由$E_{\theta_0}\phi^*(X)=\alpha$确定，则$\phi^*(X)$为水平为$\alpha$的UMP检验。
  - $\beta_{\phi^*}(\theta)$为集合$\{\theta\in\Theta \mid 0<\beta_{\phi^*}(\theta)<1\}$上的严格增函数。
  - 对任何满足$E_{\theta_0}\phi(X)=\alpha$的检验，以及$\theta<\theta_0,\theta\in\Theta$,$E_\theta\phi^*(X)\le E_\theta\phi(X)$。（MLR族也使得犯第I类错误的概率一致的小）
  - 对检验问题$H_0:\theta=\theta_0 \text{ v.s. } H_1:\theta>\theta_0$本定理结论仍然成立。
  - 对检验问题$H_0:\theta\ge\theta_0 \text{ v.s. } H_1:\theta<\theta_0$或$H_0:\theta=\theta_0 \text{ v.s. } H_1:\theta<\theta_0$只要将最上面的不等号方向变向且把第二个中的“严格增”改为“严格减”。
- **引理3.4**：设$f_i,i=1,2,\cdots,m+1$为$(\Omega,\mathscr{B},\mu)$上的可积函数，对给定常数$t_i,i=1,2,\cdots$。令可测函数集合

$$\Psi=\left\{\phi \mid 0\le\phi\le1,\int\phi f_i d\mu\le t_i,1\le t_i\le m\right\}$$

$$\Psi_0=\left\{\phi \mid 0\le\phi\le1,\int\phi f_i d\mu= t_i,1\le t_i\le m\right\}$$

  设Ψ_0非空。
  - 若存在常数$k_i,1\le i\le m$使得

$$\phi^*=\begin{cases}1 & f_{m+1}>\sum_{i=1}^m k_if_i\\ 0 & f_{m+1}<\sum_{i=1}^m k_if_i\end{cases},$$

    且$\phi^*\in\Psi_0$，则$\int\phi^*f_{m+1}d\mu=\sup_{\phi\in\Psi_0}\int\phi f_{m+1}d\mu$。此外若所有的$k_i\ge0$，则$\int\phi^*f_{m+1}d\mu=\sup_{\phi\in\Psi}\int\phi f_{m+1}d\mu$。

        - 令$M=\{(\int\phi f_1 d\mu,\int\phi f_2 d\mu,\cdots,\int\phi f_m d\mu)\mid 0\le\phi\le 1\}$，则$M$为$R^m$中的有界闭凸集。且若$(t_1,t_2,\cdots,t_m)$为$M$的内点，则存在常数$k_i,1\le i\le m$，使得上述形式的$\phi^*\in\Psi_0$。
    - **定理3.3**：设X为单参数指数分布族$f(x,\theta)=\exp[c(\theta)t(x)-b(\theta)+s(x)]$,$x\in A$，假设$c(\theta)$为$\theta$的严格增函数，考察检验问题$H_0:\theta\le\theta_1 \text{ or } \theta\ge\theta_2 \text{ v.s. } H_1:\theta_1<\theta<\theta_2$，令

$$\phi^*(X)=\begin{cases}1 & k_1<T(X)<k_2\\ \gamma_i & T(X)=k_i,i=1,2\\ 0 & T(X)<k_1 \text{ or } T(X)>k_2\end{cases}$$

        - 此时k_i,γ_i由$E_{\theta_1}\phi^*(X)=E_{\theta_2}\phi^*(X)=\alpha$确定，则$\phi^*(X)$为水平$\alpha$的UMP检验。
        - 对任何满足$E_{\theta_1}\phi(X)=E_{\theta_2}\phi(X)=\alpha$的检验以及$\theta<\theta_1 \text{ or } \theta>\theta_2$,$\theta\in\Theta,E_\theta\phi^*(X)\le E_\theta\phi(X)$。
- **一致最优无偏检验**
    - **定义3.3**：给定水平α，对检验问题$H_0:\theta\in\Theta_0 \text{ v.s. } H_1:\theta\in\Theta_1$，称检验$\phi$在水平$\alpha$下是**无偏的**，若$\beta_\phi(\theta)\le\alpha,\forall\theta\in\Theta_0$，且$\beta_\phi(\theta)\ge\alpha,\forall\theta\in\Theta_1$。
    - **定义3.4**：若对任水平α的无偏检验ϕ，存在水平α的无偏检验ϕ^*使得$\beta_\phi(\theta)\le\beta_{\phi^*}(\theta),\forall\theta\in\Theta_1$，则称$\phi^*$为水平$\alpha$的一致最优无偏检验（UMPU）。
    - **定义3.5**：给定水平α，检验ϕ称为对于集$\partial\Theta_{01}$是相似的若$\beta_\phi(\theta)=\alpha,\forall\theta\in\partial\Theta_{01}$。
    - **引理3.5**：若对每一个检验$\phi,\beta_\phi(\theta)$为θ的连续函数。若检验ϕ^*为水平α的相似检验中一致最优（功效最大），则ϕ^*为水平α的UMPU检验。
    - **定义3.6**：设统计量T对于分布族$P_\theta,\theta\in\partial\Theta_{01}$是充分的，若检验函数$\phi$满足$E_\theta(\phi(X)\mid T)=\alpha,a.s.P_\theta$则称$\phi$对于$T$在$\partial\Theta_{01}$上有**Neyman结构**。
    - **引理3.6**：设ϕ对于T在$\partial\Theta_{01}$上有Neyman结构，则ϕ对于集$\partial\Theta_{01}$相似。若T还是$P_\theta,\theta\in\partial\Theta_{01}$（有界）完全的，则两者等价。
    - **定理3.4**：对于**定理3.3**中单参数指数分布族，给定水平α，考虑检验$H_0:\theta_1<\theta<\theta_2 \text{ v.s. } H_1:\theta\le\theta_1 \text{ or } \theta\ge\theta_2$。若检验$\phi^*$有形式

$$\phi^*(X)=\begin{cases}1 & T(X)<k_1 \text{ or } T(X)>k_2\\ \gamma_i & T(X)=k_i,i=1,2\\ 0 & k_1<T(X)<k_2\end{cases}$$

      其中k_i,γ_i由$E_{\theta_1}\phi^*(X)=E_{\theta_2}\phi^*(X)=\alpha$确定，则$\phi^*(X)$为水平$\alpha$的UMPU检验。
    - **定理3.5**：对于**定理3.3**中单参数指数分布族，给定水平α，考虑检验$H_0:\theta=\theta_0 \text{ v.s. } H_1:\theta\ne\theta_0$。若检验$\phi^*$有形式

$$\phi^*(X)=\begin{cases}1 & T(X)<k_1 \text{ or } T(X)>k_2\\ \gamma_i & T(X)=k_i, i=1,2\\ 0 & k_1<T(X)<k_2\end{cases}$$

其中k_i,γ_i由$E_{\theta_0}\phi^*(X)=\alpha, E_{\theta_0}T(X)\phi^*(X)=\alpha E_{\theta_0}T(X)$确定，则$\phi^*(X)$为水平$\alpha$的UMPU检验。

- 对于多参数的检验问题参考P33-P38，冗余参数的概念。
- **One-sample Problems**（P45-P49）
- **Two-sample Problems**（P50-P56）

- **一致最优不变检验（UMPI）**（P57-P63）
  - **定义3.7**：对检验问题$H_0:\theta\in\Theta_0 \text{ v.s. } H_1:\theta\in\Theta_1$和变换群$G$
    - 称检验问题在变换群G下是不变的（invariant），如果$g^*\Theta_0=\Theta_0, g^*\Theta_1=\Theta_1, \forall g^*\in G^*$
    - 对于在群G下不变的检验问题，一个检验ϕ称为是不变检验若$\phi(gX)=\phi(X), \forall g\in G$。
    - 若对任水平α的不变检验ϕ，存在水平α的不变检验ϕ^*使得$\beta_\theta(\theta)<\beta_{\phi^*}(\theta), \forall\theta\in\Theta_1$，则称$\phi^*$为水平$\alpha$的一致最优不变检验（UMPI）。
    - 统计量$T(X)$称为在变换群G下为极大不变量若$T(gX)=T(X), \forall g\in G$，且若$T(X)=T(Y)\Rightarrow$存在$g$使得$gX=Y$。
  - 求UMPI检验时，先找一个适当的极大不变量T（不一定唯一），然后找出T的分布族，在此分布族下是否有UMP检验，若有则为UMPI检验
    - 设统计量$T(X)$为群G的极大不变量，ϕ为不变检验$\Leftrightarrow$存在可测函数h使得$\phi(X)=h(T(X))$。
    - 多看作业
    - 使用充分统计量
    - 群G在样本空间的变化能否在充分统计量T的值域空间诱导出一个变换群G^T（若不能，则T无用）
    - 转化到T后在群G^T下施以不变性要求，对此再利用前面说的方法求UMPI检验。
- **假设检验与Wald决策理论**（P64-P68）
  - 设a_0表示接受，a_1表示拒绝，设损失函数为$L(a,\theta),\theta\in\Theta$，则对任一随机化决策函数$\phi(X)$，其风险为

$$R_\theta(\phi)=E_\theta[L(a-0,\theta)(1-\phi(X))+L(a_1,\theta)\phi(X)]$$

  问题归结于找ϕ使$R_\theta(\phi)$最小。
- **基于似然的检验**

- **定义3.8**：统计量

$$\Lambda(X)=\frac{\sup_{\theta\in\Theta_0} f(X,\theta)}{\sup_{\theta\in\Theta} f(X,\theta)}=\frac{\sup_{\theta\in\Theta_0} L(X,\theta)}{\sup_{\theta\in\Theta} L(X,\theta)}$$

  称为似然比统计量，若$\Lambda<c$拒绝H_0。
- 给定水平α，若存在c_α使得$\sup_{\theta\in\Theta_0} P_\theta(\Lambda(X)<c_\alpha)=\alpha$，则LRT为水平$\alpha$的检验。
- **定理3.7**：设$X_1,X_2,\cdots,X_n$ i.i.d $\sim f(x,\theta),\theta\in\Theta$，满足定理2.6中条件$B1-B6$且$I(\theta)=J(\theta),\forall\theta\in\Theta$。若MLE$\hat{\theta}_n$满足$\sqrt{n}(\hat{\theta}_n-\theta)\xrightarrow{d.f.}N(0,[I(\theta)]^{-1})$。考察检验问题$H_0:\theta\in\Theta_0=\{\theta\in\Theta\mid g(\theta)=0\}\text{v.s.}H_1:\Theta/\Theta_0$,这里$g:R^p\to R^r(r<p)$为连续可微函数，$g=(g_1,g_2,\cdots,g_r)^T,\text{rank}\frac{\partial g(\theta)}{\partial\theta^T}=r$。则在$H_0$下，LRT统计量$\Lambda_n(X)$有

$$-2\log\Lambda_n(X)\xrightarrow{d.f.}\mathcal{X}_r^2$$

  此检验拒绝域为$-2\log\Lambda_n(X)>\mathcal{X}_r^2(\alpha)$，渐近有$\alpha$水平。
- **截断型分布的似然比**（P76）
- **Bartlett Correction**（P77）
- **Wald test**
  - $$W_n=n[g(\hat{\theta}_n)]^T\left\{\frac{\partial g(\hat{\theta}_n)}{\partial\theta^T}[I(\hat{\theta}_n)]^{-1}\left[\frac{\partial g(\hat{\theta}_n)}{\partial\theta^T}\right]^T\right\}^{-1}g(\hat{\theta}_n)>c$$

    此处$\hat{\theta}_n$为MLE。
- **Score test**
  - $$R_n=\left[S(\hat{\theta}_{n0})\right]^T\left[nI(\hat{\theta}_{n0})\right]^{-1}S(\hat{\theta}_{n0})>c$$

    其中$\hat{\theta}_{n0}$为在约束H_0下的MLE，$S(\theta)=\frac{\partial l(\theta)}{\partial\theta}$为score function。
- 在H_0下，$W_n\xrightarrow{d.f.}\mathcal{X}_r^2$，故当$W_n>\mathcal{X}_r^2(\alpha)$时拒绝$H_0$，此检验渐近地有$\alpha$水平。
- 在H_0下，$R_n\xrightarrow{d.f.}\mathcal{X}_r^2$，故当$R_n>\mathcal{X}_r^2(\alpha)$时拒绝$H_0$，此检验渐近地有$\alpha$水平。
  - W_n,R_n的定义中$nI(\theta)$用基于n个样本的Fisher信息阵$I_n(\theta)$代替。
  - 三个检验是渐近等价的。

- **定义3.9**：p-value为在观测样本X下，零假设被拒绝的最小显著性水平。p-value也称为observed significance level。

- **定义3.10**：$p(X)=\inf\{\alpha \mid \phi_\alpha(X)=1\}$称为（样本$X$）的p-value。
- **拟合优度检验**
  - **Pearson**：

$$T_n=\sum_{j=1}^{k}\frac{(\xi_{nj}-np_j)^2}{np_j}\xrightarrow{d.f.}\mathcal{X}_{k-1}^2$$

  - 由具体数据可以计算出T_n的值，记为t，从而可以计算$p(t)=P(\mathcal{X}_{k-1}^2>t)$，即为检验的$p$值。
  - 对于有限多个点的分布族F_θ，需要先利用$X_1, cdots, X_n$对θ做估计$\hat{\theta}_n$（通常是极大似然估计MLE），然后计算

$$T_n=\sum_{i=1}^{k}\frac{(\xi_{nj}-np_j(\hat{\theta}_n))^2}{np_j(\hat{\theta}_n)}\xrightarrow{d.f.}\mathcal{X}_{k-r-1}^2$$

- 列联表检验
  - $$T_n=\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(nX_{ij}-m_in_j)^2}{nm_in_j}$$
  - 当n充分大时，在原假设H_0下$T_n\xrightarrow{d.f.}\mathcal{X}_{(r-1)(s-1)}^2$。
- 基于经验分布的拟合优度检验
  - Glivenko-Cantelli定理：

$$\sup_x\left|\hat{F}_n(x)-F(x)\right|\to 0\quad \text{a.s.}$$

  - D_n, V_n, C_n, A_n等定义及性质P109-P117。
- 正态性检验
  - Q-Q plot（P120）
  - Shapiro-Wilk Test（P124）
- 非参数检验
  - 符号检验
  - 基本思路：设$X_1, X_2, \cdots, X_n$ i.i.d. $\sim F(x)$（未知），固定u，$p=F(u)$$=P(X_i\le u)$，考虑如下的检验问题（其中$p_0\in(0,1)$已知）
    - $$H_0: p\le p_0 \leftrightarrow H_1: p>p_0;$$
    - $$H_0: p=p_0 \leftrightarrow H_1: p\ne p_0。$$

令

$$\Delta_i = \begin{cases} 1, & X_i - u \le 0 \\ 0, & \text{otherwise} \end{cases}$$

则$\Delta_i \text{i.i.d} \sim \text{Bernoulli}$（$p$），然后急于之前的UMP检验以及UMPU检验进行检验。

- 置换检验（Permutation Tests）P131
- 秩检验P150

第四章

- 置信区间：称区间$[L(X), U(X)]$为θ的$100(1-\alpha)\%$置信区间，若$P_\theta(L(X) \le \theta \le U(X)) \ge 1-\alpha$，$\forall \theta \in \Theta$。
- 给定样本X，区间$[L(X), U(X)]$在样本重复下，有$100(1-\alpha)\%$的这样的区间将会含有真值。

第五章

- **Jeffreys先验**：$\pi(\theta) \propto |I(\theta)|^{\frac{1}{2}}$
- **定义5.1**：设A为决策问题的行动空间，$L(a,\theta) \ge 0$为损失函数，θ先验分布为$\Pi(\theta)$，若决策函数$\delta(x)$满足$\forall x, E[L(\delta(x),\theta) \mid X = x] = \min_{a \in A} E[L(a,\theta) \mid X = x]$，则称$\delta(x)$为Bayes决策，记为$\delta_B(x)$。
- **定义5.2**：估计$\hat{\theta}_M$称为θ（相对于损失函数L）的极小极大估计，若对任何估计$\hat{\theta}, \sup_{\theta \in \Theta} R_\theta(\hat{\theta}) \ge \sup_{\theta \in \Theta} R_\theta(\hat{\theta}_M)$
- 经验Bayes P35
- Bayes检验 P42