Decision Theory: Part 4

Dr. Qiuzhuang Sun

STAT3023

Asymptotically minimax procedures

- We have so far only looked at examples of asymptotically minimax procedures using 0-1 loss, that is:

$$L(d|\theta) = 1\{|d - \theta| > C_n\},$$

  for which the risk is the non-coverage probability of the interval:

$$d(\mathbf{X}) \pm C_n.$$

- This has been because it is easier to derive limiting (maximum) risk for Bayes procedures using a uniform prior, since procedures are level sets of truncated densities, and only convergence in distribution is needed.

Asymptotically minimax procedures

- In particular, we showed the limiting risk of the Bayes procedure $\tilde{d}(\mathbf{X})$ using a uniform prior can be derived by
    1. deriving the limiting risk of $d_{\text{flat}}(\mathbf{X})$, the Bayes procedure using the flat prior $w(\theta) = 1$;
    2. showing that with probability tending to 1,

$$\tilde{d}(\mathbf{X}) = d_{\text{flat}}(\mathbf{X}).$$

- It turns out the same is true for the case of squared error loss. In many cases, the Bayes procedure (the posterior mean) using a UNIFORM prior is "close" to that obtained using a flat prior. In particular, the limiting (rescaled) risk for the two procedures is the same.

Asymptotically minimax procedures – squared error loss

Example (**Binomial with squared error loss**): Suppose $\mathbf{X} = (X_1, \ldots, X_n)$ consists of iid Binomial$(1, \theta)$ random variables for some unknown $\theta \in (0,1)$. Consider the decision problem with decision space $\mathcal{D} = (0,1)$ and loss function $L(d|\theta) = (d - \theta)^2$. Show that for any $\alpha_0 > 0, \beta_0 > 0$, the Bayes procedure using the conjugate prior

$$w(\theta) = \frac{\theta^{\alpha_0 - 1}(1-\theta)^{\beta_0 - 1}}{B(\alpha_0, \beta_0)}$$

is asymptotically minimax. You may use the fact that

$$\lim_{n\to\infty} nE_\theta[(\tilde{d}(\mathbf{X}) - \theta)^2] = \lim_{n\to\infty} nE_\theta[(d_{\mathsf{flat}}(\mathbf{X}) - \theta)^2], \quad \forall \theta \in (\theta_0, \theta_1),$$

where $\tilde{d}(\mathbf{X})$ and $d_{\mathsf{flat}}(\mathbf{X})$ are Bayes procedures under the $U(\theta_0, \theta_1)$ prior and flat prior $w(\theta) = 1$, $\theta \in (0,1)$, respectively.

Asymptotically minimax procedures – squared error loss

Asymptotically minimax procedures – squared error loss

Asymptotically minimax procedures – squared error loss

Asymptotically minimax procedures – squared error loss

Asymptotically minimax procedures – squared error loss

Asymptotically minimax procedures – squared error loss

Asymptotically minimax procedures – squared error loss

Asymptotically minimax procedures – squared error loss

Asymptotically minimax procedures – squared error loss

Asymptotically minimax procedures – squared error loss

Example (**Normal variance (with known mean)**): Suppose $\mathbf{X} = (X_1, \ldots, X_n)$ consists of iid $N(0, \theta)$ random variables for some unknown $\theta \in (0, \infty)$. Consider the decision problem with decision space $\mathcal{D} = (0, \infty)$ and the squared error loss. Show that both the MLE and the Bayes procedure under the Inverse Gamma (conjugate) prior:

$$w(\theta) = \frac{\lambda_0^{\alpha_0} e^{-\lambda_0/\theta}}{\theta^{\alpha_0+1}\Gamma(\alpha_0)},\ \theta > 0, \quad \text{for some known } \alpha_0, \gamma_0 > 0,$$

are asymptotically minimax. You may assume for any $0 < \theta_0 < \theta_1 < \infty$:

$$\lim_{n\to\infty} nE_\theta[(\tilde{d}(\mathbf{X}) - \theta)^2] = \lim_{n\to\infty} nE_\theta[(d_{\mathsf{flat}}(\mathbf{X}) - \theta)^2], \quad \forall \theta \in (\theta_0, \theta_1),$$

where $\tilde{d}(\mathbf{X})$ and $d_{\mathsf{flat}}(\mathbf{X})$ are Bayes procedures under the $U(\theta_0, \theta_1)$ prior and flat prior $w(\theta) = 1$, $\theta > 0$, respectively.

Asymptotically minimax procedures – squared error loss

Asymptotically minimax procedures – squared error loss

Asymptotically minimax procedures – squared error loss

Asymptotically minimax procedures – squared error loss

Asymptotically minimax procedures – squared error loss

Asymptotically minimax procedures – squared error loss

Asymptotically minimax procedures – squared error loss

Asymptotically minimax procedures – squared error loss

Asymptotically minimax procedures – squared error loss

Asymptotically minimax procedures – absolute error loss

Example (**Normal mean**): Suppose $\mathbf{X} = (X_1, \ldots, X_n)$ consists of iid $N(\theta, 1)$ random variables for some unknown $\theta \in \mathbb{R}$. Consider the decision problem with decision space $\mathcal{D} = \mathbb{R}$ and loss $L(d|\theta) = |d - \theta|$. Show $\bar{X}$ is asymptotically minimax.

Asymptotically minimax procedures – absolute error loss

Asymptotically minimax procedures – absolute error loss

Asymptotically minimax procedures – absolute error loss

Asymptotically minimax procedures – absolute error loss

Example (**Uniform scale parameter**): Suppose $\mathbf{X} = (X_1, \ldots, X_n)$ consists of iid $U(0, \theta)$ random variables for some unknown $\theta > 0$. Consider the decision problem with decision space $\mathcal{D} = \Theta = (0, \infty)$ and loss $L(d|\theta) = |d - \theta|$. Compare $d_{\mathsf{ML}}(\mathbf{X}) = \hat{\theta}_{\mathsf{ML}}$, which is the MLE and a "median unbiased" version of the MLE $d_{\mathsf{med}}(\mathbf{X}) = 2^{\frac{1}{n}} X_{(n)}$. Show $d_{\mathsf{med}}(\mathbf{X})$ is asymptotically minimax. You may assume for any $0 < \theta_0 < \theta_1 < \infty$:

$$\lim_{n\to\infty} nE_\theta[(\tilde{d}(\mathbf{X}) - \theta)^2] = \lim_{n\to\infty} nE_\theta[(d_{\mathsf{flat}}(\mathbf{X}) - \theta)^2], \quad \forall \theta \in (\theta_0, \theta_1),$$

where $\tilde{d}(\mathbf{X})$ and $d_{\mathsf{flat}}(\mathbf{X})$ are Bayes procedures under the $U(\theta_0, \theta_1)$ prior and flat prior $w(\theta) = 1$, $\theta > 0$, respectively.

Asymptotically minimax procedures – absolute error loss

Asymptotically minimax procedures – absolute error loss

Asymptotically minimax procedures – absolute error loss

Asymptotically minimax procedures – absolute error loss

Asymptotically minimax procedures – absolute error loss

Asymptotically minimax procedures – absolute error loss

Asymptotically minimax procedures – absolute error loss

Asymptotically minimax procedures – absolute error loss

Asymptotically minimax procedures – absolute error loss

Summary

- We have examined many examples with a variety of loss functions
    - Squared error loss
    - Absolute error loss
    - 0-1 loss (for interval estimation)

- In "regular" cases (usually exponential family), the MLE is generally different from Bayes estimators. However:
    - The differences are mainly in the bias, not the variance
    - The bias is asymptotically negligible compared to variance
    - Both MLE and Bayes estimators are optimal (in the sense that they are asymptotically minimax) under various loss functions

Summary

- In "non-regular" cases (e.g., $U(0,\theta)$), the MLE and Bayes estimators can be "more different"
  - Bias in the MLE is of the same order as the variance, and it is not asymptotically negligible
  - Bias-corrected versions of the MLE can be asymptotically minimax
  - Bayes estimators "asymptotically" adjust for the bias, and are asymptotically minimax
- Why use asymptotically minimax as a criterion for optimality?
  - Non-asymptotic optimality results are more difficult to establish
  - It gets around the "super-efficiency" problem
  - The AMLB theorem applies to any procedures. This is rather rare in statistics. Many optimality criteria apply to a restricted class of procedures, e.g., UMP tests to 1-parameter exponential family and 1-sided alternative