Feynman Liang
STAT 210A
Lecture #9?? TODO: Check and type up previous
9/20/2016

1 Bayes Risk

Proof. Let $\delta(x)$ be any other estimator.

$$\begin{aligned} R_{Bayes}(\Lambda,\delta) &= \mathbb{E}[L(\Theta,\delta(X))] && (1.1)\\ &= \mathbb{E}[\mathbb{E}[L(\Theta,\delta)|X]] && (1.2)\\ &\geq \mathbb{E}[\mathbb{E}[L(\Theta,\delta_\Lambda(X))|X]] \qquad = R_{Bayes}(\Lambda,\delta_\Lambda) && (1.3)\\ & && (1.4) \end{aligned}$$

□

2 Prior, Posterior

Usually interpret Λ capturess prior beliefs, **prior distribution**

Definition 2.1. The *posterior distribution* is distribution of Θ given $X = x$

2.1 Densities

Write $\lambda(\theta)$ for the prior density.
$p_\theta(x)$ likelihood $P(X = x|\Theta = \theta)$.
$\lambda(\theta|x) = \frac{\lambda(\theta)p_\theta(x)}{\int_\Omega \lambda(\gamma)p_\gamma(x)d\gamma}$ the posterior
$q(x) = \int \lambda(\theta)p_\theta(x)d\theta$ the marginal.

$$\begin{aligned} \lambda(\theta|x) &= P(\Theta = \theta|X = x) && (2.1)\\ &= \frac{P(\Theta = \theta, X = x)}{P(X = x)} && (2.2) \end{aligned}$$

Bayes estimator minimizes

$$\int_\Omega L(\theta,d)\lambda(\theta|x)d\theta \tag{2.3}$$

for observed x.

2.2 Squared-Error

A "bias-variance tradeoff" for Bayes estimators

$$L(\theta, d) = (g(\theta) - d)^2 \tag{2.4}$$

$$\int (g(\theta) - d)^2 \lambda(\theta|x) d\theta = \mathbb{E}[(g(\theta) - d)^2 | X = x] \tag{2.5}$$

$$= \mathrm{Var}(g(\theta)|X = x) + (\mathbb{E}[g(\theta)|X = x] - d)^2 \tag{2.6}$$

Example 2.2 (Beta-Binomial)**.** $X|\Theta = \theta \sim \mathrm{Binom}(n, \theta)$, $P(x|\theta) = \theta^x (1-\theta)^{n-x} \binom{n}{x}$, $x = 0, 1, \cdots, n$

Prior distribution $\Theta \sim \mathrm{Beta}(\alpha, \beta)$, $P(\theta) = \theta^{\alpha-1}(1-\theta)^{\beta-1} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ where $\Gamma(t) = \int_0^\infty u^{t-1} e^{-u} du$ is the *gamma function*.

(Aside: the maximum of n uniforms $nx^{n-1} \sim \mathrm{Beta}(n, 1)$)

The posterior distribution

$$\lambda(\theta|x) = \lambda(\theta) p_\theta(x)/q(x) \tag{2.7}$$

$$= \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}\theta^x(1-\theta)^{n-x}\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\binom{n}{x}}{\int_\Omega \theta^{\alpha-1}(1-\theta)^{\beta-1}\theta^x(1-\theta)^{n-x}\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\binom{n}{x} d\theta} \tag{2.8}$$

So

$$\lambda(\theta|x) \propto_\theta \theta^{x+\alpha-1}(1-\theta)^{n-x+\beta-1} \tag{2.9}$$

$$\propto \mathrm{Beta}(x+\alpha, n-x+\beta) \tag{2.10}$$

Definition 2.3. A *conjugate prior* is a prior for which posterior has same form. (e.g. Beta is conjugate prior for Binom likelihood)

$$\implies \mathbb{E}[\Theta|X = x] = \frac{x+\alpha}{n+\alpha+\beta} \to x/n \tag{2.11}$$

Example 2.4. Normal Mean $X|\Theta = \theta \sim \mathcal{N}(\theta, \sigma^2) \propto_\theta \exp\{\frac{-(x-\theta)^2}{2\sigma^2}\}$

$\Theta \sim \mathcal{N}(\mu, \tau^2) \propto_\theta \exp\{\frac{-(\theta-\mu)^2}{2\tau^2}\}$

Then

$$\lambda(\theta|x) \propto_\theta \exp\left\{\frac{-(x-\theta)^2}{2\sigma^2} - \frac{(\theta-\mu}{2\tau^2}\right\} \tag{2.12}$$

$$\propto_\theta \exp\left\{\theta\left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right) - \theta^2\left(\frac{1}{2\sigma^2} + \frac{1}{2\tau^2}\right)\right\} \tag{2.13}$$

$$= \exp\left\{\theta\left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right)\frac{\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}}{\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}} - \frac{\theta^2}{2\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}}\right\} \tag{2.14}$$

Defining $\mu_{post} = \left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}\right) \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}$ and $\sigma^2_{post} = \frac{\sigma^2\tau^2}{\sigma^2+\tau^2}$, we have

$$\lambda(\theta|x) \propto_\theta \mathcal{N}(\frac{x\tau^2 + \mu\sigma^2}{\sigma^2 + \tau^2}, \sigma^2_{post}) \tag{2.15}$$

Letting $\sigma^2 = \sigma_0^2/n$ and $\tau^2 = \sigma_0^2/m$ (pseudo-data interpretations: n observations with mean x and m observations with mean μ)

$$\Theta|X = x \sim \mathcal{N}\left(\frac{xn + \mu m}{n+m}, \sigma_0^2 \frac{\frac{1}{mn}}{\frac{1}{m} + \frac{1}{n}}\right) \tag{2.16}$$

$$= \mathcal{N}\left(\frac{xn + \mu m}{n+m}, \frac{\sigma_0^2}{m+n}\right) \tag{2.17}$$

This is like a UMVUE but with some modification:

$$\mathbb{E}[\Theta|X = x] = x\frac{n}{n+m} + \mu\frac{m}{n+m} \tag{2.18}$$

Theorem 2.5. *The posterior mean is biased unless* $\delta_\Lambda(X) \stackrel{a.s.}{=} g(\Theta)$.

Proof.

$$\mathbb{E}[(\delta_\Lambda(X) - g(\Theta))^2] = \mathbb{E}[\delta_\Lambda(X)^2] + \mathbb{E}[g(\Theta)^2] - 2\mathbb{E}[\delta_\Lambda(x)g(\Theta)] \tag{2.19}$$

$$= (1 + 1 - 2)\mathbb{E}[\delta_\Lambda(X)^2] = 0 \tag{2.20}$$

$$\mathbb{E}[\delta(X)g(\Theta)] = \mathbb{E}[\delta(X)\mathbb{E}[g(\theta)|X]] \tag{2.21}$$

$$= \mathbb{E}[\delta(X)^2] \tag{2.22}$$

$$\mathbb{E}[\delta(X)g(\Theta)] = \mathbb{E}[g(\Theta)\mathbb{E}[\delta(X)|\Theta]] \tag{2.23}$$

$$= \mathbb{E}[g(\Theta)^2] \tag{2.24}$$

$$\implies (g(\Theta) - \delta(X))^2 \stackrel{\text{a.s.}}{=} 0 \tag{2.25}$$

□