STAT-W4109: Probability and Statistics - Fall 2012

Lecture 12

Thursday - October 11

Kamiar Rahnama Rad

- Likelihood - a model for typical observation X conditional on unknown parameter θ - $f(x|\theta)$
- Prior - uncertainty about the parameter before data are observed - $\pi(\theta)$
- Marginal distribution of the data - $m(x) = \int f(x|\theta')\pi(\theta')d\theta'$
- Goal - update the prior using the data to make the best possible estimator of θ - Bayes Rule

$$\pi(\theta|x) \quad = \quad \frac{f(x|\theta)\pi(\theta)}{m(x)} \propto f(x|\theta)\pi(\theta)$$

- Summary

$$\begin{array}{rl} \text{Likelihood} & f(x|\theta) \\ \text{Prior} & \pi(\theta) \\ \text{Joint} & h(x,\theta) = f(x|\theta)\pi(\theta) \\ \text{Marginal} & m(x) = \displaystyle\int \pi(\theta')f(x|\theta')d\theta' \\ \text{Posterior} & f(x|\theta)\pi(\theta)/m(x) \end{array}$$

- Prove that $\mathrm{E}[\theta|X]$ is Bayes optimal
- Risk: $R(\theta) = \mathrm{E}_{X|\theta}[(\hat{\theta}(X) - \theta)^2]$
- Expected Risk: $E_\theta[R(\theta)]$

$$
\begin{aligned}
& \int \pi(\theta) d\theta \Big(\int f(x|\theta)(\theta - \hat{\theta}(x))^2 dx \Big) = \int f(x|\theta)\pi(\theta)(\theta - \hat{\theta}(x))^2 d\theta dx \\
= \quad & \int \pi(x|\theta) m(x)(\theta - \hat{\theta}(x))^2 d\theta dx = \int m(x) dx \Big(\int \pi(\theta|x)(\theta - \hat{\theta}(x))^2 d\theta \Big) \\
= \quad & E_X\Big[E_{\theta|X}[(\hat{\theta}(X) - \theta)^2]\Big] = E_X\Big[E_{\theta|X}[\hat{\theta}(X)^2 + \theta^2 - 2\theta\hat{\theta}(X)]\Big] \\
= \quad & E_X\Big[E_{\theta|X}[\theta^2] + \hat{\theta}(X)^2 - 2E_{\theta|X}[\theta]\hat{\theta}(X)]\Big] \\
= \quad & E_X\Big[E_{\theta|X}[\theta^2] - (E_{\theta|X}[\theta])^2 + (E_{\theta|X}[\theta])^2 + \hat{\theta}(X)^2 - 2E_{\theta|X}[\theta]\hat{\theta}(X)]\Big] \\
= \quad & E_X\Big[VAR_{\theta|X}[\theta] + \big(\hat{\theta}(X) - E_{\theta|X}[\theta]\big)^2]\Big] \\
= \quad & VAR[\theta] + E_X\Big[\big(\hat{\theta}(X) - E_{\theta|X}[\theta]\big)^2]\Big]
\end{aligned}
$$

- Normal Likelihood :$f(x|\theta) = \mathcal{N}(\theta, \sigma^2) = \frac{e^{-\frac{1}{2\sigma^2}(x-\theta)^2}}{\sqrt{2\pi\sigma^2}}$
- Normal Prior :$\pi(\theta) = \mathcal{N}(\mu, \tau^2) = \frac{e^{-\frac{1}{2\tau^2}(\theta-\mu)^2}}{\sqrt{2\pi\tau^2}}$
- Joint distribution:

$$
\begin{aligned}
h(x,\theta) &= f(x|\theta)\pi(\theta) = \frac{e^{-\frac{1}{2\sigma^2}(x-\theta)^2}}{\sqrt{2\pi\sigma^2}} \times \frac{e^{-\frac{1}{2\tau^2}(\theta-\mu)^2}}{\sqrt{2\pi\tau^2}} \\
&= \frac{e^{-\frac{1}{2\rho}\left[\theta-\rho(\frac{x}{\sigma^2}+\frac{\mu}{\tau^2})\right]^2}}{\sqrt{2\pi\rho}} \times \frac{e^{-\frac{1}{2(\tau^2+\sigma^2)}\left[x-\mu\right]^2}}{\sqrt{2\pi(\tau^2+\sigma^2)}} \\
&= \pi(\theta|x)m(x)
\end{aligned}
$$

where $\rho = \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1}$.

- Marginal: $m(x) = \mathcal{N}\left(\mu, \sigma^2 + \tau^2\right)$.
- Posterior: $\theta|X \sim \mathcal{N}\left(\frac{\sigma^2}{\sigma^2+\tau^2}\mu + \frac{\tau^2}{\sigma^2+\tau^2}X, \frac{1}{\frac{1}{\sigma^2}+\frac{1}{\tau^2}}\right)$.

- Likelihood: $D = X_1, X_2, \cdots, X_n \sim \mathcal{N}(\theta, \sigma^2)$

$$
\begin{aligned}
f(D|\theta) &= \prod_{i=1}^{n} \frac{e^{-\frac{1}{2\sigma^2}(x_i-\theta)^2}}{\sqrt{2\pi\sigma^2}} = \frac{e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2}}{(2\pi\sigma^2)^{n/2}} \\
&= \frac{e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2}}{(2\pi\sigma^2)^{n/2}} \times e^{-\frac{1}{2(\sigma^2/n)}(\bar{x}-\theta)^2}
\end{aligned}
$$

- Normal Prior :$\pi(\theta) = \mathcal{N}(\mu, \tau^2) = \frac{e^{-\frac{1}{2\tau^2}(\theta-\mu)^2}}{\sqrt{2\pi\tau^2}}$
- Joint distribution:

$$
\begin{aligned}
h(D, \theta) &= f(D|\theta)\pi(\theta) = \frac{e^{-\frac{1}{2(\sigma^2/n)}(\bar{x}-\theta)^2}}{\sqrt{2\pi(\sigma^2/n)}} \times \frac{e^{-\frac{1}{2\tau^2}(\theta-\mu)^2}}{\sqrt{2\pi\tau^2}} \times F(D) \\
&= \frac{e^{-\frac{1}{2\rho}\left[\theta-\rho\left(\frac{\bar{x}}{(\sigma^2/n)}+\frac{\mu}{\tau^2}\right)\right]^2}}{\sqrt{2\pi\rho}} \times \frac{e^{-\frac{1}{2(\tau^2+\sigma^2/n)}\left[\bar{x}-\mu\right]^2}}{\sqrt{2\pi(\tau^2+\sigma^2/n)}} \times F(D) \\
&= \pi(\theta|D)m(D)
\end{aligned}
$$

where $\rho = \left(\frac{1}{(\sigma^2/n)} + \frac{1}{\tau^2}\right)^{-1}$.

- Posterior: $\theta|D \sim \mathcal{N}\left(\frac{\sigma^2/n}{\sigma^2/n+\tau^2}\mu + \frac{\tau^2}{\sigma^2/n+\tau^2}\bar{X}, \frac{1}{\frac{1}{\sigma^2/n}+\frac{1}{\tau^2}}\right)$.

- Likelihood: $X|\theta \sim$Binomial(n, θ)
- Prior: $\pi(\theta) \sim$ Beta$(\alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$
- Joint: $h(X, \theta) = \frac{\binom{n}{x}\theta^{\alpha+x-1}(1-\theta)^{n-x+\beta-1}}{B(\alpha,\beta)}$
- Marginal: $\binom{n}{x}\frac{B(x+\alpha,n-x+\beta)}{B(\alpha,\beta)}$
- Posterior: $\frac{\theta^{\alpha+x-1}(1-\theta)^{n-x+\beta+1}}{B(x+\alpha,n-x+\beta)}$
- Prior Expectation: $\mathrm{E}[X] = \frac{\alpha}{\alpha+\beta}$
- Prior Variance: $\mathrm{VAR}[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ is $O(1/\alpha^2)$ for large α and fixed β
- Posterior Expectation :

$$E[\theta|X] = \frac{\alpha + X}{\alpha + \beta + n} \quad = \quad \left(\frac{\alpha}{\alpha + \beta + n}\right)\frac{\alpha}{\alpha + \beta} + \left(1 - \frac{\alpha}{\alpha + \beta + n}\right)\frac{X}{n}$$

- Likelihood: $D = X_1, X_2, \cdots, X_n \sim \mathcal{P}(\theta)$

$$f(D|\theta) = \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!} = \frac{e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

- Gamma Prior :$\pi(\theta) = \mathcal{G}(\alpha, \beta) = \theta^{\alpha-1} e^{-\beta\theta} \frac{\beta^\alpha}{\Gamma(\alpha)}$
- Joint distribution:

$$\begin{aligned} h(D,\theta) &= f(D|\theta)\pi(\theta) = \frac{e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \times \theta^{\alpha-1} e^{-\beta\theta} \frac{\beta^\alpha}{\Gamma(\alpha)} \\ &= e^{-(n+\beta)\theta}\theta^{\alpha-1+\sum x_i} \times \Big(\text{Independent of } \theta\Big) \\ &= \pi(\theta|D) m(D) \end{aligned}$$

- Posterion distribution: $\pi(\theta|D) = \mathcal{G}\big(\alpha + \sum X_i, n + \beta\big)$

- Posterion distribution: $\pi(\theta|D) = \mathcal{G}\big(\alpha + \sum X_i, n + \beta\big)$
- Prior mean: E$[\theta]$= $\frac{\alpha}{\beta}$
- Observed mean $\bar{X} = \frac{1}{n}\sum_i X_i$
- Posterior mean: E$[\theta|D] = \frac{\alpha+\sum X_i}{n+\beta}$

$$= \quad \frac{n}{n+\beta}\Big(\frac{\sum X_i}{n}\Big) + \frac{\beta}{n+\beta}\Big(\frac{\alpha}{\beta}\Big)$$

- as the sample size increase the prior becomes irrelevant
- prior can be improper but posterior stays proper

We have seen three convenient examples for which the posterior distribution remained in the same family as the prior distribution. In such a case, the effect of likelihood is only to update the prior parameters and not to change prior's functional form. We say that such priors are conjugate with the likelihood. Conjugacy is popular because of its mathematical convenience; once the conjugate pair likelihood/prior is found, the posterior is calculated with relative ease.