Stat210B: Theoretical Statistics **Lecture Date: April 17, 2007**

Lecture 26 – Optimality of Estimators

Lecturer: Michael I. Jordan *Scribe: Kasper Daniel Hansen*

1 Overview

We are considering the following problem: we are trying to estimate $\psi(\theta)$ for a model parametrized by θ using a given sequence of estimators (T_n), and we are interested in assessing whether the sequence of estimators are optimal in some sense.

Because of the example set forth by Hodges estimator, we do not study pointwise convergence of $\sqrt{n}(T_n - \psi(\theta))$. Instead we assume we can obtain the following limit result

$$\sqrt{n}\Big(T_n - \psi(\theta + \frac{h}{\sqrt{n}})\Big) \xrightarrow{\theta + \frac{h}{\sqrt{n}}} L_{\theta,h}$$

and we wish to utilize this limit result to say something about the optimality of the sequence (T_n).

We hence need to consider two aspects

- In what sense is a limit distribution $L_{\theta,h}$ "good".
- In what sense is it enough to merely study the limit distribution, when trying to assess the properties of the sequence.

This lecture corresponds to sections 8.1, 8.3-8.7 in (van der Vaart, 1998)

2 The Setup

Theorem 1. *Assume the following:*

- *Let $(P_\theta : \theta \in \Theta)$ be a statistical model which is QMD at a point θ.*
- *Assume that the model has a non-singular Fisher information $I(\theta)$ at θ.*
- *Assume that the function $\psi : \Theta \mapsto \mathcal{R}^j$ is differentiable at θ.*

Let T_n be an estimator in the experiment $(P^n_{\theta + \frac{h}{\sqrt{n}}} : h \in \mathcal{R}^k)$ such that

$$\sqrt{n}\Big(T_n - \psi(\theta + \frac{h}{\sqrt{n}})\Big) \xrightarrow{\theta + \frac{h}{\sqrt{n}}} L_{\theta,h}$$

holds for every h, then there exists a randomized statistic T in the limit experiment $(N(h, I(\theta)^{-1}), h \in \mathcal{R}^k)$ such that

$$T - \dot{\psi} h \sim L_{\theta,h}$$

Proof. Apply van der Vaart (1998, Thm 7.1) to the sequence

$$\sqrt{n}\Big(T_n - \psi(\theta)\Big) = \sqrt{n}\Big(T_n - \psi(\theta + \frac{h}{\sqrt{n}})\Big) - \underbrace{\sqrt{n}(\psi(\theta + \frac{h}{\sqrt{n}}) - \psi(\theta))}_{\to \dot{\psi} h}$$

From the theorem we get a randomized statistic T such that

$$\sqrt{n}\Big(T_n - \psi(\theta)\Big) \xrightarrow{\theta + \frac{h}{\sqrt{n}}} T$$

But T must have the same distribution as $L_{\theta,h} + \dot{\psi} h$, hence $T - \dot{\psi} h \sim L_{\theta,h}$. □

The following is a heuristic outline of what we want to do: under the assumptions of the theorem, the distribution of

$$\sqrt{n}\Big(T_n - \psi(\theta + \frac{h}{\sqrt{n}})\Big)$$

is approximately equal to the distribution of

$$T - \dot{\psi} h$$

This means that if we find a good estimator T of $\dot{\psi} h$ in the limit experiment we should be able to say that T_n is a good estimator of $\psi(\theta + \frac{h}{\sqrt{n}})$.

One candidate for a best estimator in the limit experiment is $\dot{\psi} X$, where X is an observation from $N(h, I(\theta)^{-1})$, because

- It is unbiased, and minimum variance amongst the unbiased estimators.
- It is best equivariant
- It is best minimax.

(see below for a discussion of the properties).

Now the best estimator for $\dot{\psi} h$ has the following distribution

$$\dot{\psi} X \sim N(\dot{\psi} h, \dot{\psi}(\theta) I(\theta)^{-1} \dot{\psi}(\theta)^T)$$

This must be the best limit distribution for a given h. If we set $h = 0$ in the equation above, we recover the original θ. We conclude, heuristically, that the best limit distribution for $\sqrt{n}(T_n - \psi(\theta))$ is $N(0, \dot{\psi}(\theta) I(\theta)^{-1} \dot{\psi}(\theta)^T)$.

3 Properties of $\dot{\psi} X$

The following section is devoted to investigate the properties of $\dot{\psi} X$ in the limit experiment and to show that for a number of criterias, this is indeed the best estimator.

More precisely we will study estimators of Ah in the experiment $N(h, \Sigma)$ with Σ being non-singular and A being some matrix.

3.1 Equivariance

Definition 2. An estimator T for Ah is called *equivariant in law* if the distribution of $T - Ah$ does not depend on h

Remark 3. Since $\dot{\psi}X - \dot{\psi}h \sim N(0, \dot{\psi}(\theta)I(\theta)^{-1}\dot{\psi}(\theta)^T)$, the estimator $\dot{\psi}X$ is equivariant in law for estimating $\dot{\psi}h$.

We will now try to characterize the *best* equivariant in law estimator.

Theorem 4. *The distribution of any equivariant in law estimator of Ah can be decomposed as $L = N(0, A\Sigma A^T) * M$ for some probability measure M. The only estimator for which $M = \delta_0$ is AX.*

Remark 5. If $M = \delta_0$ then $L = N(0, A\Sigma A^T)$.

A bowl shaped loss function ℓ is a function with values in $[0, \infty]$ such that the sublevel sets $\{x : \ell(x) \le c\}$ are convex and symmetric about the origin.

Theorem 6. *For any bowl shaped loss function ℓ on $\mathcal{R}^k$, every possible probability measure M and any covariance matrix Σ, we have*

$$\int \ell \, dN(0, \Sigma) \le \int \ell \, d(N(0, \Sigma) * M)$$

The first theorem tell us that the law of an equivariant in law estimator is equal to a certain kind of convolution. The second theorem tells us that "convolution decreases concentration", with concentration measured by a loss function. Taken together, the theorems tells us that the best (in concentration sense) distribution is $N(Ah, A\Sigma)$, which is only obtained by the estimator AX. The conclusion is that AX is the best equivariant in law estimator.

3.2 Mimimax

An estimator T is minimax (relative to a given loss function) for the parameter Ah if it minimizes the maximum risk

$$\sup_h E_h \ell(T - Ah)$$

Since AX is equivariant in law, the distribution of $\ell(AX - Ah)$ does not depend on h, hence $E_h \ell(AX - Ah) = E_0 \ell(AX)$ or

$$\sup_h E_h \ell(AX - Ah) = E_0 \ell(AX)$$

We now give a lower bound on the maximum loss

Theorem 7. *For any bowl shaped loss function ℓ, the maximum risk of any estimator is bounded below by $E_0 \ell(AX)$. Consequently AX is a minimax estimator for Ah. If Ah is real and $E_0(AX)^2 \ell(AX) < \infty$ then AX is the only minimax estimator for Ah up to changes on sets of probability zero.*

3.3 Conclusion

We have shown that the best estimator of $\dot{\psi}h$ indeed is $\dot{\psi}X$ based on a variety of criterias.

4 Asymptotics

We now try to make the asymptotic part of the heuristic argument rigorously. We will consider two approaches (1) only consider sequences of estimators that are *regular* and (2) consider sequences that are locally minimax.

4.1 Regular estimators

Definition 8. A sequence of estimators (T_n) are called *regular* (or asymptotically equivariant in law) for estimating $\psi(\theta)$, if

$$\sqrt{n}\Big(T_n - \psi(\theta + \frac{h}{\sqrt{n}})\Big) \xrightarrow{\theta+\frac{h}{\sqrt{n}}} L_\theta$$

for every h.

Remark 9. The real property is that the limiting distribution L_θ does not depend on h.

Theorem 10. *Under the same assumptions of Theorem 1, if the sequence of estimators (T_n) are regular with limit distribution L_θ, then*

$$L_\theta = N(0, \dot{\psi}(\theta)I(\theta)^{-1}\dot{\psi}(\theta)^T) * M_\theta$$

for some probability measure M_θ.

Remark 11. Following the theorem above regarding convolutions and loss functions, we conclude that the best possible limit distribution is $N(0, \dot{\psi}(\theta)I(\theta)^{-1}\dot{\psi}(\theta)^T)$.

Proof. This is an easy consequence of the theorems above. Because of theorem 1 we have that there exists a randomized statistic T such that $T - \dot{\psi}h \sim L_{\theta,h} = L_\theta$. Hence T is equivariant in law for the limit experiment so we have the representation above by theorem 4. □

How does this assumption deal with Hodges estimator? Well, Hodges estimator is not regular, so by a priorily only considering regular estimators, we are by definition excluding Hodges estimator from being "good".

4.2 Local Minimax

Theorem 12. *Let the experiment $(P_\theta : \theta \in \Theta)$ be QMD at θ with non-singular Fisher information $I(\theta)$ and assume that ψ is differentiable at θ. Then, for any bowl-shaped loss function*

$$\sup_I \liminf_{n\to\infty} \sup_{h\in I} E_{\theta+\frac{h}{\sqrt{n}}}\,\ell\left(\sqrt{n}\Big(T_n - \psi(\theta + \frac{h}{\sqrt{n}})\Big)\right) \geq \int \ell\, dN(0, \dot{\psi}(\theta)I(\theta)^{-1}\dot{\psi}(\theta)^T)$$

with I belonging to the set of finite subsets of $\mathcal{R}^k$.

Proof. Deferred to next lecture □

Remark 13. This is one of the most famous theorems in the class and is considered a major accomplishment for the theory of optimal estimators.

References

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.