7.2.3 Bayes Estimators

Classical approach. θ is unknown but fixed

Bayesian approach. θ can be described by a ~~prior~~ distribution, $\pi(\theta)$ also called a prior distribution.

Given a sample $X_1, \dots, X_n$, the prior can be updated to $\pi(\theta | \underline{x})$ as follows.

$$\pi(\theta|\underline{x}) = \frac{f(\theta, \underline{x})}{f(\underline{x})} = \frac{f(\underline{x}|\theta) \cdot \pi(\theta)}{f(\underline{x})}$$

$\pi(\theta|\underline{x})$ is called the posterior and

$$f(\underline{x}) = \int f(\underline{x}, \theta)\, d\theta = \int f(\underline{x}|\theta)\, \pi(\theta)\, d\theta$$

Example. Binomial Bayes estimation

Assume $X_1, \dots, X_n \sim$ Bernoulli(p). Let $Y = \sum_{i=1}^{n} X_i \sim B(n, p)$

Assume the prior for p is beta(α, β) i.e

$$\pi(p) = f(p|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1}, \quad p \in (0,1), \ \alpha > 0, \ \beta > 0$$

$$f(y, p) = f(y|p) \cdot \pi(p) = \binom{n}{y} p^{y} (1-p)^{n-y} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\, \Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1}$$

$$= \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} p^{y+\alpha-1} (1-p)^{n-y+\beta-1}$$

$$f(y) = \binom{m}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \underbrace{\int_0^1 p^{y+\alpha-1} (1-p)^{m-y+\beta-1}\, dp}_{\dfrac{\Gamma(y+\alpha)\cdot\Gamma(m-y+\beta)}{\Gamma(m+\alpha+\beta)}}$$

$$= \binom{m}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \cdot \frac{\Gamma(y+\alpha)\;\Gamma(m-y+\beta)}{\Gamma(m+\alpha+\beta)}$$

and $$\pi(p \mid y) = \frac{f(y,p)}{f(y)} = \frac{\binom{m}{y} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\, p^{y+\alpha-1} (1-p)^{m-y+\beta-1}}{\binom{m}{y} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\; \dfrac{\Gamma(y+\alpha)\;\Gamma(m-y+\beta)}{\Gamma(m+\alpha+\beta)}}$$

$$= \frac{\Gamma(m+\alpha+\beta)}{\Gamma(y+\alpha)\;\Gamma(m-y+\beta)}\; p^{y+\alpha-1} (1-p)^{m-y+\beta-1} \quad \text{i.e } \beta(y+\alpha,\, m-y+\beta)$$

Note that $$f(y,p) = \underbrace{\binom{m}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} \cdot \frac{\Gamma(y+\alpha)\,\Gamma(m-y+\beta)}{\Gamma(m+\alpha+\beta)}}_{f(y)} \cdot \underset{\pi(p \mid y)}{\frac{\Gamma(m+\alpha+\beta)}{\Gamma(y+\alpha)\,\Gamma(m-y+\beta)} \cdot p^{y+\alpha-1} (1-p)^{m-y+\beta}}$$

and $$\int f(y,p)\, dp = f(y) \underbrace{\int \pi(p \mid y)\, dp}_{= 1}$$

$$E[p \mid y] = \frac{y+\alpha}{\alpha+\beta+m} \quad \text{and} \quad \frac{y+\alpha}{\alpha+\beta+m} = \hat{p}_B \text{ is called the}$$

Bayes estimator of p.

The prior distribution has a mean $\frac{\alpha}{\alpha+\beta}$

Ignoring the prior gives $\hat{p}_e = \frac{y}{n}$

$$\hat{p}_{Be} = \frac{1}{\alpha+\beta+n}(y+\alpha) = \frac{1}{\alpha+\beta+n}\left(\frac{ny}{n} + \frac{(\alpha+\beta)\cdot\alpha}{\alpha+\beta}\right)$$

$$= \frac{n}{\alpha+\beta+n}\left(\frac{y}{n}\right) + \frac{\alpha+\beta}{\alpha+\beta+n}\left(\frac{\alpha}{\alpha+\beta}\right)$$

As n increases $\frac{n}{\alpha+\beta+n} \rightarrow 1$ and $\frac{\alpha+\beta}{\alpha+\beta+n} \rightarrow 0$

Definition 7.2.15

Let $\mathcal{F}$ denote the class of pdfs/pmfs $f(x|\theta)$

A class, Π, of prior distributions is a conjugate family for $\mathcal{F}$ if the posterior distribution is in the class Π for all $f \in \mathcal{F}$, all priors in Π and all x.

The beta-family is a conjugate family for the binomial distribution.

Example 7.2.16

$X \sim N(\theta, \sigma^2)$. $\pi(\theta) \sim N(\mu, \tau^2)$, μ, τ^2 and σ^2 are known. We get $f(x,\theta) = f(x|\theta)\cdot\pi(\theta) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma\tau} e^{-\frac{1}{2}\left\{\frac{(x-\theta)^2}{\sigma^2} + \frac{(\theta-\mu)^2}{\tau^2}\right\}}$

$$f(x) = \int_{-\infty}^{\infty} f(x,\theta)\, d\theta.$$

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

$$= \theta^2\left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right) - 2\theta\left[\frac{\mu}{\tau^2} + \frac{x}{\sigma^2}\right] + \left[\frac{\mu^2}{\tau^2} + \frac{x^2}{\sigma^2}\right]$$

$$= \frac{\sigma^2+\tau^2}{\sigma^2\tau^2}\left[\left[\theta^2 - 2\theta\left[\frac{\mu\sigma^2 + x\tau^2}{\sigma^2+\tau^2}\right] + \left(\frac{\mu\sigma^2+x\tau^2}{\sigma^2+\tau^2}\right)^2\right] - \left[\frac{\mu\sigma^2+x\tau^2}{\sigma^2+\tau^2}\right]^2 + \frac{\mu^2\sigma^2 + x^2\tau^2}{\sigma^2+\tau^2}\right]$$

$$= \frac{\sigma^2+\tau^2}{\sigma^2\tau^2}\left[\theta - \frac{\mu\sigma^2+x\tau^2}{\sigma^2+\tau^2}\right]^2 + \frac{(x-\mu)^2}{\sigma^2+\tau^2}$$

$$\Rightarrow f(x,\theta) = \underbrace{\frac{1}{\sqrt{2\pi}} \cdot \frac{\sqrt{\sigma^2+\tau^2}}{\sigma\tau} e^{-\frac{\sigma^2+\tau^2}{2\sigma^2\tau^2}\left[\theta - \frac{\mu\sigma^2+x\tau^2}{\sigma^2+\tau^2}\right]^2}}_{\pi(\theta|x)} \cdot \underbrace{\frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{\sigma^2+\tau^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2+\tau^2}}}_{f(x)}$$

and $E[\theta|x] = \frac{\tau^2}{\sigma^2+\tau^2}x + \frac{\sigma^2}{\sigma^2+\tau^2}\mu$ $\qquad$ τ large $\Rightarrow$ heavy weight on x

$Var[\theta|x] = \frac{\sigma^2\tau^2}{\sigma^2+\tau^2} < \sigma^2$ $\qquad$ σ^2-large = more weight on the prior μ.