Lecture 21: Bayesian estimation

- With squared error loss, data X, prior $\pi(\theta)$ the Bayes estimate of θ is
$$\mathrm{E}\left(\theta|X\right) = \int \theta\pi(\theta|X)d\theta.$$
- For data $X = (X_1, \ldots, X_n)$ with the X_i iid $(N(\mu, \sigma^2)$ and a $N(\mu, \tau^s)$ prior I computed the joint density of X and μ, the marginal density of X and the posterior density of μ given X.
- Basic result: if Y_1, Y_2 are independent $N(\nu_i, \gamma_i^2)$ and $Y = X_1 + X_2$ then
$$\mathrm{E}(Y_1|Y) = \nu_1 + \frac{\mathrm{Cov}(Y_1, Y)}{\mathrm{Var}(Y)}(Y - \nu_1 - \nu_2) = \nu_1 + \frac{\gamma_1^2}{\gamma_1^2 + \gamma_2^2}((Y - \nu_1 - \nu_2).$$
- Note that
$$\mu, \bar{X} - \mu, \text{ and } (X_1 - \mu, \ldots, X_n - \mu)$$
are independent. (MVN so just compute covariances.)
- So:
$$\mathrm{E}(\mu|X_1, \ldots, X_n) = \mathrm{E}(\mu|X_1 - \bar{X}, \ldots, X_n - \bar{X}, \bar{X}) = \mathrm{E}(\mu|\bar{X}).$$

Consequences

- For each $w \in (0,1)$ and each μ_0 the estimate $w\bar{X} + (1-w)\mu_0$ is admissible.
- True also for $w = 1$ – $\bar{X}$ is admissible.
- $\bar{X}$ is Bayes for improper prior $\pi(\mu) \equiv 1$.
- Sometimes improper prior leads to proper posterior – like here.
- Not enough to prove $\bar{X}$ is admissible but it is.
- If $X_1, \ldots, X_n$ are iid $\mathrm{MVN}_p(\mu, I)$ then $\bar{X}$ is inadmissible for $p \geq 3$ and squared error loss $L(\hat{\mu}, \mu) = ||\hat{\mu} - \mu||^2$.

Binomial example / minimax

- For data $X = (X_1, \ldots, X_n)$ with the X_i iid Bernoulli(θ) and a Beta(α, β) prior I computed the joint density of X and θ, the marginal density of X and the posterior density of θ given X.
- Using $S = \sum_i X_i$:

$$\begin{aligned} f(x|\theta)\pi(\theta) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha+S-1}(1-\theta)^{\beta+n-S-1} \\ f(x) &= \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+S)\Gamma(\beta+n-S)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+n)} \\ \pi(\theta|X) &= \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+S)\Gamma(\beta+n-S)}\theta^{\alpha+S-1}(1-\theta)^{\beta+n-S-1} \end{aligned}$$

- I defined conjugate prior families.

Lecture 21

- In both cases the estimate is a weighted average of the usual estimate and the prior mean.
- The Bernoulli (Binomial) example illustrates the way we deal with mixed discrete and continuous variables.
- Many prediction, interpolation and extrapolation schemes are based on conditional expectations.
- I looked at a couple of improper priors leading, via Bayes' formula, to real posteriors.

Minimax

- The Bayes estimate of θ is $\delta(X) \equiv w\hat{\theta} + (1-w)\theta_0$ where

$$\hat{\theta} = X/n \text{ and } \theta_0 = \alpha/(\alpha+\beta).$$

- This estimate has mean squared error

$$R(\theta,\delta) = w^2\frac{\theta(1-\theta)}{n} + (1-w)^2(\theta-\theta_0)^2$$

- I showed how to find w and θ_0 so that the risk function is constant.
- For this choice δ is minimax.
- This estimate is not a good one.

Coverage in text and notes

- Chapters 11 and 12.
- Course slides on Bayes estimation.