STAT 206B
Chapter 4: Bayesian Point Estimation
Chapter 5: Hypothesis Testing & Confidence Regions

Winter 2022

† Bayesian Inference

- The posterior distribution supposedly contains all the available information about θ.
- The *entire* posterior distribution $\pi(\theta \mid x)$ is the extensive summary of the information available on the parameter θ.
- A visual inspection of the graph of the posterior will often provide the best insight concerning θ (at least in low dimensions)
- More standard uses of the posterior are still helpful e.g. point estimation, interval estimation, testing, prediction...
- CR Chapter 4 and JB Chapter 4.3

† **Bayesian Point Estimation:** the simplest inferential use of the posterior distribution

- Report a point estimate for $h(\theta)$, with an associated measure of accuracy

  $\Rightarrow$ Find $\pi(h(\theta) \mid x)$ and then the *Bayes rule* d, i.e., a solution of

  $$\min \mathrm{E}^{\pi}\left\{L(\theta, d) \mid x\right\} \quad \text{for } d \in \mathcal{D} \text{ and } \theta \in \Theta.$$

  ⋆⋆ Recall we found the Bayes actions under standard loss functions such as the quadratic loss, the absolute error loss and the 0-1 loss.

  ⋆⋆ The mean and median of the posterior are frequently better estimates of θ than the mode (i.e., MAP).

† Estimation Error

- We evaluate the precision of $\delta^\pi(x)$
- For example, we may use the posterior squared error:

$$\mathsf{E}^\pi[(\delta^\pi(x) - h(\theta))^2 \mid x].$$

★★ If we use $\mathsf{E}^\pi[h(\theta) \mid x]$ as the estimate of $h(\theta)$, report $\sqrt{\mathsf{Var}^\pi(h(\theta) \mid x)}$ as the standard error (posterior standard deviation).

- **JB Example 1** (p136) Consider the situation wherein a child is given an intelligence test. Assume that the tet result X is $\mathrm{N}(\theta, 100)$, where θ is the true IQ (intelligence) level of the child, as measured by the test. Assume also that, in the population as a whole, θ is distributed according to a $\mathrm{N}(100, 225)$ distribution. Suppose that we observe a student who scores 115 on the test.

  ⋆⋆ We can find

  $$\theta \mid x \sim \mathrm{N}((1/100+1/225)^{-1}(x/100+100/225), (1/100+1/225)^{-1}).$$

  $\Rightarrow$ $\mu^{\pi}(115) = 110.39$ and $\sqrt{V^{\pi}(115)} = \sqrt{69.23} = 8.32$.

- **JB Example 8**(p137) Assume $X \sim \mathrm{N}(\theta, \sigma^2)$ (σ^2 known) and the noninformative prior $\pi(\theta) = 1$ is used, then the posterior distribution of θ given x is $\mathrm{N}(x, \sigma^2)$. Hence the posterior mean is $\mu^\pi(x) = x$ and the posterior variance and standard deviation are σ^2 and σ, respectively.

  $\star\star$ The same as the usual classical estimate with standard error.

  $\star\star$ Their interpretations are different!

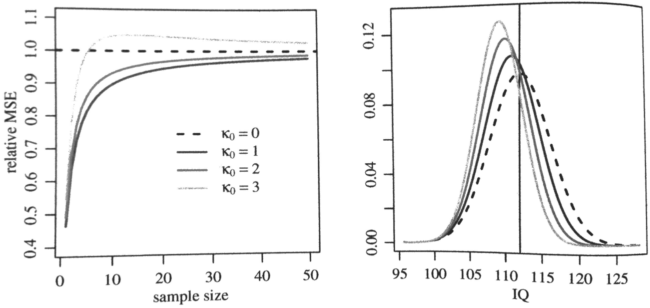
- Sampling Properties
  - ⋆⋆ Sampling properties: behavior of an estimator under hypothetically repeatable surveys or experiments.
  - ⋆⋆ Suppose $\theta_0 =$ the true value of the population mean.
  - ⋆⋆ To evaluate how close an estimator $\delta(x)$ is likely to be to θ_0, we use the mean square error(MSE)

$$\begin{aligned} \mathrm{MSE}(\delta \mid \theta_0) &= \mathrm{E}\{(\delta - \theta_0)^2 \mid \theta_0\} \\ &= \mathrm{E}\{(\delta - m)^2 \mid \theta_0\} + \mathrm{E}\{(m - \theta_0)^2 \mid \theta_0\} \\ &= \mathrm{Var}(\delta \mid \theta_0) + \mathrm{Bias}^2(\delta \mid \theta_0), \end{aligned}$$

where $m = \mathrm{E}(\delta \mid \theta_0)$

- **PH p82** Recall the IQ example (similar but different!).
  - ⋆⋆ $X \sim \mathrm{N}(100, 225)$ for the general population.
  - ⋆⋆ Suppose that we sample n individuals from a particular town and estimate θ, the town-specific mean IQ score based on the sample of size n.
  - ⋆⋆ In fact, people in the town are extremely exceptional so $\theta_0 = 112$ and $\sigma^2 = 169$.
  - ⋆⋆ Consider $x_i \mid \theta \overset{iid}{\sim} \mathrm{N}(\theta, \sigma^2)$, where $\sigma^2 = 169$ but θ is unknown
  - ⋆⋆ Assume $\theta \sim \mathrm{N}(\mu_0, \tau_0^2)$, where $\tau_0 = \sigma/\sqrt{\kappa_0}$
  - ⋆⋆ For Bayesian inference, we lack the information about the town a natural choice of $\mu_0 = 100$.

- **PH p82** Example: IQ Scores.
  - ⋆⋆ Let $\kappa_0 = \sigma^2/\tau^2$ and compare $\mathrm{MSE}(\delta_n^\pi \mid \theta_0)$ and $\mathrm{MSE}(\delta_n \mid \theta_0)$ by varying n and κ_0.
  - ⋆⋆ MSE errors and sampling distribution of different $\delta_n^\pi(x)$



- Comments on unbiasedness

  ⋆⋆ No Bayes estimate with respect to the squared error loss can be unbiased, except in a case when its Bayes' risk is 0 (that is, the perfect estimation is possible).

  $\Leftrightarrow$ If $\delta^\pi(x)$ is unbiased for θ, then $\delta^\pi(x)$ is not Bayes under the squared error loss unless its Bayes risk is zero.

  For your practice, show this.

  ⋆⋆ **No problem!** Even frequentist agree that insisting on unbiasedness can lead to bad estimators, and that in their quest to minimize the risk by trading off between variance and bias-squared a small dosage of bias can help.

† Interval Estimation (CR 5.5 and JB 4.3.2)

- $(1-\alpha)100\%$ confidence intervals (CI's)–Classical interval estimate

  ⋆⋆ Generate data from the assumed model many times and for each data set to exhibit the CI.

  ⋆⋆ Now, the proportion of CIs covering the unknown parameter "tends to" $1-\alpha$.

- We will construct $C_x \subset \Theta$ where θ should be with high probability.

  ⋆⋆ The distribution used to assess the credibility of an interval estimator is the posterior distribution.

† Credible Sets

- Credible Set: Assume the set C_x is a subset of Θ. Then C_x is a credible set with credibility $(1-\alpha)\cdot 100\%$ if

$$P^{\pi}(\theta \in C_x \mid x) = \mathrm{E}^{\pi}\{1(\theta \in C_x) \mid x\} = \int_{C_x} \pi(\theta \mid x) d\theta > 1-\alpha.$$

  ⋆⋆ If the posterior is discrete, then the integral becomes sum.

- Bayesian interpretation of a credible set C_x is natural: The probability of a parameter belonging to the set C_x is $1-\alpha$.

  ⋆⋆ The frequentist CI is random but our credible interval is fixed given data.

† Credible Sets (contd)

- For a given posterior function such set is not unique.

  ⋆⋆ *Q: How to choose one particular set?*

- For a given credibility level $(1-\alpha)100\%$, the shortest credible set is of interest.

- The size of a set is simply its total length if the parameter space Θ is one dimensional, total area, if Θ is two dimensional, and so on.

- To minimize the size, sets should correspond to highest posterior probability (density) areas.

† Credible Sets (contd)

- The $(1-\alpha)100\%$ HPD (high posterior density) credible set for parameter θ is a set C_x, subset of Θ of the form

$$C_x = \{\theta \in \Theta \mid \pi(\theta \mid x) \geq k_\alpha\},$$

  where k_α is the largest constant for which

$$P^\pi(\theta \in C_x \mid x) \geq 1-\alpha.$$

- Geometrically, if the posterior density is cut by a horizontal line at the height k_α, the set C is projection on the θ axis of the part of line inside the density, i.e., the part that lies below the density.

- See **Def 5.5.2**

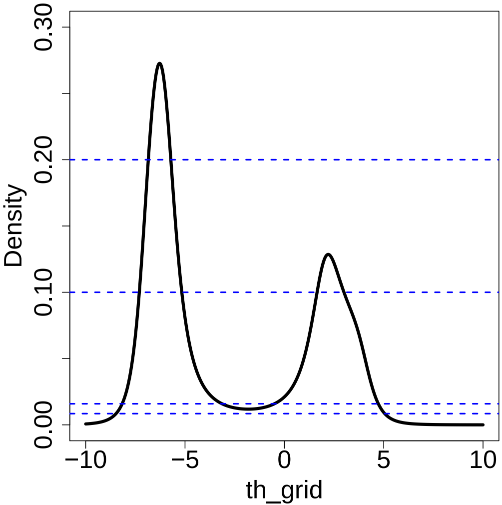
- **Example 5.5.3** Consider $x \sim \mathrm{N}(\theta, \sigma^2)$. Consider $\theta \sim \mathrm{N}(0, \tau^2)$. Find the $100(1-\alpha)\%$ HPD credible interval.

$\star\star$ Find the $100(1-\alpha)\%$ HPD credible interval with $\pi(\theta) \propto 1$.

$\star\star$ Note that we can use improper priors in this setting and do not encounter the same difficulties as when testing the point-null hypothesis.

- **JB Example 10** (p141, with a slight change) Assume that four observations, $x_i = 2, -7, 4, -6$, $i = 1, \ldots, 4$ are sampled from Cauchy $C(\theta, 1)$ distribution with parameter of interest θ ($f(x \mid \theta) = 1/\{\pi(1 + (x - \theta)^2)\}$. Consider the flat prior $\pi(\theta) = 1$. Sketch the posterior.

- **Example** (contd) The posterior is bimodal!



• **Example** (contd) Four horizontal lines at levels $k = 0.008475$, 0.0159, 0.1, and 0.2 are shown. These lines determine four credible sets,

⋆⋆ $k_{0.01} = 0.008475$: $[-8.498, 5.077]$ with $P^{\theta|X}(8.498 \leq \theta \leq 5.077) = 99\%$;

⋆⋆ $k_{0.05} = 0.0159$: $[-8.189, -3.022] \cup [-0.615, 4.755]$ with posterior credibility of 95%;

⋆⋆ $k = 0.1$: $[-7.328, -5.124] \cup [1.591, 3.120]$ with posterior credibility of 64.2%;

⋆⋆ $k = 0.2$: $[-6.893, -5.667]$ with posterior credibility of 31.3%.

- **Example** (contd)
  - ⋆⋆ Observe for $\alpha = 0.05$ and 0.1, the credible intervals consist of two separate intervals.
  - ⋆⋆ This may indicate that the prior is not agreeing with the data (unimodal in the prior vs bimodal in data).
  - ⋆⋆ There is no frequentist counterpart for the CI for θ in the above model.

- **Example** Let $x \mid \theta$ be the shifted exponential with density

$$f(x \mid \theta) = \exp\{-(x-\theta)\}1(\theta \leq x).$$

Let θ be half-Cauchy,

$$\pi(\theta) = \frac{2}{\pi(1+\theta^2)}, \quad \theta > 0.$$

Find the posterior and show that $(1-\alpha)100\%$ HPD credible set is of the form $[\beta, x]$ for some $\beta \in (0, x)$.

- **Example** Let $\eta = e^{\theta}$ and find the posterior $\pi^{\star}(\eta \mid x)$. Show that $\pi^{\star}(\eta \mid x)$ is decreasing in η and that the credible set for η is of the form $[1, \gamma]$, for some $\gamma < e^{x}$.

$\star\star$ Transform the interval of η back to the space of θ and observe $[\log 1, \log \gamma] = [0, \beta'] \neq [\beta, x]$.

- One undesirable property of credible sets is the lack of invariance with respect to monotone transformations.

- For a solution, read JB pages 144-145.

- A HPD credible sets can be found for multivariate cases. See JB p143

† Predictive Inference

- Predict a random variable $y \sim g(y \mid \theta)$ based on observations of $x \sim f(x \mid \theta)$.

  ⋆⋆ no need to be $g = f$

  ⋆⋆ easily can be extended to the case of $y \sim g(y \mid \theta, x)$.

- Find the predictive density of y given x, when the prior for θ is π,

$$p(y \mid x) = \int_{\Theta} g(y \mid \theta)\pi(\theta \mid x) d\theta.$$

† Predictive Inference (contd)

- Point estimation: use the loss function and find the Bayes actions minimizing $\mathrm{E}(L(y, a) \mid x)$.
- Posterior predictive interval for y.