Bayes estimators

Given θ, $X_1, \cdots, X_n$ iid $\sim f(x|\theta)$.
prior $g(\theta)$.
joint distribution: $f(x_1, \cdots, x_n|\theta)g(\theta)$
posterior $g(\theta|x_1, \cdots, x_n)$.
Estimator $T = h(X_1, \cdots, X_n)$.
Loss: $L(t, \theta)$
Risk: $R(T, \theta) = E_T L(T, \theta)$
Bayes risk of T is
$R(T, g) = \int_\Theta R(T, \theta)g(\theta)d\theta$.
T is Bayes estimator if
$R(T, g) = \inf_T R(T, g)$.

To minimize Bayes risk, we only need to minimize the conditional expected loss given each x observed.

Binomial model

$X \sim bin(x|n,\theta)$
prior $g(\theta) \sim unif(0,1)$.
posterior $g(\theta|x) \sim beta(x+1, n-x+1)$.
Beta distribution:
$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, 0 \le x \le 1, E(X) = \frac{\alpha}{\alpha+\beta}$.
For squared error loss, the Bayes estimator is
posterior mean $\frac{x+1}{n+2}$.
If X is a random variable, choose c
$\min E(X-c)^2$.
$c = E(X)$

Prior distributions

Where do prior distributions come from?
* a prior knowledge about θ
* population interpretation–(a population of possible θ values).
* mathematical convenience (conjugate prior)
conjugate distribution– the prior and the posterior distribution are in the same parametric family.

Conjugate prior distribution

Advantages:
* mathematically convenient
* easy to interpret
* can provide good approximation to many prior opinions
(especially if we allow mixtures of distributions from the conjugate family)
Disadvantages:
may not be realistic

Binomial model

$X|\theta \sim bin(n, \theta)$
$\theta \sim beta(\alpha, \beta)$.
Beta distribution:
$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$,
$0 \leq x \leq 1, E(X) = \frac{\alpha}{\alpha+\beta}$.
$g(\theta|x) = beta(x+\alpha, n-x+\beta)$.
posterior mean is $\hat{\theta} = \frac{x+\alpha}{n+\alpha+\beta}$.

Binomial model: continued

Under loss function $L(t,\theta) = \frac{(t-\theta)^2}{\theta(1-\theta)}$.
We need to minimize
$g(t) = E_{\theta|X} L(t,\theta) = E_{\theta|X} \frac{(t-\theta)^2}{\theta(1-\theta)}$
$= t^2 E_{\theta|X} \frac{1}{\theta(1-\theta)} - 2t E_{\theta|X} \frac{1}{1-\theta} + E_{\theta|X} \frac{\theta}{1-\theta}$.
It is a quadratic function of t.
The minimizer is
$t^* = \frac{E_{\theta|X} \frac{1}{1-\theta}}{E_{\theta|X} \frac{1}{\theta(1-\theta)}} = \frac{\alpha+x-1}{\alpha+\beta+n-2}$.

exercise

Let $X_1, \cdots, X_n$ be iid Poisson (λ), and let λ have a gamma (α, β) distribution.

1). Find the posterior distribution of λ.

2). Calculate the posterior mean and variance.

3). Find the Bayes estimator under the loss $L(t, \lambda) = \frac{(t-\lambda)^2}{\lambda}$.

Solutions

1). The joint distribution of $X_1, \cdots, X_n$ given λ is
$p(x_1, \cdots, x_n|\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum_i x_i}}{\prod_i x_i!}$
the posterior distribution of λ is
$g(\lambda|x_1, \cdots, x_n) \propto p(x_1, \cdots, x_n|\lambda)g(\lambda)$
$\propto \frac{e^{-n\lambda}\lambda^{\sum_i x_i}}{\prod_i x_i!} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$
$\propto \lambda^{\sum_i x_i + \alpha - 1} e^{-(\beta+n)\lambda} \propto$ gamma $(\sum_i x_i + \alpha, \beta + n)$.
2). The posterior mean of λ is $E(\lambda|x_1, \cdots, x_n) = \frac{\sum_i x_i + \alpha}{\beta + n}$,
and the posterior variance is $Var(\lambda|x_1, \cdots, x_n) = \frac{\sum_i x_i + \alpha}{(\beta+n)^2}$.

3). For simplicity, write $x = (x_1, \cdots, x_n)$, we need to minimize
$E_{\lambda|x} L(t,\lambda) = E_{\lambda|x} \frac{(t-\lambda)^2}{\lambda} = E_{\lambda|x}(\frac{1}{\lambda}t^2 - 2t + \lambda)$
which is a quadratic function of t. The minimizer is
$t^* = \frac{1}{E_{\lambda|x}\frac{1}{\lambda}}$.
Note $E_{\lambda|x}\frac{1}{\lambda} = \int \frac{1}{\lambda} \frac{(\beta+n)^{\sum_i x_i+\alpha}}{\Gamma(\sum_i x_i+\alpha)} \lambda^{\sum_i x_i+\alpha-1} e^{-(\beta+n)\lambda} d\lambda =$
$\frac{\Gamma(\sum_i x_i+\alpha-1)(\beta+n)^{\sum_i x_i+\alpha}}{\Gamma(\sum_i x_i+\alpha)(\beta+n)^{\sum_i x_i+\alpha-1}} = \frac{\beta+n}{\sum_i x_i+\alpha-1}$.
and the Bayes estimator is $t^* = \frac{\sum_i x_i+\alpha-1}{\beta+n}$.

Let Θ consists of two points, $\Theta = \{\frac{1}{3}, \frac{2}{3}\}$.
Let $\mathcal{A}$ be the real line and $L(\theta, a) = (\theta - a)^2$.
A coin is tossed once and the probability of head is θ. Consider the set of nonrandomized decision rules which are functions from the set $\{H, T\}$ to $\mathcal{A}$: $d(H) = x, d(T) = y$.
Find the Bayes rule with respect to the prior distribution giving prob. $1/2$ to $\theta = 1/3$ and prob. $1/2$ to $\theta = 2/3$.

Exercise

Suppose the conditional distribution of given θ is $X|\theta \sim bin(3, \theta)$, and the prior distribution of θ is uniform on (0, 1). i.e., $g(\theta) = 1, 0 \leq \theta \leq 1$.

Suppose $X = 2$ is observed.

1). Find the joint distribution of X and θ.

2). Find the probability $P(X = 2)$.

3). Derive the conditional distribution of θ given $X = 2$.

4). Find the conditional mean and variance of θ given $X = 2$.