Statistical Inference
Lecture Five

https://people.bath.ac.uk/masss/APTS/apts.html

Simon Shaw

University of Bath

APTS, 14-18 December 2020

Overview of Lecture Five

In Lecture Four we considered a Bayesian statistical decision problem, $[\Theta, \mathcal{D}, \pi(\theta), L(\theta, d)]$.

- The risk of decision $d \in \mathcal{D}$ under the distribution $\pi(\theta)$ is $\rho(\pi(\theta), d) = \int_\theta L(\theta, d)\pi(\theta)\, d\theta$.
- A decision $d^* \in \mathcal{D}$ for which $\rho(\pi, d^*) = \rho^*(\pi)$ is a Bayes rule.
- The Bayes rule for the posterior decision respects the strong likelihood principle.

Today, we'll look at decision theory from a classical perspective.

- The classical risk for the model $\mathcal{E} = \{\mathcal{X}, \Theta, f_X(x \,|\, \theta)\}$ is

$$R(\theta, \delta) \;=\; \int_X L(\theta, \delta(x)) f_X(x \,|\, \theta)\, dx.$$

- A decision rule δ_0 is admissible if there is no decision rule δ_1 which dominates it.
- Wald's Complete Class Theorem, CCT: a decision rule is admissible if and only if it is a Bayes rule for some prior distribution.
- Admissible decision rules respect the SLP.

Admissible rules

- Bayes rules rely upon a prior distribution for θ: the risk is a function of d only.
- In classical statistics, there is no distribution for θ and so another approach is needed.

Definition (The classical risk)

For a decision rule $\delta(x)$, the classical risk for the model $\mathcal{E} = \{\mathcal{X}, \Theta, f_X(x \,|\, \theta)\}$ is

$$R(\theta, \delta) \;=\; \int_X L(\theta, \delta(x)) f_X(x \,|\, \theta)\, dx.$$

- The classical risk is thus, for each δ, a function of θ.

Example

Let $X = (X_1, \ldots, X_n)$ where $X_i \sim N(\theta, \sigma^2)$ and σ^2 is known. Suppose that $L(\theta, d) = (\theta - d)^2$ and consider a conjugate prior $\theta \sim N(\mu_0, \sigma_0^2)$. Possible decision functions include:

1. $\delta_1(x) = \overline{x}$, the sample mean.
2. $\delta_2(x) = \text{med}\{x_1, \ldots, x_n\} = \tilde{x}$, the sample median.
3. $\delta_3(x) = \mu_0$, the prior mean.
4. $\delta_4(x) = \mu_n$, the posterior mean where

$$\mu_n = \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1} \left(\frac{\mu_0}{\sigma_0^2} + \frac{n\overline{x}}{\sigma^2}\right),$$

the weighted average of the prior and sample mean accorded to their respective precisions.

Example - continued

The respective classical risks are

1. $R(\theta, \delta_1) = \frac{\sigma^2}{n}$, a constant for θ, since $\overline{X} \sim N(\theta, \sigma^2/n)$.
2. $R(\theta, \delta_2) = \frac{\pi\sigma^2}{2n}$, a constant for θ, since $\tilde{X} \sim N(\theta, \pi\sigma^2/2n)$ (approximately).
3. $R(\theta, \delta_3) = (\theta - \mu_0)^2 = \sigma_0^2 \left(\frac{\theta-\mu_0}{\sigma_0}\right)^2$.
4. $R(\theta, \delta_4) = \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-2} \left\{\frac{1}{\sigma_0^2}\left(\frac{\theta-\mu_0}{\sigma_0}\right)^2 + \frac{n}{\sigma^2}\right\}$.

Which decision do we choose? We observe that $R(\theta, \delta_1) < R(\theta, \delta_2)$ for all $\theta \in \Theta$ but other comparisons depend upon θ.

- The accepted approach for classical statisticians is to narrow the set of possible decision rules by ruling out those that are obviously bad.

Definition (Admissible decision rule)

A decision rule δ_0 is inadmissible if there exists a decision rule δ_1 which dominates it, that is

$$R(\theta, \delta_1) \ \leq \ R(\theta, \delta_0)$$

for all $\theta \in \Theta$ with $R(\theta, \delta_1) < R(\theta, \delta_0)$ for at least one value $\theta_0 \in \Theta$. If no such δ_1 exists then δ_0 is admissible.

- If δ_0 is dominated by δ_1 then the classical risk of δ_0 is never smaller than that of δ_1 and δ_1 has a smaller risk for θ_0.
- Thus, you would never want to use δ_0.[1]
- The accepted approach is to reduce the set of possible decision rules under consideration by only using admissible rules.

[1] Here I am assuming that all other considerations are the same in the two cases: e.g. for all $x \in \mathcal{X}$, $\delta_1(x)$ and $\delta_0(x)$ take about the same amount of resource to compute.

- We now show that admissible rules can be related to a Bayes rule δ^* for a prior distribution $\pi(\theta)$.

Theorem

If a prior distribution $\pi(\theta)$ is strictly positive for all Θ with finite Bayes risk and the classical risk, $R(\theta, \delta)$, is a continuous function of θ for all δ, then the Bayes rule δ^* is admissible.

Proof (Robert, 2007)

Letting $f(\theta, x) = f_X(x \mid \theta)\pi(\theta)$ we have

$$\begin{aligned}
\mathbb{E}\{L(\theta, \delta(X))\} &= \int_x \int_\theta L(\theta, \delta(x)) f(\theta, x)\, d\theta dx \\
&= \int_\theta \left\{ \int_x L(\theta, \delta(x)) f_X(x \mid \theta)\, dx \right\} \pi(\theta)\, d\theta \\
&= \int_\theta R(\theta, \delta)\pi(\theta)\, d\theta
\end{aligned}$$

Proof continued

- Suppose that the Bayes rule δ^* is inadmissible and dominated by δ_1.
- Thus, in an open set C of θ, $R(\theta, \delta_1) < R(\theta, \delta^*)$ with $R(\theta, \delta_1) \leq R(\theta, \delta^*)$ elsewhere.
- Consequently, $\mathbb{E}\{L(\theta, \delta_1(X))\} < \mathbb{E}\{L(\theta, \delta^*(X))\}$ which is a contradiction to δ^* being the Bayes rule. $\square$

- The relationship between a Bayes rule with prior $\pi(\theta)$ and an admissible decision rule is even stronger.
- The following result was derived by Abraham Wald (1902-1950)

Wald's Complete Class Theorem, CCT

In the case where the parameter space Θ and sample space $\mathcal{X}$ are finite, a decision rule δ is admissible if and only if it is a Bayes rule for some prior distribution $\pi(\theta)$ with strictly positive values.

- An illuminating blackboard proof of this result can be found in Cox and Hinkley (1974, Section 11.6).
- There are generalisations of this theorem to non-finite decision sets, parameter spaces, and sample spaces but the results are highly technical.
- We'll proceed assuming the more general result, which is that a decision rule is admissible if and only if it is a Bayes rule for some prior distribution $\pi(\theta)$, which holds for practical purposes.

So what does the CCT say?

1. Admissible decision rules respect the SLP. This follows from the fact that admissible rules are Bayes rules which respect the SLP. This provides support for using admissible decision rules.
2. If you select a Bayes rule according to some positive prior distribution $\pi(\theta)$ then you cannot ever choose an inadmissible decision rule.