C2_Bayes Theorem, Prior and Decision Theory

1. The Bayes Concept

1.1 Mariginal PDF/PMF and Likelihood

For data $x_1, \ldots, x_n$ from a random sample, $f(x \mid \theta)=\prod_{i=1}^n h\left(x_i \mid \theta\right)$, and

Likelihood: $L(\theta) =_\kappa \prod_{i=1}^n h\left(x_i \mid \theta\right)$,

where h **is the marginal PDF** or PMF of X_i

1.2 Posterior

* **Posterior:** $g(\theta \mid x) \propto L(\theta) g(\theta)$,

which can be stated in words as the posterior is proportional to the likeithood times the prior.

- **if we interest on parametersn in theta:**

  Suppose $\theta=(\gamma, \phi) \in \Gamma \times \Phi$ and interest centers on inference about γ, then we can simply eliminate ϕ from the posterior by marginalizing it out. This gives the **posterior distribution of** γ **:**
  $g(\gamma \mid x)=\int_{\Phi} g(\gamma, \phi \mid x) d \phi=\int_{\Phi} g(\gamma \mid x, \phi) g(\phi \mid x) d \phi$

  - If the prior distribution and likelihood factors into
    $g(\gamma, \phi)=g(\gamma) g(\phi)$ and $L(\gamma, \phi)=L(\gamma) L(\phi)$,

    then γ and ϕ are posteriori independent, i.e., $g(\gamma, \phi \mid x)=g(\gamma \mid x) g(\phi \mid x)$

1.3 Sufficiency

* **Concept:** $t(x)$ is called a sufficient statistic for θ.

* **likelihood function** $L(\theta)=e(t(x) \mid \theta)$. This implies that $g(\theta \mid x)$ depends on x only through $t(x)$.

- **Examples**

  Example: Suppose that X has a **binomial distribution** with n trials and probability of success
  θ. Then, $t(x)=\bar{x}=\sum_{i=1}^n x_i / n$ is a sufficient statistic.
  Example 2: Suppose that $X_1, \ldots, X_n$ are iid **gamma** random variables, i.e., the sampling model is $f(x \mid \gamma, \phi)=\prod_{i=1}^n \frac{1}{\Gamma(\gamma) \phi^\gamma} x_i^{\gamma-1} e^{-x_i / \phi}=\frac{1}{\Gamma(\gamma) \phi^\gamma}(\tilde{x})^{n(\gamma-1)} e^{-n \bar{x} / \phi}$ where
  $\tilde{x}=\left(\prod_{i=1}^n x_i\right)^{1 / n}$. Note that $(\tilde{x}, \bar{x})$ is a sufficient statistic for (γ, ϕ)
  Example 3: Suppose that $X_1, \ldots, X_n$ are iid **Poisson** random variables with mean θ, i.e.,
  $f(x \mid \theta)=\prod_{i=1}^n \frac{e^{-\theta} \theta^{x_i}}{x!}=\kappa e^{-n \theta} \theta^{n \bar{x}}$ Note that $\bar{x}$ is a sufficient statistic for θ

2. Parametric Inference

2.1 Point estimation

- **Example:**

  Let $X_1, \ldots, X_n \sim$ Bernoulli (θ). Write $\dot{x} = \sum_{i=1}^{n} x_i$. Thus, $L(\theta) = \theta^{\dot{x}}(1-\theta)^{n-\dot{x}}$
  Let the prior for θ be $g(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$ i.e.we denote as Beta (a, b).

  The prior mean of θ is $a/(a+b)$.

  The posterior distribution: $g(\theta \mid x) \propto \theta^{\dot{x}+a-1}(1-\theta)^{n-\dot{x}+b-1}$, which is Beta $(\dot{x}+a, n-\dot{x}+b)$.

  Thus, $\bar{\theta} = \frac{\dot{x}+a}{n+a+b}$

2.2 Interval estimation

- **Concept:** Under a Bayesian approach, confidence intervals are replaced by **credible intervals**.
- For a real parameter θ, a $1-\alpha$ credible interval is formed by two values $\underline{\theta}$ and $\bar{\theta}$ such that $P(\underline{\theta} \leq \theta \leq \bar{\theta}) = \int_{\underline{\theta}}^{\bar{\theta}} g(\theta \mid x) d\theta = 1-\alpha$
- Equal tail probability credible interval $\int_{-\infty}^{\theta} g(\theta \mid x) d\theta = \int_{\bar{\theta}}^{\infty} g(\theta \mid x) d\theta = \alpha/2$.

2.3 Hypothesis testing

- **Concept:**

  The ratio of the posterior odds to the prior odds is the **Bayes factor** in favor of H_0
  $B(x) = \frac{P(\theta \in \Theta_0 | x)/P(\theta \in \Theta_1 | x)}{P(\theta \in \Theta_0)/P(\theta \in \Theta_1)}\}$

  - The Bayes factor quantifies the evidence in the data χ in favor of H_n.

2.4 Predictive Inference

- **Aim:**

  In many instances, our interest centers on predicting a set of unobserved $Y_1, \ldots, Y_k$ random variables given the data $X_1 = x_1, \ldots, X_n = x_n$

- **How ?** : predict based on the predictive distribution:
  $f(y \mid x) = \int_{\Theta} f(y \mid \theta) g(\theta \mid x) d\theta$
  - Usually, $f(y \mid x, \theta) = f(y \mid \theta)$ and $f(y \mid \theta)$

3. Prior Distributions

3.1 Concepts:

prior distributions for two cases:

1. No prior information is available or when prior knowledge is of little significance compared to information from the data: uninformative and these are commonly called **Noninformative priors.(Jeffrey's Prior**)
2. **Natural conjugate priors**. In this case, the prior distribution is assumed to be from a parametric family, and its parameters are chosen so that there is a match between elicited summary measures of the decision maker's prior distribution and the corresponding summary measures of the natural conjugate prior

3.2 Jeffrey's Prior -- Noninformative priors

- **Aim:**

  Suppose $\psi=\xi(\theta)$, However it is not easy to calculate prior and posterior directly from ψ. Thus, calculate prior and posterior from θ, and due to its invarient in one-to-one transformation. **We can get prior or posterior of ψ.**

- **Steps:**
  - Suppose $\psi=\xi(\theta)$
  - Fisher's infomation: $I(\theta)=E\left[\left(\frac{\partial \ln f(X|\theta)}{\partial \theta}\right)^2 \mid \theta\right]=-E\left[\frac{\partial^2 \ln f(X|\theta)}{\partial \theta^2} \mid \theta\right]$
  - Jeffrey's prior: $g(\theta) \propto \sqrt{I(\theta)}$
  - Interest prior: $h(\psi) \propto g(\theta)\left|\frac{d\theta}{d\psi}\right|$

- **Examples:**

  (a) Suppose $X \mid \theta \sim$ Poisson (θ) (Poisson distribution with mean θ). Derive the Jeffrey's prior for θ
  (b) Suppose $X \mid \theta \sim \operatorname{Gamma}(\gamma, 1/\theta)$, where γ is known. Derive the Jeffrey's prior for θ.
  Note: You only need to give the prior probability density functions (PDF's) up to a proportionality constant. Answer:

  (a) The PMF for the **Poisson distribution** is $f(x \mid \theta)=\frac{\exp(-\theta)\theta^x}{x!}, x \in\{0,1,\ldots\}$. We have

$$I(\theta)=-E\left[\frac{\partial^2 \ln f(X \mid \theta)}{\partial \theta^2} \mid \theta\right]=-E\left[\frac{\partial^2}{\partial \theta^2}(-\theta+X \ln \theta) \mid \theta\right]=-E\left[-\frac{X}{\theta^2} \mid \theta\right]=\frac{1}{\theta}, \theta>0$$

  since $E(X \mid \theta)=\theta$. Thus, the **Jeffrey's prior** is

$$g(\theta) \propto \sqrt{I(\theta)} = \theta^{-\frac{1}{2}}, \theta > 0$$

This is an *improper prior.*

(b) The PDF for the **Gamma distribution** is $f(x \mid \theta) = \frac{1}{\Gamma(\gamma)} \theta^\gamma x^{\gamma-1} e^{-\theta x} I(x > 0)$. We have

$$I(\theta) = -E\left[\frac{\partial^2 \ln f(X \mid \theta)}{\partial \theta^2} \mid \theta\right] = -E\left\{\frac{\partial^2}{\partial \theta^2}[\gamma \ln \theta - \theta X + (\gamma - 1)\ln(X) - \ln(\Gamma(\gamma))] \mid \theta\right\} =$$
$$-E\left[-\frac{\gamma}{\theta^2} \mid \theta\right] = \frac{\gamma}{\theta^2}, \theta > 0$$

Thus, the **Jeffrey's prior** is

$$g(\theta) \propto \sqrt{I(\theta)} \propto \theta^{-1}, \theta > 0$$

This is an *improper prior*

4. Natural Conjugate Priors

4.1 Concept

- **Family of prior distributions** $\mathcal{H} = \{g_a(\theta) : a \in \mathcal{A}\}$
- **Posterior: if** $g(\theta) \in \mathcal{H} \Rightarrow g(\theta \mid x) \propto f(x \mid \theta) g(\theta) \in \mathcal{H}$

4.2 Examples

- Exe Q5(b) -- Prove geometric and beta are natural conjugate pairs
    - **The Geometric distribution:** $h(x_i \mid \theta) = \theta(1-\theta)^{x_i - 1}, x_i \in \{1, 2, \ldots\}$
    - If we choose $\mathrm{Beta}(a, b)$ **as the prior** for θ, then the
    - **Posterior distribution** for θ:
      $g(\theta \mid x) \propto L(\theta) g(\theta) \propto \theta^n (1-\theta)^{\dot{x}-n} \theta^{a-1} (1-\theta)^{b-1} = \theta^{n+a-1}(1-\theta)^{\dot{x}-n+b-1}$
    - **Thus, we see that** $\theta \mid x \sim \mathrm{Beta}(n + a, \dot{x} - n + b)$,*
    - which implies that the $\mathrm{Beta}(a, b)$ distribution is a natural conjugate family for the Geometric (θ) sampling model.