Lecture 6: Posterior distribution and posterior consistency

Tianpei Xie

Aug.11st., 2015

Contents

1 Prior, Posterior, and Predictive Distributions

- **Definition** (***Prior Distribution***) [Ghosh and Ramamoorthi, 2003, Schervish, 2012]
Let $(\Omega, \mathscr{P}, \mathbb{P})$ be a *probability space*, and let $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ and $(\Theta, \mathcal{B}(\Theta))$ be *Borel spaces*. Let $X : \Omega \to \mathcal{X}$ and $\theta : \Omega \to \Theta$ be *measurable*. Then θ is called a ***parameter*** and Θ is called a ***parameter space***.

  The *conditional distribution for X given θ* is called ***a parametric family of distributions*** of X. *The parametric family* is denoted by

$$\mathcal{P}_0 = \left\{\mathcal{P}_{\theta_0} : \forall A \in \mathcal{B}(\mathcal{X}), \mathcal{P}_{\theta_0}(A) = \mathbb{P}\left\{X \in A | \theta = \theta_0\right\}, \theta_0 \in \Theta\right\}.$$

  We also use the symbol $\mathbb{P}_{\theta_0}[X \in A]$ to stand for $\mathcal{P}_{\theta_0}(A)$. ***The prior distribution*** of θ is *the probability measure* μ_θ over $(\Theta, \mathcal{B}(\Theta))$ induced by θ from $\mathbb{P}$.

- **Definition** (***Likelihood Function and Conditional Density***) [Schervish, 2012]
Suppose that each $\mathcal{P}_{\theta_0}$, when considered as a measure on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, is *absolutely continuous* with respect to a measure ν on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Let

$$f_{X|\theta}(x|\theta_0) = \frac{d\mathcal{P}_{\theta_0}}{d\nu}.$$

  We can assume that $f_{X|\theta}$ is ***measurable*** with respect to ***the product σ-field*** $\mathcal{B}(\mathcal{X}) \otimes \mathcal{B}(\Theta)$. This will allow us to integrate this function with respect to measures on *both* $\mathcal{X}$ and Θ. The function $f_{X|\theta}(x|\theta_0)$, considered as ***a function of*** θ after $X = x$ is observed, is often called ***the likelihood function*** $L(\theta)$.

  For each $\theta \in \Theta$, the function $f_{X|\theta}(x|\theta_0)$ is ***the conditional density*** with respect to ν of X given $\theta = \theta_0$. That is for each $A \in \mathcal{B}(\mathcal{X})$,

$$\mathcal{P}_\theta(A) = \int_A f_{X|\theta}(x|\theta) d\nu(x)$$

- **Definition** (***Marginal Distribution***) [Schervish, 2012]
Let μ_X be ***the marginal distribution*** so that

$$\mu_X(A) = \mathbb{P}\left[X \in A\right]$$

  Using *Tonelli's theorem*, we have

$$\mu_X(A) = \int_\Theta \left[\int_A f_{X|\theta}(x|\theta) d\nu(x)\right] d\mu_\theta(\theta) = \int_A \left[\int_\Theta f_{X|\theta}(x|\theta) d\mu_\theta(\theta)\right] d\nu(x)$$

  It follows that μ_X is *absolutely continuous* with respect to ν with *density*

$$f_X(x) = \int_\Theta f_{X|\theta}(x|\theta) d\mu_\theta(\theta)$$

  This density is often called ***the (prior) predictive density*** *of X* or ***the marginal density*** *of X*.

- **Theorem 1.1** *(**Bayes' Theorem**). [Schervish, 2012]*
  *Suppose that X has a parametric family $\mathcal{P}_0$ of distributions with **parameter space** Θ. Suppose that $\mathcal{P}_\theta \ll \nu$ in $\mathcal{X}$ for all $\theta \in \Theta$, and let $f_{X|\theta}(x|t)$ be the **conditional density** (with respect to ν) of X given $\theta = t$. Let μ_θ be the **prior distribution** of θ.*

  *Let $\mu_{\theta|X}(\cdot|x)$ denote **the conditional distribution of** θ **given** $X = x$. Then $\mu_{\theta|X} \ll \mu_\theta$, a.s. **with respect to the marginal of** X, and the Radon-Nikodym **derivative** is*

  $$\frac{d\mu_{\theta|X}}{d\mu_\theta}(\theta|x) = \frac{f_{X|\theta}(x|\theta)}{\int_\Theta f_{X|\theta}(x|t)d\mu_\theta(t)} \tag{1}$$

  *for those x such that the **denominator** is **neither** 0 **nor infinite**. The **prior predictive probability** of the set of x values such that the denominator is 0 or infinite is 0, hence the posterior can be defined arbitrarily for such x values.*

- **Definition** (***Posterior Distribution***) [Schervish, 2012]
  The conditional distribution of θ given $X = x$ is called ***the posterior distribution** of θ*, denoted as $\mu_{\theta|X}$.

  1. $\mu_{\theta|X}(\cdot|x)$ is a ***measure*** on $\mathcal{B}(\Theta)$ given $X = x$;

     The Bayes theorem confirms that $\mu_{\theta|X} \ll \mu_\theta$ and for all $B \in \mathcal{B}(\Theta)$, and the following equation holds *almost surely with respect to* μ_X

     $$\mu_{\theta|X}(B|x) = \int_B \frac{f_{X|\theta}(x|s)}{\int_\Theta f_{X|\theta}(x|t)d\mu_\theta(t)} d\mu_\theta(s), \quad \mu_X\text{-a.s.} \tag{2}$$

  2. $\mu_{\theta|X}(B|\cdot)$ is a $\mathcal{B}(\mathcal{X})$-***measurable function*** of X given any $B \in \mathcal{B}(\Theta)$;

  3. $\mu_{\theta|X}(B|\cdot)$ is ***integrable*** with respect to ***marginal distribution*** μ_X and

     $$\mathbb{P}[X \in A, \theta \in B] = \int_A \mu_{\theta|X}(B|x)d\mu_X(x) \tag{3}$$

References

Jayanta K Ghosh and RV Ramamoorthi. *Bayesian nonparametrics*. Springer Science & Business Media, 2003.

Mark J Schervish. *Theory of statistics*. Springer Science & Business Media, 2012.