STA414: Lecture 7 Quiz (Asymptotic Properties)

Instructor: Thibault Randrianarisoa

Questions

Question 1 In the frequentist analysis of a Bayesian posterior, which of the following best describes the core mathematical assumption being made?

a) Both the parameter θ and the data X are treated as random variables governed by a joint distribution $P(X, \theta)$.

b) The posterior distribution is evaluated assuming the data is generated from a fixed, true data-generating distribution P_{θ_0}, while treating the posterior itself as a random measure.

c) The prior distribution Π is updated iteratively until it exactly matches the empirical distribution of the data X.

d) The denominator of Bayes' formula is assumed to be exactly zero to establish the asymptotic bounds of the posterior.

Question 2 Suppose you are modeling a parameter $\theta \in \mathbb{R}$ and you choose a prior distribution Π that follows a Chi-square distribution with k degrees of freedom (χ^2_k). If the true data-generating parameter happens to be $\theta_0 = -3$, what will happen to the posterior distribution $\Pi[\cdot|X]$ as the sample size $n \to \infty$?

a) It will be consistent and converge to a normal distribution centered at -3 because the sample data will eventually overwhelm the prior according to the BvM theorem.

b) It will be inconsistent at $\theta_0 = -3$.

c) It will concentrate all its mass exactly at 0.

d) It will converge to a normal distribution centered at the maximum likelihood estimator, but with an artificially inflated variance due to the prior form.

Question 3 The Bernstein-von Mises theorem establishes that $d_{TV}(\Pi[\cdot|X], \mathcal{N}(\hat{\theta}_n(X), \frac{I(\theta_0)^{-1}}{n})) \xrightarrow{\mathbb{P}} 0$ under P_{θ_0}. When this theorem holds, which of the following is a direct consequence for large n?

a) The posterior distribution becomes exactly a normal distribution, completely independent of the choice of prior Π.

b) The area between the density curve of the posterior and the density curve of $\mathcal{N}(\hat{\theta}_n(X), \frac{I(\theta_0)^{-1}}{n})$ shrinks to zero in probability.

c) The prior distribution must have a normal form $\mathcal{N}(a, \sigma^2)$ for the theorem to apply.

d) The maximum likelihood estimator $\hat{\theta}_n(X)$ converges to the true parameter θ_0 at a rate of $1/n$.

Question 4 In the proof of posterior consistency for finite discrete models, it is essential that the Hellinger affinity $\rho(P_j, P_{\theta_0}) < 1$ for all $j \neq \theta_0$. What underlying assumption about the statistical model directly guarantees this strict inequality?

a) The model is identifiable, meaning different parameters produce strictly different probability measures.

b) The prior distribution assigns equal probability $\pi_j = 1/k$ to all possible parameters.

c) The denominator of Bayes' formula is non-zero almost surely under the marginal distribution of X.

d) The data X_i are independent and identically distributed according to a normal distribution.

Solutions

1. Correct Answer: b) The posterior distribution is evaluated assuming the data is generated from a fixed, true data-generating distribution P_{θ_0}, while treating the posterior itself as a random measure.
Reasoning: The Bayesian framework is used strictly to define the posterior distribution $\Pi[\cdot|X]$. Once defined, this distribution is studied in probability under P_{θ_0}, with $\theta_0 \in \Theta$ fixed. This means assuming the frequentist hypothesis where $X_1, ..., X_n$ are independent and identically distributed according to P_{θ_0}.

2. Correct Answer: b) It will be inconsistent at $\theta_0 = -3$ because the χ^2 prior assigns strictly zero probability mass to any negative value.
Reasoning: The support of a χ^2_k distribution is limited to non-negative numbers ($[0, \infty)$). If the true parameter is $\theta_0 = -3$, the prior density around θ_0 is zero. Because the posterior density is proportional to the prior times the likelihood, it will also be zero for all negative values. Therefore, the posterior cannot concentrate around $\theta_0 = -3$ and is completely inconsistent, mirroring the lecture example of a prior on $[0, 1]$ evaluated at $\theta_0 = 2$.

3. Correct Answer: b) The area between the density curve of the posterior and the density curve of $\mathcal{N}(\hat{\theta}_n(X), \frac{I(\theta_0)^{-1}}{n})$ shrinks to zero in probability.
Reasoning: The total variation distance $d_{TV}(P, Q)$ is defined as half the integral of the absolute difference between the probability densities p and q. Therefore, the theorem mathematically states that the L_1 distance (the area between the curves) between the posterior and the specified normal distribution converges to zero in probability under P_{θ_0}.

4. Correct Answer: a) The model is identifiable, meaning different parameters produce strictly different probability measures.
Reasoning: The proof explicitly relies on the fact that if a model is identifiable, the Hellinger distance between two different measures cannot be zero, meaning they are not equal. Because they are not equal, the Hellinger affinity $\rho(P_j, P_{\theta_0})$ is strictly less than 1 for all $j \neq \theta_0$.