STAT 561: Statistical Inference II

Shihao Tong

University of British Columbia

Update: February 27, 2022

本文件主要参考资料为 *Junshao* 《数理统计》, *JiahuaChen* Stat560 and 561 notes，以及 lecture 所讲内容。涉及引用之处遂佐以链接以溯源。谬误难免，免责于吾。

1 Fundamental Measure Theory

Definition 1.1 *(σ-Algebra & measurable space) Given sample space (outcome space) Ω, let $\mathcal{F}$ be the set of subsets of Ω. Then $\mathcal{F}$ is called the σ-Algebra if the following holds:*

- *Empty set is in $\emptyset \in \mathcal{F}$*
- *Closed under complement. If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$*
- *Closed under union. If $A_i \in \mathcal{F}, i = 1, 2, \ldots, n$, then $\cup_i^n A_i \in \mathcal{F}$*

Then the tuple $(\Omega, \mathcal{F})$ is called the measurable space.

一个 Ω 的最大的 σ 代数是其 power set，即其所有子集的集合。最小的是 $\{\emptyset, \Omega\}$。包含 A 的最小的 σ-algebra 用 $\sigma(A) = \{A, A^c, \emptyset, \Omega\}$ 来表示。注意这是定义的“可测空间”，下面才是“测度空间”。常用的一个可测空间是 *Borel σ-field*。一个简单的例子说明其性质

Example 1.1 *Let $\Omega = \{a, b, c, d\}, A = \big\{\{a\}, \{b\}\big\}$. Then the smallest σ field generated by set A becomes*

$$\sigma(A) = \big\{\emptyset, \Omega, \{a\}, \{b\}, \{a, b\}, \{b, c, d\}, \{a, c, d\}, \{c, d\}\big\}$$

注意$A \subseteq \sigma(A)$，而不是 $A \in \sigma(A)$.

Definition 1.2 *(Borel σ-field) The Borel σ-algebra of $\mathbb{R}$, written $\mathcal{B}$, is the σ-algebra generated by the open sets. That is, if O denotes the collection of all open subsets of $\mathbb{R}$, then $\mathcal{B} = \sigma(O)$.*

个人认为最好区别一下 *collection* 和 *set* 这两个说法。前者只是一堆东西，而后者“集合”则需要满足最基本的几个性质（i.e 互异，无序，确定性）

Definition 1.3 *(Measure) Given measure space* $(\Omega, \mathcal{F})$*. Then a set function (i.e domain is set of set)* $\mathcal{V}$ *defined on* $\mathcal{F}$ *is called a measure*

- 测度非负性 $0 \geq \mathcal{V}(A) \geq \infty$
- 空集测度为零 $\mathcal{V}(\emptyset)$
- 测度可数可加性 *If* $A_i \in \mathcal{F}, i = 1, 2, ..., n$*, and* A_i, A_j *are disjoint (i.e* $A_i \cap A_j = \emptyset$ *for* $i \neq j$*), then*
$$\mathcal{V}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathcal{V}(A_i) \tag{1}$$

由于测度可以取 ∞，运算时我们遵循对无穷的一般处理方法，但是对于 $\infty - \infty$ 和 ∞/∞ 没有定义。下面的两个重要的常用测度。

Example 1.2 *(Counting measure) Given measurable space* $(\Omega, \mathcal{F})$*. Let* $\mathcal{V}(A)$ *be the number of elements of* $A \in \mathcal{F}$*. Then* $\mathcal{V}$ *defined on* $\mathcal{F}$ *is called the counting measure.*

Example 1.3 *(Lebesgue Measure) There is a unique measure m on* $(\mathbb{R}, \mathcal{B})$ *that satisfies*

$$m([a, b]) = b - a$$

for every finite interval $[a, b]$*, This is called Lebesgue Measure. If we restrict m to the measurable space* $([0, 1], \mathcal{B}_{[0,1]})$ *then m is a probability measure.*

2 Basic for Inference

Let $\mathbf{X} = (X_1, X_2, ..., X_n)$ be iid sample from a model $\mathcal{F}$.

Definition 2.1 *(Statistic) A statistic is a measurable function of data which does not depend on any unknown parameters.*

即统计量不能包含有关 paramter 的信息，如期望，方差等。

Definition 2.2 *(Sufficient Statistics) Let T(x) be a statistic. If the conditional distribution of X given T does not depend on unknown parameter values, we say T is a sufficient statistics.*

目的是为了 inference 随机变量 X 分布所依赖的某个 parameter θ。Sample **X** 中包含的信息可以分成两部分：一部分是与 θ 有关的，一部分是无关的。如果一个统计量是充分的，则 sample 中所有有关 θ 的信息都包含在 $T(x)$ 里了，用 $T(x)$ 或是所有 sample 来对 θ 做推断没有区别。在没有 T 的 condition 之前，X 是 depends on θ 的，而 $T(X)$ 就像是一个 filter，过滤掉了有关 θ 的信息，则剩余的 X (i.e conditional X) 就与 θ 无关了。

Lemma 2.1 *(Factorization) Iff the density function of X can be written as*

$$f(x; \theta) = h(x)g(T(x); \theta) \tag{2}$$

then $T(x)$ is sufficient statistics for θ.

Definition 2.3 *(Minimum sufficient) Sufficient statistics $T(x)$ is minimum sufficient if T is the function of every other sufficient statistic.*

Sufficient 是指统计量包含关于参数的充分的信息，但有可能信息中有冗余的部分。Minimum sufficient 仍然可能包含冗余信息，因此下面引入 completeness。

Definition 2.4 *(Completeness) Statistics $T(x)$ is said to be complete if $\mathbb{E}(g(T)) = 0$ implies $g(\cdot) = 0$ almost surely for any function g Check*

如果存在一个 $g(\cdot)$ 使得 $\mathbb{E}(g(T)) = 0$，那么就说明统计量 T 中仍然包含有关于 θ 的信息从而就不完备。

Example 2.1 *(Sufficient Stat) Let $X \sim Exp(\lambda)$ and we have n i.i.d samples. Then*

$$f(\boldsymbol{x};\lambda) = \lambda^n \exp\{-\lambda \sum_{i=1}^{n} x_i\} \prod_{i=1}^{n} \mathbb{I}_{[0,\infty)}(x_i)$$

where the statistics $T(x) = \sum_{i=1}^{n} x_i$ is sufficient.

Definition 2.5 *(UMVUE) An unbiased estimator $\widehat{\theta}$ is uniformly minimum variance estimator of θ, if for any other unbiased estimator $\widetilde{\theta}$, s.t*

$$\mathrm{Var}_\theta(\widehat{\theta}) \le \mathrm{Var}_\theta(\widetilde{\theta})$$

for all $\theta \in \Theta$.

The subscript θ is used to represent the variance calculation is based on true parameter θ.

3 Exponential Family

Definition 3.1 *Suppose there exists a real valued function $\eta(\theta), T(x), A(\theta)$ and $h(x)$ s.t.*

$$f(x;\theta) = \exp\{\eta(\theta)T(x) - A(\theta)\}\, h(x) \tag{3}$$

we say $\{f(x;\theta) : \theta \in \Theta \subset \mathbb{R}\}$ is a one-parameter exponential family.

即一个 r.v 的 pdf(pmf) 可以写成 (1) 中的形式，则该 r.v 就是指数分布族的一个。可以简单做一下运算，如果假设所有的 sample 都是 i.i.d 的，则其 joint density 仍然属于指数分布族。θ 的变化 span 了整个 set。

Example 3.1 *Suppose $X_1, X_2, \ldots, X_n$ are i.i.d from Binomial (m, θ) There joiny deusity function* *is*

$$\begin{aligned} f(x_1, x_2, \ldots, x_n; \theta) &= \prod_{i=1}^{n} \binom{m}{x_i} \theta^{x_i} (1-\theta)^{m-x_i} \\ &= \left[\prod_{i=1}^{n} \binom{m}{x_i}\right] \theta^{\sum x_i} (1-\theta)^{\sum m-x_i} \\ &= \left[\prod_{i=1}^{n} \binom{m}{x_i}\right] \left(\frac{\theta}{1-\theta}\right)^{\sum x_i} (1-\theta)^{-mn} \\ &= \underbrace{\left[\prod_{i=1}^{n} \binom{m}{x_i}\right]}_{h(x)} \exp\left\{ \underbrace{\sum x_i}_{T(x)} \underbrace{\ln \frac{\theta}{1-\theta}}_{\eta(\theta)} - \underbrace{mn \ln^{1-\theta}}_{A(\theta)} \right\} \end{aligned}$$

Note 此处假设 m 是已知的，只有 θ 是 *parameter*。对于 *Binomial Distribution* 我们有时候也把 $\eta(\theta)$ 这个 *log-odd* 写作

$$\theta = \frac{\exp(\eta)}{1 + \exp(\eta)}$$

则其表达形式成为

$$g(x_1, ..., x_n; \theta) = \exp\{\eta T(x) - nm \ln(1 + \exp(\eta))\} h(x)$$

用相同方法可知，*Poisson, Negative Binomial* 都属于指数分布族。

Definition 3.2 *(Support of r.v) Let X be a random variable or vector. The support of X of that of its distribution is the set of all x s.t* $\forall \delta > 0$,

$$\mathbb{P}\{X \in (x - \delta, x + \delta)\} > 0$$

直观上定义 support 的目的是为了说明 $X = x$ 是可能发生的。对于离散的情况很容易理解，而对于连续的随机变量（i.e 如 $Z \sim N(0, 1)$ 则对于任意的 z，$\mathbb{P}(Z = z) = 0$）我们则需要先如上所述，讲 x expand 到一个很小的区间上讨论。例如，standard exponential distribution has support $[0, \infty)$，standard normal has support $\mathbb{R}$。因此可以大致将 support 理解为

$$\{x : f(x; \theta) > 0\}$$

If two distributions belong the same one-parameter exponential family, then they have the same support. 同一指数族（i.e 同样的 T, η, A 等实函数，不同的 θ）中的分布，其 support 不依赖于参数 θ。因此 $X \sim \text{Uniform}(0, \theta)$ 就不属于指数分布族，因为其 support 依赖于参数 θ.

Definition 3.3 *Suppose there exists a real vector function* $\eta(\boldsymbol{\theta}), T(x), A(\boldsymbol{\theta})$ *and* $h(x)$ *s.t.*

$$f(x; \boldsymbol{\theta}) = \exp\left\{\sum_{j=1}^{k} \eta_j(\boldsymbol{\theta}) T_j(x) - A(\boldsymbol{\theta})\right\} h(x) \tag{4}$$

where

- *$\eta_.(\boldsymbol{\theta})$ and $A(\boldsymbol{\theta})$ are all maps from $\Theta \to \mathbb{R}$*
- *$T_.(x)$ and $h(x)$ are all maps from $\mathbb{R}^p \to \mathbb{R}$*

we say $\{f(x;\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^k\}$ is a multi-parameter exponential family. Defenition from here page 18

将定义拓展至多参数的情况，这样就可以说明，诸如 normal distribution 是指数分布族了。注意此处的 k，即其和最多有 $\boldsymbol{\theta}$ 的 conponet 一样多的项和。还有一种定义是

$$f(x;\boldsymbol{\theta}) = a(x)b(\boldsymbol{\theta})\exp\left\{\sum_{j=1}^{k}\eta_j(\boldsymbol{\theta})T_j(x)\right\} \tag{5}$$

Example 3.2 *(Normal distribution) Let $X \sim N(\mu, \sigma^2)$. Let $X_1, ..., X_n$ are i.i.d samples. Then there joint density becomes*

$$\begin{aligned}\phi\left(x_1,\ldots,x_n;\mu,\sigma^2\right) &= (2\pi)^{-n/2}\sigma^{-n}\exp\left\{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right\}\\ &= (2\pi)^{-n/2}\exp\left\{\frac{\mu}{\sigma^2}\sum_{i=1}^{n}x_i - \frac{1}{2\sigma^2}\sum_{i=1}^{n}x_i^2 - \frac{n\mu^2}{2\sigma^2} - n\log\sigma\right\}.\end{aligned}$$

It is clear that this fit definition 3.3.

对于指数分布族，有如下性质可以将指数与 $b(\theta)$

Lemma 3.1 *If we have*

$$\frac{d}{d\boldsymbol{\theta}_j}\int f(x;\boldsymbol{\theta})dx = \int\frac{\partial}{\partial\boldsymbol{\theta}_j}f(x;\boldsymbol{\theta})dx$$

where $\boldsymbol{\theta}_j$ is the jth component of $\boldsymbol{\theta}$. Then $\forall j, 1 \le j \le k$ we have

$$\mathbb{E}_{\boldsymbol{\theta}}\left(\sum_{\ell=1}^{k}T_\ell(x)\frac{\partial}{\partial\boldsymbol{\theta}_j}\eta(\boldsymbol{\theta})\right) = -\frac{\partial}{\partial\boldsymbol{\theta}_j}\log b(\boldsymbol{\theta}) \tag{6}$$

前提条件并不 trival，导数并不能由外面放到积分里面。

Proof. LHS 是关于 x 的积分，对于 $\forall\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^k$ 都有

$$\int f(x;\boldsymbol{\theta})dx = 1$$

所以变化 $\boldsymbol{\theta}$ 积分值并不变化，所以 LHS 永远为 0。对于 RHS,

$$\frac{\partial}{\partial\boldsymbol{\theta}_j}f(x;\boldsymbol{\theta}) = a(x)\exp\left\{\sum_{\ell=1}^{k}T_\ell(x)\eta_\ell(\theta)\right\}\cdot\left\{b'(\theta) + b(\theta)\sum_{\ell=1}^{k}T_\ell(x)\frac{\partial}{\partial\theta_j}\eta_\ell(\theta)\right\}$$

对其进行积分并令结果为 0

$$\int a(x)b'(\theta)\cdot\sum_{\ell=1}^{k}T_\ell(x)\eta_\ell(\theta)dx = \int a(x)b(\theta)\sum_{\ell=1}^{k}T_\ell(x)\eta_\ell(\theta)\cdot\sum_{\ell=1}^{k}T_\ell(x)\frac{\partial}{\partial\theta_j}\eta_\ell(\theta)dx$$

两边同乘 $b(\theta)$，移项得

$$-b'(\theta)\cdot 1 = b(\theta)\mathbb{E}_{\boldsymbol{\theta}}\left(\sum_{\ell=1}^{k} T_\ell(x)\frac{\partial}{\partial \boldsymbol{\theta}_j}\eta(\boldsymbol{\theta})\right)$$

4 Choice of Estimator

在有多个 estimator 时我们需要将他们进行比较，常用的就是 MSE 因为其综合度量了 bias 和 variance。首先介绍几个不等式。

Theorem 4.1 *(Holder's Inequality) Let* $\mathbb{E}(|X|^p) < \infty, \mathbb{E}(|Y|^q) < \infty$ *for* p, q *where* $1/p + 1/q = 1$, *then*

$$|\mathbb{E}(XY)| \le \mathbb{E}(|XY|) \le \left\{\mathbb{E}(|X|^p)\right\}^{\frac{1}{p}}\left\{\mathbb{E}(|Y|^q)\right\}^{\frac{1}{q}} \tag{7}$$

Proof. We show the first inequality by

$$-|XY| \le XY \le |XY| \implies -\mathbb{E}(|XY|) \le \mathbb{E}(XY) \le \mathbb{E}(|XY|) \implies |\mathbb{E}(XY)| \le \mathbb{E}(|XY|)$$

For the second inequality, we appeal to Yong's inequality which is,if we have $1/p + 1/q = 1$ then $\forall a, b \ge 0$ we have

$$ab \le \frac{a^p}{p} + \frac{b^q}{q} \tag{8}$$

apply this with

$$a = \frac{|X|}{\{\mathbb{E}(|X|^p)\}^{\frac{1}{p}}} \quad b = \frac{|Y|}{\{\mathbb{E}(|Y|^q)\}^{\frac{1}{q}}}$$

then

$$\frac{|X|\,|\ Y\ |}{\mathbb{E}(|X|^p)^{\frac{1}{p}}\,\mathbb{E}(|Y|^q)^{\frac{1}{q}}} \leqslant \frac{|x|^p}{\mathbb{E}(|x|^p)}\cdot\frac{1}{p} + \frac{|Y|^q}{\mathbb{E}(|Y|^q)}\cdot\frac{1}{q}$$

take expectation on both hand sides (notice $|X||Y| = |XY|$)

$$\frac{\mathbb{E}(|XY|)}{\mathbb{E}(|X|^p)^{\frac{1}{p}}\,\mathbb{E}(|Y|^q)^{\frac{1}{q}}} \le \frac{1}{p} + \frac{1}{q} = 1$$

Theorem 4.2 *(Cauchy-Schwarz inequality) The probability version of C-S is*

$$\mathbb{E}(|XY|)^2 \le \mathbb{E}(|X|^2)\mathbb{E}(|Y|^2) \tag{9}$$

The poof is simply using Holder's inequality and let $p = q = 2$. This equality then easily derive the covariance inequality

$$|\mathrm{Cov}\,(X, Y)|^2 \le \mathbb{E}\left((X - \mathbb{E}(X))^2\right)\mathbb{E}\left((Y - \mathbb{E}(Y))^2\right) = \mathrm{Var}\,(X)\,\mathrm{Var}\,(Y) \tag{10}$$

Remark *(Mean Square Error Decomposition) The MSE can be decomposed into*

$$\mathrm{MSE}(\widehat{\theta}_n, \theta) = \mathbb{E}\left((\widehat{\theta}_n - \theta)^2\right) = \mathrm{Var}\left(\widehat{\theta}_n\right) + \mathrm{Bias}^2\left(\widehat{\theta}_n\right) \tag{11}$$

括号中的 $(\widehat{\theta}_n - \theta)^2$ 是一个 loss function，也可以是别的 loss function 比如 $|\widehat{\theta}_n - \theta|$ 或者是 0-1 loss. Theoretically the best estimator can be defined as: $\widetilde{\theta}_n$ is the best estimator if

$$MSE(\widetilde{\theta}_n, \theta) \leq MSE(\widehat{\theta}_n, \theta)\ \forall \widehat{\theta}_n$$

But this is hopeless, since if we set the estimator as a constant equal to the true parameter θ then the MSE = 0.

Definition 4.1 *(Convex Function) The function $\phi(x)$ is said to be convex if $\forall \alpha \in [0, 1]$, and let a, b be two distinct points in the domain, then*

$$\phi\left(\alpha a + (1-\alpha)b\right) \leq \alpha\phi(a) + (1-\alpha)\phi(b)$$

If the $\leq$ is strictly less than, then it is said to be strictly convex.

Theorem 4.3 *Let $\phi(x)$ to be convex, then $\forall t$ in the domain, $\exists \ell(x)$, which is a line $\ell(x) = a(x-t) + \phi(t)$ s.t*

$$\phi(x) \geq \ell(t)$$

even if t is not differentiable.

即对于 convex 的函数上的所有的点，都存在一条切线位于 $\phi(x)$ 之下，即使 t 点不可导。Recall *Jensen's Inequality*

Theorem 4.4 *(Jensen's Inequality) Let $\phi(\cdot)$ be convex. Let X be a random variable. Then*

$$\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X))$$

Proof. 利用 Theorem 4.3，因为 ϕ convex, 则存在直线 $\ell(x;t) = a(x-t) + \phi(t) \leq \phi(x)\forall x$。令 $t = \mathbb{E}(X)$ 则有

$$\phi(X) \geq a(X - \mathbb{E}(X)) + \phi(\mathbb{E}(X))$$

Then take expectation on both sides we get the results.

Definition 4.2 *(Risk) 此处讨论的 loss 指 estimator 在估计真是 parameter 时的 loss，定义为*

$$\mathcal{R}(\theta, \widehat{\theta}_n) = \mathbb{E}_\theta\left(L(\theta, \widehat{\theta}_n)\right)$$

假设 loss function $L(\theta, \widehat{\theta}_n)$ is **strictly convex in the second argument** $\widehat{\theta}_n$.

下面的定理说明，以 risk 作为判定标准，给定任意的 estimator 总会存在一个比它更好（risk 更小）的 estimator

Theorem 4.5 *Let $T(x)$ be a sufficient statistics and $\widehat{\theta}_n$ is a given estimator. Then the new estimator*

$$\widetilde{\theta}_n(t) = \mathbb{E}_\theta(\widehat{\theta}_n \mid T = t) = \mathbb{E}(\widehat{\theta}_n \mid T = t)$$

(assume the loss function here is strictly convex) satisfies $\mathcal{R}(\theta, \widetilde{\theta}_n) < \mathcal{R}(\theta, \widehat{\theta}_n)$ unless $\widehat{\theta}_n = \widetilde{\theta}_n$ the = can be achieved.

Proof. Consider loss function $L(\theta, \cdot)$ strictly convex. Then using Jensen's inequality we have

$$L\left(\theta, \widetilde{\theta}_n\right) = L\left(\theta, \mathbb{E}\left(\hat{\theta}_n \mid T = t\right)\right) < \mathbb{E}\left(L\left(\theta, \hat{\theta}_n\right) \mid T = t\right)$$

then take expectation w.r.t T on both sides we get

$$\mathcal{R}(\theta, \widetilde{\theta}_n) < \mathcal{R}(\theta, \widehat{\theta}_n)$$

Notice the RHS is by using the iterated law of expectation.

Theorem 4.6 （关于完备充分统计量的唯一无偏估计存在性）*Suppose T is a sufficient and complete statistics, then if $\exists$ a unbiased estimator for θ, then $\exists$ a unique unbiased estimator for θ that is a function of T and it is UMVUE (check).*

Proof. Let $\widehat{\theta}_1$ is unbiased (not a function of T, it is simply unbiased). Then we come up with, naturally

$$\widehat{\theta}_2 = \mathbb{E}(\widehat{\theta}_1 \mid T) = f(T)$$

is unbiased (iterated expectation). Assume there exists another unbiased estimator which is also a function of T which is $\widehat{\theta}_3 = g(T)$, then

$$\mathbb{E}(\widehat{\theta}_2 - \widehat{\theta}_3) = \mathbb{E}(f(T) - g(T)) = 0$$

then appeal to the completeness of T, $f(T) - g(T) = 0$, so $f(T) = g(T)$. this is not a good proof, see the hand script

注意重点是存在 estimator 是关于 T 的函数。下面一个例子是如何寻找 Poisson 分布的 $e^{-\lambda}$ UMVUE。

Example 4.1 *Let $X_1, ..., X_n$ are i.i.d Poisson(λ) distribution and we want an unbiased estimator for $e^{-\lambda}$) while $e^{-\overline{X}_n}$ is biased for $e^{-\lambda}$. Then we observe that*

$$e^{-\lambda} = Pr_\lambda(X = 0)$$

so we can try to find an unbiased estimator for the probability of $X = 0$ so it should be

$$h(x_1, x_2, ..., x_n) = \mathbb{I}(X = 0) = \begin{cases} 1, & X = 0 \\ 0, & o/w \end{cases}$$

it is easy to check the expectation of $h(x)$ exactly the $e^{-\lambda}$ so unbiased.Unfinished

对于一个 estimator 我们可以找到一个合适的 lower bound 来衡量其精确程度即 Carmer Rao lower bound

Theorem 4.7 *(Carmer-Rao Lower bound) Let some regular conditions be hold:*

- $\{x : f(x;\theta) > 0\}$ *the set does not depend on* θ
- *The parameter set* Θ *is an open*
- *The pdf* $f(x;\theta)$ *is differentiable and finite*
- *The differential operator is exchangeable that is*

$$\frac{d}{d\boldsymbol{\theta}_j}\int f(x;\boldsymbol{\theta})dx = \int \frac{\partial}{\partial\boldsymbol{\theta}_j} f(x;\boldsymbol{\theta})dx$$

- *The differential operator is exchangeable even for function which is*

$$\frac{d}{d\boldsymbol{\theta}_j}\int h(x)f(x;\boldsymbol{\theta})dx = \int h(x)\frac{\partial}{\partial\boldsymbol{\theta}_j} f(x;\boldsymbol{\theta})dx$$

- *For any* $h(x)$*, an estimator of parameter* θ *3its second moment is finite which is*

$$\mathbb{E}_\theta(h(x)^2) < \infty$$

then we have

$$\operatorname{Var}(h(x)) \geq \frac{1}{\mathcal{I}(\theta)}\left[\frac{d}{d\theta}\mathbb{E}_\theta(h(x))\right]^2 \tag{12}$$

Proof. Starting with the last condition

$$\begin{aligned}\frac{d}{d\theta}\mathbb{E}_\theta(h(x)) &= \int h(x)\frac{d}{d\theta}f(x;\theta)dx = \int h(x)\frac{\frac{\partial}{\partial\theta}f(x;\theta)}{f(x;\theta)}\cdot f(x;\theta)dx \\ &= \mathbb{E}_\theta\left(h(x)\cdot\frac{\partial}{\partial\theta}\log f(x;\theta)\right)\end{aligned}$$

Easy to check that $\mathbb{E}_\theta\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right) = 0$, So

$$= \operatorname{Cov}_\theta\left(h(x), \frac{\partial}{\partial\theta}\log f(x;\theta)\right)$$

then in short we have

$$\begin{aligned}\left[\frac{d}{d\theta}\mathbb{E}_\theta(h(x))\right]^2 &= \operatorname{Cov}_\theta\left(h(x), \frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2 \\ &\leqslant \operatorname{var}(h(x))\cdot\operatorname{var}\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right) \\ &= \operatorname{var}(h(x))\cdot\mathbb{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right) \\ &= \operatorname{var}(h(x))\cdot\mathcal{I}(\theta)\end{aligned}$$

如果可以证明 $h(x)$ 是无偏估计量，则化简为

$$\operatorname{Var}(h(x)) \geq \frac{1}{\mathcal{I}(\theta)}$$

注意上述的 lemma 对于 sample 是多个的时候也适用，推导过程说明了当 n 个 sample 是 i.i.d 的时候 lower bond 中的分子是没有变化的，因为假设条件中的最后一个条件可以直接改为

$$\frac{\partial}{\partial\theta}\mathbb{E}_\theta(h(\boldsymbol{x})) = \int h(\boldsymbol{x})\frac{\partial}{\partial\theta}f(\boldsymbol{x};\theta)d\boldsymbol{x}$$

发生变化的只有分子，所以 lemma 延伸为

Theorem 4.8 *(Carmer-Rao Lower bond Multiple i.i.d Samples) Let $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ be vector of i.i.d samples and the estimator is based on $\boldsymbol{X}$. Then we have*

$$\operatorname{Var}(h(\boldsymbol{x})) \geq \frac{1}{n\mathcal{I}(\theta)}\left[\frac{d}{d\theta}\mathbb{E}_\theta(h(\boldsymbol{x}))\right]^2 \tag{13}$$

Proof. Only shows the denominator is $n\mathcal{I}(\theta)$. Since X_i are i.i.d so

$$f(\boldsymbol{x};\theta) = \prod_{i=1}^n f(x_i;\theta)$$

then we have the expectation of square of score function to be

$$\begin{aligned}\mathbb{E}_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(\boldsymbol{x};\theta)\right)^2\right) &= \mathbb{E}_\theta\left(\left(\sum_{i=1}^n\frac{\partial}{\partial\theta}\log f(x_i;\theta)\right)^2\right)\\ &= \sum_{i=1}^n\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(x_i;\theta)\right)^2\right]\end{aligned}$$

the last equation is from directly expand the square of sum which is

$$\left(\sum_{i=1}^n\frac{\partial}{\partial\theta}\log f(x_i;\theta)\right)^2 = \sum_{i=1}^n\left(\frac{\partial}{\partial\theta}\log f(x_i;\theta)\right)^2 + \sum_{i\neq j}\left(\frac{\partial}{\partial\theta}\log f(x_i;\theta)\right)\left(\frac{\partial}{\partial\theta}\log f(x_j;\theta)\right)$$

For the crossing term, since they are independent, the covariance = 0. Also we have shown their expectation are 0 for each x, so the expectation of product is 0.

Recall that for MLE we have an asymptotic behavior which is

$$\sqrt{n}(\widehat{\theta}_n - \theta) \longrightarrow_d N(0, \mathcal{I}^{-1}(\theta))$$

however we should know that variance does not have such a behavior, $\operatorname{Var}\left(\sqrt{n}\widehat{\theta}_n\right) \longrightarrow_d \mathcal{I}^{-1}(\theta)$ Check this.

CR lower bound 说明如果 estimator 的 variance 证明后等于这个 lower bound 那么这个 estimator 是具有最小 variance 的 estimator。如果进一步，这个 estimator 是无偏的，那么由于 MSE = Bias + Variance，所以它就是 UMVUE。

Example 4.2 *Let* $x_1, x_2 \ldots x_n \sim$ *Poisson(*λ*) i.i,d,* $f(x;\lambda) = \lambda^x e^{-\lambda}/x!$ *First compute* $\mathcal{I}(\theta)$*, easy to get* $\mathcal{I}(\lambda) = 1/\lambda$*. let* $\widehat{\lambda}_n = \overline{X}_n$ *It's variance is*

$$\operatorname{Var}\left(\hat{\lambda}_n\right) = \frac{1}{n^2} \cdot n \cdot \operatorname{Var}(x_i) = \frac{\lambda}{n}$$

The CR lower bound is

$$\operatorname{Var}\left(\hat{\lambda}_n\right) \geqslant \frac{1}{n\mathcal{I}(\lambda)} = \frac{\lambda}{n}$$

So its variance reach the lower bound. $\widehat{\lambda}_n$ *is UMVUE*

5 Hypothesis Testing

统计推断可以概述为三个部分, 都是从样本推断整体

- **Estimates:** 或者说是 point estimate，即通过 sample 估计参数的一个值
- **Inference:** 有了 point estimate，我们希望给这个 r.v 一个区间来衡量精确程度，所以 assign 给它一个概率分布如 CLT 等等渐进结论
- **Hypothesis test:** inference 给出的概率分布往往是不对的，所以引入假设检验来回答 yes or no 这个问题

Hypothesis test 包含两部分，一是 Hypothesis

$$H_0 : \theta \in \Theta_0$$
$$H_a : \theta \in \Theta_a$$

而是 decision rule

$$\phi(\mathbf{X}) = \begin{cases} \text{yes (reject) or 1,} & \text{if...} \\ \text{no (not reject) or 0,} & \text{otherwise} \end{cases}$$

这里的 $\mathbf{X}$ 是 sample，即可以看出 decision rule 是 random 的（因为 sample 是随机的）。自然的存在两种错误，弃真（type I) 和取伪（type II）错误。Let's take the 01 definition of the decision rule.

Definition 5.1 *(Power function) Let $\phi(X)$ be a statistics, and θ_0 and θ_1 are the null and alternative hypothesis parameter. Then*

$$\beta(\phi, \theta) = \mathbb{E}_\theta(\phi(x))$$

is called the power function.

注意 power function 可以看做是关于 θ 的函数，当带入 θ_0 时他表示的就是一类错误的概率，带入 θ_1 时表示的则是 power。

Definition 5.2 *(Size of a Test) Let β be power function. The size is the probability of committing type I error, so it is*

$$\alpha = \sup_{\theta \in \Theta_0} \beta(\phi(X), \theta)$$

If simple test, then size equals significance level?

5.1 Uniformly Most Powerful Test (UMP)

Definition 5.3 *(UMP) A test ϕ_* of size α is uniformly most powerful (UMP) test if and only if*

$$\beta(\phi^*, \theta) \geq \beta(\phi, \theta)$$

for all $\theta \in \Theta_a$ and ϕ of level α.

即在任意的 $\theta \in \Theta_a, \phi_*$ 都使得 power 最大，大致可以理解为

$$\beta(\phi^*, \theta) = \sup_{\phi} \beta(\phi, \theta), \quad \forall \theta \in \Theta_a$$

由于两种 error 不能被同时 minimize，所以我们通常给其中一个 error 设定一个小的上限，然后通过某种方法 minimize 另一种 error，如我们控制第一类错误

$$\sup_{\theta \in \Theta_0} \beta(\phi(X), \theta) \leq \alpha$$

where α is a given level. 如果 null 只包含一个 population 那么 size 就等于 α. 下面的 N-P 引理即基于该逻辑。

Theorem 5.1 *(Neyman-Pearson Lemma) Suppose that $\mathcal{P}_0 = \{P_0\}$ and $\mathcal{P}_1 = \{P_1\}$. Let f_j be the p.d.f. of P_j w.r.t. a σ-finite measure ν (e.g., $\nu = P_0 + P_1$), $j = 0, 1$.*

1. *(Existence of a UMP test). For every α, there exists a UMP test of size α, which is equal to*

$$T_*(X) = \begin{cases} 1 & f_1(X) > c f_0(X) \\ \gamma & f_1(X) = c f_0(X) \\ 0 & f_1(X) < c f_0(X) \end{cases}$$

where $\gamma \in [0, 1]$ and $c \geq 0$ are some constants chosen so that $E[T_*(X)] = \alpha$ when $P = P_0$ ($c = \infty$ is allowed).

2. *(Uniqueness). If T_{**} is a UMP test of size α, then*

$$T_{**}(X) = \begin{cases} 1 & f_1(X) > c f_0(X) \\ 0 & f_1(X) < c f_0(X) \end{cases} \quad a.s.\ P.$$

Note

- 此处的定义摘自 *Jun Shao 6.1.1*，其中 P 是一个由其参数 *index* 的 *population*，$X \subset P \in \mathcal{P} = \mathcal{P}_0 \cup \mathcal{P}_1$ 要 *test* 的是 p 属于 $\mathcal{P}_0$ 还是 $\mathcal{P}_1$。这种 *test* 是 *simple test*。
- 该定理的用处在于如果可以找到这样的 c 和 γ，则我们就可以确定，这个 *test* 是 *UMP*。其方法是计算 $\mathbb{E}[T_*(X)] = \alpha$ 来解 c。
- 其中设定 $E[T_*(X)] = \alpha$ 就是在控制第一类错误来 *maximize power*。
- *a.s* P 的意思是 $\mathbb{P}(\phi(X) = 1 \text{ or } 0) = 13$。注意这里的 *a.s* 并不是渐进理论中的 *a.s convergence*，这里的 n，*sample size* 是 *fixed* 的。
- 注意，这个 *test* 的形式是永远不变的，其拒绝规则永远是 $f_1/f_0 > c$ 大于某个值。需要变的是其 *ratio* 的增减性要与 *test* 的方向一致。参见例题 *Assignment1,Q6.*

Example 5.1 *Consider simple test*

$$H_0 : \theta = 1 \quad H_a : \theta = 2$$

The NP test, which is UMP, will be the same as

$$H_0 : \theta \leq 1 \quad H_a : \theta > 1$$

We should also make sure that

$$\sup_{\theta \in \Theta_0} \beta(\phi(X), \theta) = \alpha$$

that is $\beta(\phi, \theta)$ is non-decreasing in θ

Definition 5.4 *(Monotonicity Likelihood Ratio for One-sided Test) Let $f(x, \theta)$, $\theta \in \Theta \subseteq \mathbb{R}$ with hypothesis to be*

$$H_0 : \theta \leq \theta_0 \ \ H_a : \theta > \theta_0$$

this is a one sided hypothesis test. If, $\forall \theta_1 < \theta_2$, we have

$$\frac{f(x, \theta_2)}{f(x, \theta_1)} = b(T(x))$$

as a monotone non-decreasing function in T.

Intuition: 其目的是为了制定 rejection 的规则，即需要量化标准什么时候要拒绝原假设。Likelihood ratio 是 alternative 比 null，其比值越大越接近备择假设的 population，当大到一定程度我们就要拒绝原假设即 reject。

Example 5.2 *Let $\theta_2 > \theta_1$. For exponential family*

$$\frac{f(x, \theta_2)}{f(x, \theta_1)} = \frac{b(\theta_2)}{b(\theta_1)} \exp\{T(x)(c(\theta_2) - c(\theta_1))\}$$

so the family has monotonicity likelihood ratio property as long as the function $c(\cdot)$ is nondecreasing in θ.

Then we extend the NP test to those with MLR property.

Proposition 5.2 *Let $f(x, \theta)$ has monotone likelihood ratio then $\forall h(\cdot)$ which is monotone non-decreasing, then*

$$g(\theta) = \mathbb{E}_\theta [h(T(X))]$$

is monotone and non-decreasing in θ.

Using the above property we can extend the NP test to be

Theorem 5.3 *If the function $f(x, \theta)$ has the monotone likelihood ratio property, then it can be written as*

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

and $\beta(\phi, \theta)$ is monotone in θ.

即如果对 ratio 这个 function 我们能找到一个关于他的 statistics 的单调函数，那么 test 可以化简为直接关于这个 statistics 的 test。总结下 One-Sided test，即可以用 NP Lemma 的类型：假设 $\theta_a > \theta_0$

- Simple test

$$H_0 : \theta = \theta_0, \quad H_a : \theta = \theta_a$$

- Same side as the direction of θ_a and θ_0

$$H_0 : \theta = \theta_0, \quad H_a : \theta > \theta_0$$

- Extend the null

$$H_0 : \theta \leq \theta_0, \quad H_a : \theta > \theta_0$$

以上三种假设最终得到的 UMP test 是一样的。下面讨论 two sided 的情况。先总结 two-sided hypothesis test 的形式。Let $\theta_1 < \theta_2$,

$$H_0 : \theta \notin (\theta_1, \theta_2) \ v.s \ H_a : \theta \in (\theta_1, \theta_2) \tag{14}$$

$$H_0 : \theta \in [\theta_1, \theta_2] \ v.s \ H_a : \theta \notin [\theta_1, \theta_2] \tag{15}$$

$$H_0 : \theta = \theta_0 \ v.s \ H_a : \theta \neq \theta_0 \tag{16}$$

下面只讨论 exponential family 的 test 情况。

5.2 UMP and UMPU for Exponential Family

所有的 UMP 都是 UMPU，但有时候 UMP 不存在。因此我们可以在 unbiased 的 test 中找到 UMP 的 test，称为 UMPU test。这个逻辑与我们再找 UMVUE 时是一样的：当 general 的最好的 test 不存在时，我们 impose 一些条件，在这个条件下我们寻找最好的 test。

Definition 5.5 *(Unbiasedness of a Test) The test ϕ satisfying the following test*

$$\beta(\phi, \theta) \leq \alpha \quad \forall \theta \in \Theta_0$$

$$\beta(\phi, \theta) \geq \alpha \quad \forall \theta \in \Theta_a$$

is said to be unbiased. So any UMP is UMPU.

Definition 5.6 *(Similar Test) Consider hypothesis test*

$$H_0 : \theta \in \Theta_0 \quad H_a : \theta \in \Theta_1$$

Let $\overline{\Theta}_{01} = \Theta_0 \cap \Theta_1$. A test ϕ is similar on $\overline{\Theta}_{01}$ if and only if

$$\beta(\phi, \theta) = \alpha, \quad \forall \theta \in \overline{\Theta}_{01}$$

(The intersection is not accurate. It can be boundary points of the two sets)

Definition 5.7 *(Natural Exponential Family, Shao §2.1.3) In lecture it is given as*

$$f(x;\theta,\phi) = a(x)b(\theta,\phi)\exp\{\theta T(x) + \phi' U(x)\}$$

where $\theta \in \mathbb{R}$, $\phi \in \mathbb{R}^p$. We focus on a subset of the natural exponential family

$$f(x;\theta) = a(x)b(\theta)\exp\{\theta T(x)\}$$

which is one-parameter.

Theorem 5.4 *(UMPU Test for One-parameter NATURAL Exponential Family) Let $f(x,\theta)$ be the pdf of single parameter exponential family, that is $\phi = 0$ in definition 5.7.*

1. *Let the hypothesis to be*

$$H_0 : \theta \in [\theta_1, \theta_2] \quad H_a : \theta \notin [\theta_1, \theta_2]$$

then UMPU test of size α is

$$\phi(T(x)) = \begin{cases} 1 & T < K_1 \text{ or } T > K_2 \\ \gamma_i & T = K_i,\ i = 1,2 \\ 0 & K_1 < T < K_2 \end{cases}$$

where

$$\alpha = \mathbb{E}_{\theta_1}[\phi(T)] = \mathbb{E}_{\theta_2}[\phi(T)]$$

2. *Let the hypothesis to be*

$$H_0 : \theta \notin [\theta_1, \theta_2] \quad H_a : \theta \in [\theta_1, \theta_2]$$

the interval can be open. Then the UMPU test of level α is

$$\phi(T(x)) = \begin{cases} 1 & K_1 < T < K_2 \\ \gamma_i & T = K_i,\ i = 1,2 \\ 0 & K_1 \text{ or } T > K_2 \end{cases}$$

where

$$\alpha = \mathbb{E}_{\theta_1}[\phi(T)] = \mathbb{E}_{\theta_2}[\phi(T)]$$

3. *Let the hypothesis to be*

$$H_0 : \theta = \theta_0 \quad H_a : \theta \neq \theta_0$$

the interval can be open. Then the UMPU test of level α is

$$\phi(T(x)) = \begin{cases} 1 & T < K_1 \text{ or } T > K_2 \\ \gamma_i & T = K_i \\ 0 & K_1 < T < K_2 \end{cases}$$

where

$$\alpha = \mathbb{E}_{\theta_0}[\phi(T)]$$

and

$$\mathbb{E}_{\theta_0}(T\phi(T)) = \alpha\mathbb{E}_{\theta_0}(T)$$

Note 对于假设 (15) 和 (16) 他们 test 形式一样，其原因是可以将 $\theta = \theta_0$ 看作是一个 interval 长度趋近于 0 的渐进行为。同时注意，做 test 的时候对于其中的 $a(x)$ 的部分无需担心，因为 theorem 确保了 UMPU 是关于 T 的 test。

在实际计算 critical value 和 γ 的时候往往经常由于无法找到关于 T 统计量的分布而无法理论计算，因此需要 simulation。具体的 simulation 方法如下：首先从给定的 pdf 中 pdf $f(x;\theta)$ 中计算出 CDF $F(\cdot)$. 由于 F 中带入的是概率且 X 是连续型变量，所以我们从 $Y \sim \mathcal{U}([0,1])$ 中 generate 出一系列值并带入到 $F(\cdot)$ 中就得到了一系列的 X 值。This is in short

$$Y = F_\theta^{-1}(\mathcal{U}) \longrightarrow Y \sim F_\theta$$

If X is not continuous, then need to define an inverse that accounting the max of the uniform r.v

Example 5.3 *Jan 31st Video*

5.3 Likelihood Ratio Test

LRT can be thought as an extension of NP lemma. If we know there exists an optimal test (i.e UMP, UMPU) then the LHR will coincides with the optimal test.

Definition 5.8 *(Generalized Likelihood Ratio Test) Let the hypothesis to be*

$$H_0 : \theta \in \Theta_0 \quad H_a : \theta \in \Theta \backslash \Theta_0$$

The generalized likelihood ratio is

$$\lambda(X) = \frac{\sup_{\theta\in\Theta_0} \ell(X;\theta)}{\sup_{\theta\in\Theta} \ell(X;\theta)}$$

Then reject H_0 if

$$\lambda(X) < c$$

for some $c \in [0,1]$ and satisfies

$$\sup_{\theta\in\Theta_0} P_{\theta_0}(\lambda(X) < c) = \alpha \tag{17}$$

其逻辑是从 NP 引理延伸而来，详情见 Shao 428 页。

Note *The Θ_0 and Θ_a can be any form and not have to be simple hypothesis. If they contains only one point, then it becomes exactly the NP test. Also if the $\lambda(X)$ is well defined then $\lambda(X) \leq 1$ since $\Theta_0 \subseteq \Theta$. In this test, the all possible θ must full in one of Θ_0 and Θ_a. Even when c.d.f of $\lambda(X)$ is continuous or randomized LR test are introduced, such an c satisfying (17) still does no exist. When a UMP and UMPU test exists, the LR test is often same as the optimal test.*

Example 5.4 *(LR test for Uniform Distribution) Let $X_1, ..., X_n$ i.i.d from $\mathcal{U}([0,\theta])$ with $\theta > 0$. We want to test*

$$H_0 : \theta = \theta_0 \quad H_a : \theta \neq \theta_0$$

Since uniform is not a member of exponential family, so the previous UMPU result is not applicable (but we are not sure if there exists an UMP test or not, there may exists). So we do LR test here.

In this case, $\Theta_0 = \{\theta_0\}$ and $\Theta = \mathbb{R}^+$. Usually the LR test require maximize the likelihood twice, once for the numerator (restricted under Θ_0) and once for the denominator (unrestricted). Here since Θ_0 only contains one point so the supreme is attained at that single point. The likelihood function here is

$$\begin{aligned}\ell(\theta) = \ell(X;\theta) &= \frac{1}{\theta^n}\prod_{i=1}^{n}\mathbb{I}_{[0,\theta]}(X_i)\\ &= \frac{1}{\theta^n}\mathbb{I}_{(0,\infty)}(X_{(1)}) * \mathbb{I}_{(-\infty,\theta]}(X_{(n)})\\ &= \frac{1}{\theta^n}\mathbb{I}_{(0,\infty)}(X_{(1)}) * \mathbb{I}_{[X_{(n)},\infty)}(\theta)\end{aligned}$$

The last step is to rewrite the expression to have the likelihood be a function of parameter. It is easy to see the function is monotonically decreasing with θ. So

$$\widehat{\theta}_{MLE} = X_{(n)} = \max\{X_1, ..., X_n\}$$

for uniform distribution. Then we have

$$\lambda(X) = \frac{\ell(X;\theta_0)}{\ell(X;\widehat{\theta}_{MLE})} = \frac{(1/\theta_0)^n\mathbb{I}_{[X_{(n)},\infty)}(\theta_0)}{(1/X_{(n)})^n * 1} = \left(\frac{X_{(n)}}{\theta_0}\right)^n \mathbb{I}_{[X_{(n)},\infty)}(\theta_0)$$

Then find c s.t

$$P_{\theta_0}\left(\left(\frac{X_{(n)}}{\theta_0}\right)^n \mathbb{I}_{[X_{(n)},\infty)}(\theta_0) \leq c\right) = \alpha$$

This is where we control the type one error as before. Let

$$G_{X(n)} = \left(\frac{X_{(n)}}{\theta_0}\right)^n \mathbb{I}_{[X_{(n)},\infty)}(\theta_0)$$

(notice $X_{(n)}$ is where randomness comes from). Then the rejection region becomes

$$\left\{X_{(n)} : G(X_{(n)}) \leq c\right\} = \left[0, G^{-1}(c)\right] \cup [\theta_0, \infty) = \left[0, c^{1/n}\right] \cup [\theta_0, \infty)$$

then

$$\underbrace{P_{\theta_0}(0 \leq X_{(n)} \leq \theta_0 c^{1/n})}_{=c} + \underbrace{P_{\theta_0}(X_{(n)} \geq \theta_0)}_{=0} = \alpha$$

so $c = \alpha$.

Example 5.5 *(Normal mean hypothesis)*

Lemma 5.5 *(Theorem 6.5 from Shao) The following quantity has an asymptotic property*

$$-2\log\lambda_n \longrightarrow_d \chi_r^2 \tag{18}$$

where r is the dimension of known parameter, $\lambda_n = \lambda(X)$ if the following regular conditions holds:

- *Θ is an open set*
- *$f(x;\theta)$ is twice differentiable*
- *Exchangeable integral and differentiable operator*
- *$\mathcal{I}_1(\theta)$ is positive definite*
- *The following inequality hold for any ϵ*

$$\sup_{||\gamma-\theta||\leq\epsilon} \left\| \frac{\partial^2}{\partial\theta\partial\theta'} \log f(x;\theta)\Big|_{\theta=\gamma} \right\| \leq h_\theta(x) \in L^1$$

Lemma 5.5 的意义在于，当一个 likelihood ratio 的分布极其繁琐难以找到时，我们可以用其渐进分布以近似。其证明主要逻辑是

$$\frac{1}{\sqrt{n}}\mathcal{I}_1^{-1}(\theta)S_n(\theta) \approx \sqrt{n}(\widehat{\theta}_n - \theta)$$

约等于指 same asymptotic distribution。则有

$$\frac{1}{\sqrt{n}}S_n(\theta) = \mathcal{I}_1(\theta)\sqrt{n}(\widehat{\theta}_n - \theta) + o_p(1)$$

So the lemma tells us to reject H_0 if

$$-2\log\lambda(X) > q$$

where q is some quantile of χ_r^2 under $P(-2\log\lambda(X) > q) = \alpha$.