Bayes Procedures

MIT 18.655

Dr. Kempthorne

Spring 2016

Outline

Bayes Procedures

Decision Problem: Basic Components

- $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$: parametric model.
- $\Theta = \{\theta\}$: Parameter space.
- $\mathcal{A}\{a\}$: Action space.
- $L(\theta, a)$: Loss function.
- $R(\theta, \delta) = E_{X|\theta}[L(\theta, \delta(X))]$

Decision Problem: Bayes Components

- π : Prior distribution on Θ
- $r(\pi, \delta) = E_\theta[R(\theta, \delta)] = E_\theta[E_{X|\theta}[L(\theta, \delta(X))]] = E_{X,\theta}[L(\theta, \delta(X))]$
  "Bayes risk of the procedure $\delta(X)$ with respect to the prior π."
- $r(\pi) = \inf_{\delta \in \mathcal{D}} r(\pi, \delta)$: Minimum Bayes risk.
- δ_π : $r(\pi, \delta_\pi) = r(\pi)$
  " Bayes rule with respect to prior π and loss $L(\cdot, \cdot)$."

Bayes Procedures: Interpretations of Bayes Risk

Bayes risk

$$r(\pi, \delta) = \int_\Theta R(\theta, \delta)\pi(d\theta)$$

- $\pi(\cdot)$ weights $\theta \in \Theta$ where $R(\theta, \delta)$ matters.
- $\pi(\theta) = constant$: weights θ uniformly.
  Note: uniform weighting depends on parametrization.
- Interdependence of specifying the loss function and prior density:

$$\begin{aligned} r(\pi, \delta) &= \int_\Theta \int_{\mathcal{X}} [L(\theta, \delta(x))\pi(\theta)]p(x \mid \theta)dx\, d\theta. \\ &= \int_\Theta \int_{\mathcal{X}} [L^*(\theta, \delta(x))\pi^*(\theta)]p(x \mid \theta)dx\, d\theta. \end{aligned}$$

for $L * (\cdot, \cdot)$, $\pi^*(\cdot)$ such that

$$L^*(\theta, \delta(x))\pi^*(\theta) = L(\theta, \delta(x))\pi(\theta)$$

Bayes Procedures: Quadratic Loss

Quadratic Loss: Estimating $q(\theta)$ with $a \in \mathcal{A} = \{q(\theta), \theta \in \Theta\}$.

$$L(\theta, a) = [q(\theta) - a]^2.$$

- Bayes risk: $r(\pi, \delta) = E([q(\theta) - \delta(X)]^2)$
- Bayes risk as expected Posterior Risk:
$$\begin{aligned} r(\pi, \delta) &= E_X(E_{\theta|X} L(\theta, \delta(X))) \\ &= E_X(E_{\theta|X}([q(\theta) - \delta(x)]^2)) \end{aligned}$$
- Bayes decision rule specified by minimizing:
$$E_{\theta|X}([q(\theta) - \delta(x)]^2 \mid X = x)$$
for each outcome $X = x$, which is solved by
$$\begin{aligned} \delta_\pi(x) &= E_{\theta|X=x}[q(\theta) \mid X = x] \\ &= \frac{\int_\Theta q(\theta) p(x \mid \theta)\pi(\theta) d\theta}{\int_\Theta p(x \mid \theta)\pi(\theta) d\theta} \end{aligned}$$

Bayes Procedure: Quadratic Loss

Example 3.2.1 $X_1, \ldots, X_n$ iid $N(\theta, \sigma^2)$, $\sigma^2 > 0$, known.

Prior Distribution:

$$\pi : \theta \sim N(\eta, \tau^2).$$

Posterior Distribution:

$$\theta \mid X = x :\sim N(\eta_*, \tau_*^2)$$

where
$$\eta_* = [\frac{1}{\sigma^2/n}\overline{X} + \frac{1}{\tau^2}\eta]/[\frac{1}{\sigma^2/n} + \frac{1}{\tau^2}]$$
$$\tau_*^2 = [\tfrac{1}{\sigma^2/n} + \tfrac{1}{\tau^2}]^{-1}.$$

Bayes Procedure: $\delta_\pi(X) = E[\theta \mid x] = \eta_*$

Observations:

- Posterior risk: $E[L(\theta, \delta_\pi) \mid X = x] = \tau_*^2$ (constant!)
  $\Longrightarrow$ Bayes risk: $r(\pi, \delta_\pi) = \tau_*^2$.
- MLE $\delta_{MLE}(\mathbf{x}) = \bar{\mathbf{x}}$ has
  Constant Risk: $R(\theta, \overline{\mathbf{X}}) = \sigma^2/n$
  $\Longrightarrow$ BayesRisk: $r(\theta, \overline{\mathbf{X}}) = \sigma^2/n \ (> \tau_*^2)$
- $\lim_{\tau \to \infty} \delta_\pi(\mathbf{x}) = \bar{x}$ and $\lim_{\tau \to \infty} \tau_*^2 = \sigma^2/n$.

Bayes Procedure: General Case

Bayes Risk and Posterior Risk:

$$\begin{aligned} r(\pi,\delta) &= E_\theta[R(\theta,\delta(X)] = E_\theta[E_{X|\theta}[L(\theta,\delta(X))]] \\ &= E_X[E_{\theta|X}[L(\theta,\delta(X))]] \\ &= E_X[r(\delta(x) \mid x)] \end{aligned}$$

where

$$r(a \mid x) = E[L(\theta, a) \mid X = x] \qquad \text{(Posterior risk)}$$

Proposition 3.2.1 Suppose $\delta^* : \mathcal{X} \to \mathcal{A}$ is such that

$$r(\delta^*(x) \mid a) = \inf_{a \in \mathcal{A}}\{r(a \mid x)\}$$

Then δ^* is a Bayes rule.

Proof. For any procedure $\delta \in \mathcal{D}$,

$$\begin{aligned} r(\pi,\delta) &= E_X[r(\delta(x) \mid x)] \\ &\geq E_X[r(\delta^*(x) \mid x)] \\ &= r(\pi,\delta^*). \end{aligned}$$

Bayes Procedures for Problems With Finite Θ

Finite Θ Problem

- $\Theta = \{\theta_0, \theta_1, \ldots, \theta_K\}$
- $\mathcal{A} = \{a_0, a_1, \ldots, a_q\}$ (q may equal K or not)
- $L(\theta_i, a_j) = w_{ij}$, for $i = 0, 1, \ldots, K$, $j = 0, 1, \ldots, q$
- Prior distribution: $\pi(\theta_i) = \pi_i \geq 0$, $i = 1, \ldots, K$ ($\sum_0^K \pi_i = 1$).
- Data/Random variable: $X \sim P_\theta$ with density/pmf $p(x \mid \theta)$.

Solution:

- Posterior probabilities:

$$\pi(\theta_i \mid X = x) = \frac{\pi_i p(x \mid \theta_i)}{\sum_{j=0}^K \pi_j p(x \mid \theta_j)}$$

- Posterior risks:

$$r(a_j \mid X = x) = \frac{\sum_{i=0}^K w_{ij} \pi_i p(x \mid \theta_i)}{\sum_{i=0}^K \pi_i p(x \mid \theta_i)}$$

- Bayes decision rule: $\delta^*(x)$ satisfies

$$r(\delta^*(x) \mid x) = \min_{0 \leq j \leq K} r(a_j \mid x)$$

Finite Θ Problem: Classification

Classification Decision Problem

- $p = q$, identify $\mathcal{A}$ with Θ
- Loss function:

$$L(\theta_i, a_j) = w_{ij} = \begin{cases} 1 & \textit{if} \quad i \neq j \\ 0 & \textit{if} \quad i = j \end{cases}$$

- Bayes procedure minimizes Posterior Risk

$$\begin{aligned} r(\theta_i \mid x) &= P[\theta \neq \theta_i \mid x] \\ &= 1 - P[\theta = \theta_i \mid x] \end{aligned}$$

$\Longrightarrow \delta^*(x) = \theta_i \in \mathcal{A}$, that maximizes $P[\theta = \theta_i \mid x]$.

Special case: Testing Null Hypothesis vs Alternative

- $p = q = 1$
- $\pi_0 = \pi$, $\pi_1 = 1 - \pi_0$
- Testing $\Theta_0 = \{\theta_0\}$ versus $\Theta_1 = \{\theta_1\}$
- Bayes rule chooses $\theta = \theta_1$ if $P[\theta = \theta_1 \mid x] > P[\theta = \theta_0 \mid x]$.

Finite Θ Problem: Testing

Equivalent Specifications of Bayes Procedure: $\delta^*(x)$

- Minimizes
$$\begin{aligned} r(\pi, \delta) &= \pi R(\theta_0, \delta) + (1-\pi) R(\theta_1, \delta) \\ &= \pi P(\delta(X) = \theta_1 \mid \theta_0) + (1-\pi) P(\delta(X) = \theta_0 \mid \theta_1) \end{aligned}$$
- Chooses $\theta = \theta_1$ if
  - $P[\theta = \theta_1 \mid x] > P[\theta = \theta_0 \mid x]$
  - $(1-\pi) p(x \mid \theta_1) > \pi p(x \mid \theta_0)$
  - $\dfrac{p(x \mid \theta_1)}{p(x \mid \theta_0} > \pi/(1-\pi)$ (Likelihood Ratio)
  - $\dfrac{(1-\pi)}{\pi} \times \dfrac{p(x \mid \theta_1)}{p(x \mid \theta_0)} > 1$ (Bayes Factor)
- The procedure $\delta*$ solves:
$$\begin{array}{llll} \textit{Minimize}: & P(\delta(X) = \theta_0 \mid \theta_1) & & P(\textit{Type II Error}) \\ \textit{Subject to}: & P(\delta(X) = \theta_1 \mid \theta_0) \leq & \alpha & P(\textit{Type I Error}) \end{array}$$
i.e., minimizes the Lagrangian:
$$P(\delta(X) = \theta_1 \mid \theta_0) + \lambda P(\delta(X) = \theta_0 \mid \theta_1)$$
with Lagrange multiplier $\lambda = (1-\pi)/\pi$.

Estimating Success Probability With Non-Quadratic Loss

Decision Problem: $X_1, \ldots, X_n$ iid $Bernoulli(\theta)$

- $\Theta = \{\theta\} = \{\theta : 0 < \theta < 1\} = (0, 1)$
- $\mathcal{A} = \{a\} = \Theta$
- Loss equal to relative-squared-error:

$$L(\theta, a) = \frac{(\theta - a)^2}{\theta(1-\theta)}, \; 0 < \theta < 1 \text{ and } a \text{ real.}$$

Solving the Decision Problem

- By sufficiency, consider decision rules based on the sufficient statistic

$$S = \sum_1^n X_i \sim Binomial(n, \theta).$$

- For a prior distribution π on Θ, with density $\pi(\theta)$, denote the density of the posterior distribution by
$\pi(\theta \mid s) = [\pi(\theta)\theta^s(1-\theta)^{(n-s)}] / \int_\Theta [\pi(t)t^s(1-t)^{(n-s)}]dt$

Solving the Decision Problem (continued)

- The posterior risk $r(a \mid S = k)$ is

$$\begin{aligned} r(a \mid S = k) &= E[L(\theta, a) \mid S = k) \\ &= E[\frac{(\theta - a)^2}{\theta(1-\theta)} \mid S = k] \\ &= E[\frac{\theta}{(1-\theta)} - \frac{2}{1-\theta} a + \frac{a^2}{\theta(1-\theta)} \mid S = k] \\ &= E[\frac{\theta}{(1-\theta)} \mid k] - aE[\frac{2}{1-\theta} \mid k] + a^2 E[\frac{1}{\theta(1-\theta)} \mid k] \end{aligned}$$

  which is a parabola in a minimized at

$$a = \frac{E[\frac{1}{1-\theta} \mid k]}{E[\frac{1}{\theta(1-\theta)} \mid k]}$$

  This defines the Bayes rule $\delta^*(S)$ for $S = k$ (if the expectations exist).

- The Bayes rule can be expressed in closed form when the prior distribution is

  $\theta \sim Beta(r, s)$ has closed form solution

$$\delta^*(k) = \frac{\beta(r+k, n-k+s-1)}{\beta(r+k-1, n-k+s-1)} = \frac{(r+k-1)}{n+r+s-2}$$

- For $r = s = 1$, $\delta^*(k) = k/n = \overline{X}$ (for $k = 0$, $a = 0$ directly)

Bayes Procedures With Hierarchical Prior

Example 3.2.4 Random Effects Model

- $X_{ij} = \mu + \Delta_i + \epsilon_{ij}$, $i = 1, \dots, I$ and $j = 1, \dots, J$
  ϵ_{ij} are iid $N(0, \sigma_e^2)$.
- Δ_i iid $N(0, \sigma_\Delta^2)$ independent of the ϵ_{ij}
- $\mu \sim N(\mu_0, \sigma_\mu^2)$.

Bayes Model: Specification I

- Prior distribution on $\theta = (\mu, \sigma_e^2, \sigma_\Delta^2)$
  $\mu \sim N(\mu_0, \sigma_\mu^2)$, independent of σ_e^2 and σ_Δ^2.
  $\pi(\theta) = \pi_1(\mu)\pi_2(\sigma_e^2)\pi_3(\sigma_\Delta^2)$.
- Data distribution: $X_{ij} \mid \theta$ are jointly normal random variables with

$$E[X_{ij} \mid \theta] = \mu$$
$$Var[X_{ij} \mid \theta] = \sigma_\Delta^2 + \sigma_e^2$$
$$Cov[X_{ij}, X_{kl} \mid \theta] = \begin{cases} \sigma_\Delta^2 + \sigma_e^2 & if \quad i = k, j = l \\ \sigma_\Delta^2 & if \quad i = k, j \neq l \\ 0 & if \quad i \neq k \end{cases}$$

Bayes Procedures With Hierarchical Prior

Example 3.2.4 Random Effects Model

- $X_{ij} = \mu + \Delta_i + \epsilon_{ij}$, $i = 1, \ldots, I$ and $j = 1, \ldots, J$
  ϵ_{ij} are iid $N(0, \sigma_e^2)$.
- Δ_i iid $N(0, \sigma_\Delta^2)$ independent of the ϵ_{ij}
- $\mu \sim N(\mu_0, \sigma_\mu^2)$.

Bayes Model: Specification II

- Prior distribution on $(\mu, \sigma_e^2, \sigma_\Delta^2, \Delta_1, \ldots, \Delta_I)$
  $$\begin{aligned}\pi(\theta) &= \pi_1(\mu) \cdot \pi_2(\sigma_e^2) \cdot \pi_3(\sigma_\Delta^2) \cdot \prod_{i=1}^I \pi_4(\Delta_i \mid \sigma_\Delta^2) \\ &= \pi_1(\mu) \cdot \pi_2(\sigma_e^2) \cdot \pi_3(\sigma_\Delta^2) \cdot \prod_{i=1}^I \phi_{\sigma_\Delta}(\Delta_i)\end{aligned}$$
- Data distribution: $X_{ij} \mid \theta$ are *independent* normal random variables with
  $$E[X_{ij} \mid \theta] = \mu + \Delta_i,\ i = 1, \ldots I, \text{ and } j = 1, \ldots, J$$
  $$Var[X_{ij} \mid \theta] = \sigma_e^2$$

Bayes Procedures with Hierarchical Prior

Issues:

- Decision problems often focus on single Δ_i
- Posterior analyses then require marginal posterior distributions; e.g.
  $$\pi(\Delta_1 = d_1 \mid \mathbf{x}) = \int_{\{\theta:\Delta_1 = d_1\}} \pi(\theta \mid \mathbf{x}) \prod_{\{i \text{ except } \Delta_1\}} d\theta_i$$
- Approaches to computing marginal posterior distributions
  - Direct computation (conjugate priors)
  - Markov-Chain Monte Carlo (MCMC): simulations of posterior distributions.

Equivariance

Definition

- $\hat{\theta}_M$: estimator of θ applying methodolgy M.
- $h(\theta)$: one-to-one function of θ (a reparametrization).
- $\hat{\theta}_M$ is equivariant if
$$\widehat{h(\theta)}_M = h(\hat{\theta}_M)$$

Equivariance of MLEs and Bayes Procedures

- MLEs are equivariant.
- Bayes procedures not necessarily equivariant
  - For squared error loss, the Bayes procedure is mean of posterior distribution.
  - With non-linear reparametrization $h(\cdot)$,
$$E[h(\theta) \mid x] \neq h(E[\theta \mid x]).$$

Bayes Procedures and Reparametrization

Reparametrization of Bayes Decision Problems

- Reparametrization in Bayes decision analysis is not just a transformation-of-variables exercise with the joint/posterior distributions.
- The loss function should be transformed as well.
  If $\phi = h(\theta)$ and $\Phi = \{\phi : \phi = h(\theta), \theta \in \Theta\}$ then
  $$L^*[\phi, a] = L[h^{-1}(\phi), a].$$
- The decision analysis should be independent of the parametrization.

Equivariant Bayesian Decision Problems

Equivariant Loss Functions

- Consider a loss function for which:
  $L(h(\theta), a) = L(\theta, a)$, for all one-to-one functions $h(\cdot)$.
- Such a loss function is *equivariant*
- General class of equivariant loss functions:
  $L(\theta, a) = Q(P_\theta, P_a)$
  E.g., Kullback-Leibler divergence loss:
  $$L(\theta, a) = -E[log(\frac{p(x \mid a)}{p(x \mid \theta)}) \mid \theta]$$
  Loss $\equiv$ probability-weighted log-likelihood ratio.
  For canonical exponential family:
  $$L(\eta, a) = \sum_{j=1}^{k} [\eta_j - a_j) E[T_j(X) \mid \eta] + A(\eta) - A(a)$$