DSA 8505: Bayesian Statistics
Marking Scheme

Master of Science in Data Science & Analytics

14th Feb 2025

1. **Bayesian Credible Interval**

Step 1: Understanding the Posterior Distribution

We are given that the posterior distribution of θ is:

$$\theta|x \sim \mathcal{N}(\mu, \sigma^2).$$

This means that θ follows a normal distribution with:

- **Mean:** μ (posterior mean),
- **Variance:** σ^2 (posterior variance),
- **Standard deviation:** σ (posterior standard deviation).

$\theta \Rightarrow [\theta_L, \theta_U]$

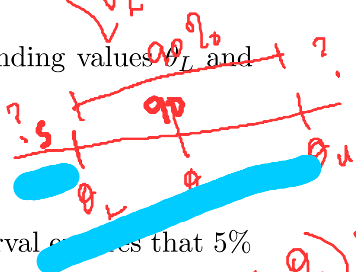
$\theta|x$

$[\theta_L \leq \theta \leq \theta_U | x]$

Step 2: Definition of Equal-Tailed Credible Interval

The **90% equal-tailed credible interval** for θ is determined by finding values θ_L and θ_U such that:

$$P(\theta_L \leq \theta \leq \theta_U|x) = 0.90.$$



Since the normal distribution is symmetric, the **equal-tailed** interval ensures that 5% of the probability mass is in each tail:

$$P(\theta \leq \theta_L|x) = 0.05, \quad P(\theta \geq \theta_U|x) = 0.05.$$

$Pr(\theta \leq \theta_L) = 5\%$

$Pr(\theta \leq \theta_U) = 95\%$

This means we need to find the **5th percentile** and **95th percentile** of the normal posterior distribution.

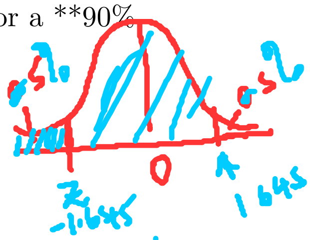
Step 3: Computing the Credible Interval

For a **standard normal distribution** $Z \sim \mathcal{N}(0, 1)$, the **critical values** for a **90% interval** are:

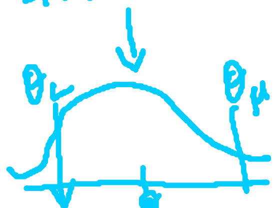
$Pr(Z \leq -1.645) = 5\%$

$Pr(Z \leq 1.645)$

$$z_{\alpha/2} = z_{0.05} \approx -1.645, \quad z_{1-\alpha/2} = z_{0.95} \approx 1.645.$$



Since θ follows $\mathcal{N}(\mu, \sigma^2)$, we apply the transformation:



$Z = \frac{\theta - \mu}{\sigma}$

$z\sigma = \theta - \mu$

$\theta = \mu \pm z\sigma$

$$\theta_L = \mu + z_{0.05} \cdot \sigma = \mu - 1.645\sigma.$$

$$\theta_U = \mu + z_{0.95} \cdot \sigma = \mu + 1.645\sigma.$$

Thus, the **90% credible interval** for θ is:

$[\theta_L, \theta_U]$

$$(\mu - 1.645\sigma, \mu + 1.645\sigma).$$

Final Answer

$$\theta \in (\mu - 1.645\sigma, \mu + 1.645\sigma)\,.$$

2. **Bayesian Updating with Beta Prior**

Step 1: Identify the Prior Distribution

$p(\theta)\quad \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$

$p(\theta) = (K)\,\theta^{\alpha-1}(1-\theta)^{\beta-1}$

The prior distribution for θ is given as:

$$\theta \sim \text{Beta}(\alpha, \beta) = \text{Beta}(3, 3).$$

This means the prior probability density function (PDF) is:

$$\boxed{p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1},} \quad 0 \leq \theta \leq 1.$$

Step 2: Identify the Likelihood Function

Since the outcome of each die roll is independent and follows a **Bernoulli distribution** (success: rolling a six), the likelihood function for observing **6 sixes in 15 rolls** follows a **Binomial distribution**:

$$p(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}.$$

$$\boxed{p(x|\theta) \propto \theta^x (1-\theta)^{n-x}}$$

Substituting $x = 6$ and $n = 15$:

$$p(6|\theta) \propto \theta^6(1-\theta)^9.$$

Step 3: Compute the Posterior Distribution

Using Bayes' theorem:

$$p(\theta|x) \propto p(\theta) \cdot p(x|\theta).$$

Substituting the **prior** and **likelihood**:

$$p(\theta|x) \propto \theta^{(3-1)}(1-\theta)^{(3-1)} \cdot \theta^6(1-\theta)^9.$$

$p(\theta|x) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1} \cdot \theta^x \cdot (1-\theta)^{n-x}$

$p(\theta|x) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$

$$p(\theta|x) \propto \theta^{(3+6-1)}(1-\theta)^{(3+9-1)}$$

$\alpha - 1 = 8$ | $\beta - 1 = 11$

$\alpha = 9$ | $\beta = 12$

$$p(\theta|x) \propto \theta^{8}(1-\theta)^{11}.$$

$\therefore pr(\theta|x) \sim B(9,12)$

Comparing with the **Beta distribution form**:

$$\text{Beta}(\alpha', \beta') \propto \theta^{\alpha'-1}(1-\theta)^{\beta'-1}.$$

We identify that:

$$\alpha' = 3 + 6 = 9, \quad \beta' = 3 + 9 = 12.$$

Final Answer: Posterior Distribution

$$\theta|x \sim \text{Beta}(9, 12).$$

This is the **updated belief** about θ after observing the data.

3. **Bayes Estimators**
   Given:

   $E(\theta|x) = \mu = 10$

   $$\pi(\theta|x) \sim N(10, 16).$$

   (a) **Bayes estimator under squared error loss** [**1 mark**] The Bayes estimator under squared error loss is the posterior mean:

   $$\hat{\theta}_{\text{Bayes}} = \boxed{E[\theta|x]} = 10.$$

   (b) **Bayes estimator under absolute error loss** [**1 mark**] The Bayes estimator under absolute error loss is the posterior median:

   $$\hat{\theta}_{\text{Bayes}} = \text{Median}(\theta|x) = 10.$$

   Since the normal distribution is symmetric, the mean and median are the same.

4. **MAP Estimation vs. Bayes Estimator**

   (a) **Difference between MAP and Bayes estimator** - The MAP estimate is the mode (most probable value), which is $\theta = 6$. - The Bayes estimator under squared error loss is the posterior mean, which may be different from the mode.

   (b) **Choosing estimate under squared error loss** - Since squared error loss minimizes expected error, we choose the posterior mean $\hat{\theta} = 4.5$.

5. **Bayesian Estimation with Normal Prior and Likelihood**

   (a) **Derivation of posterior distribution**
   We are given:

   - **Prior:**

     $$\theta \sim N(\mu_0, \sigma_0^2)$$

   - **Likelihood:**

     $$x|\theta \sim N(\mu, \sigma^2).$$

• posterior $\theta|x$??

$\pi(\theta|x) \propto p(\theta)\, pr(x|\theta)$

Step 1: Compute the Posterior using Bayes' Theorem

By Bayes' theorem, the posterior distribution is proportional to the product of the prior and the likelihood:

$$\pi(\theta|x) \propto p(x|\theta)\pi(\theta).$$

Expanding each term:

- **Prior distribution:**

$$\pi(\theta) = \boxed{\frac{1}{\sqrt{2\pi\sigma_0^2}}}\exp\left(-\frac{(\theta-\mu_0)^2}{2\sigma_0^2}\right).$$

$$\pi(\theta) \propto \exp\left[-\frac{1}{2}\left(\frac{\theta-\mu_0}{\sigma_0}\right)^2\right]$$

- **Likelihood function:**

$$p(x|\theta) = \boxed{\frac{1}{\sqrt{2\pi\sigma^2}}}\exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right).$$

$$p(x|\theta) \propto \exp\left[-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2\right]$$

$$\pi(\theta|x) \propto \pi(\theta)\cdot p(x|\theta)$$

Taking the product:

$$\pi(\theta|x) \propto \exp\left(-\frac{(\theta-\mu_0)^2}{2\sigma_0^2} - \frac{(x-\theta)^2}{2\sigma^2}\right).$$

$$e^a \; e^b = e^{a+b}$$

Step 2: Completing the Square

Expanding the exponents:

$$-\frac{(\theta-\mu_0)^2}{2\sigma_0^2} - \frac{(x-\theta)^2}{2\sigma^2}.$$

$$ax^2+bx+\frac{b^2}{4a^2}$$

Rewriting each quadratic term:

$$-\frac{\theta^2-2\mu_0\theta+\mu_0^2}{2\sigma_0^2} - \frac{x^2-2x\theta+\theta^2}{2\sigma^2}.$$

$$ax^2+bx$$

$$\left(ax^2+\frac{b}{a}x\right)$$

Rearranging:

$$-\frac{\theta^2}{2\sigma_0^2} + \frac{\mu_0\theta}{\sigma_0^2} - \frac{\mu_0^2}{2\sigma_0^2} - \frac{x^2}{2\sigma^2} + \frac{x\theta}{\sigma^2} - \frac{\theta^2}{2\sigma^2}.$$

Factor out θ^2:

$$-\theta^2\left(\frac{1}{2\sigma_0^2}+\frac{1}{2\sigma^2}\right) + \theta\left(\frac{\mu_0}{\sigma_0^2}+\frac{x}{\sigma^2}\right) + \text{(constant terms)}.$$

Since this is the exponent of a normal distribution, the posterior mean and variance are:

$$-\frac{1}{2}\left[\left(\frac{x-\theta}{\sigma}\right)^2\right]$$

Step 3: Compute Posterior Mean and Variance

- **Posterior Variance**:

$$\boxed{\sigma_{\text{post}}^2 = \frac{1}{\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2}}}.$$

- **Posterior Mean**:

$$\mu_{\text{post}} = \sigma_{\text{post}}^2\left(\frac{\mu_0}{\sigma_0^2}+\frac{x}{\sigma^2}\right).$$

$$\boxed{\sigma^2_{post} = \frac{1}{\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2}}}$$

Substituting σ_{post}^2:

$$\mu_{\text{post}} = \boxed{\frac{\frac{\mu_0}{\sigma_0^2}+\frac{x}{\sigma^2}}{\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2}}}.$$

$$\pi(\theta|x) \sim N(\mu_{post}, \sigma^2_{post})$$

Therefore, the posterior distribution is:

$$\theta|x \sim N\left(\frac{\frac{\mu_0}{\sigma_0^2}+\frac{x}{\sigma^2}}{\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2}}, \frac{1}{\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2}}\right).$$

(b) **Posterior mean minimizes squared error loss** We want to show that the **posterior mean** is the Bayesian estimator under **squared error loss**. That is, given the loss function:

$$L(\theta,\hat{\theta}) = (\theta-\hat{\theta})^2,$$

$E(x) = \int_{-\infty}^{\infty} x\, f(x)\, dx$

$E(L)$

we aim to minimize the expected posterior loss:

$$R(\hat{\theta}) = \boxed{E_\pi\left[(\theta-\hat{\theta})^2|x\right]}.$$

$\int_{-\infty}^{\infty} (\theta-\hat{\theta})^2|x \cdot \pi(\theta|x)\, d\theta$

posterior distr.

Step 1: Expected Posterior Loss Function

The expected squared error loss is:

$$R(\hat{\theta}) = E_\pi\left[(\theta-\hat{\theta})^2|x\right].$$

Expanding the expectation:

$$R(\hat{\theta}) = \int_{-\infty}^{\infty} \boxed{(\theta-\hat{\theta})^2}\pi(\theta|x)d\theta.$$

≈ 0

Step 2: Expanding the Integral

$a^2 + 2ab + b^2$

Expanding the squared term:

$$(\theta-\hat{\theta})^2 = (\underbrace{\theta - E[\theta|x]}_{a} + \underbrace{E[\theta|x]-\hat{\theta}}_{b})^2.$$

$(a+b)^2 = a^2+2ab+b^2$

$E\, ax = a\, Ex$

Expanding further:

$$E_\pi (\theta-\hat{\theta})^2 = \underbrace{(\theta-E[\theta|x])^2}_{a^2} + \underbrace{2(\theta-E[\theta|x])}_{2ab}\underbrace{E[\theta|x]-\hat{\theta}}_{???} + \underbrace{(E[\theta|x]-\hat{\theta})^2}_{b^2}.$$

Step 3: Taking Expectation

$E\, 2k[\theta - E\theta|x]$

$E_\pi(\theta - E\theta(x)] = 0$

Taking expectation on both sides:

$2k E[\theta - h]$

$$E_\pi\left[(\theta-\hat{\theta})^2|x\right] = E_\pi\left[(\theta-E[\theta|x])^2|x\right] + 2E_\pi\left[(\theta-E[\theta|x])(E[\theta|x]-\hat{\theta})|x\right] + (E[\theta|x]-\hat{\theta})^2.$$

$2k\{E\theta|x) - h$

$h - h = 0$

The second term is zero because:

$$E_\pi[(\theta-E[\theta|x])] = 0.$$

$E\{E(x)\}$

$E(x)$

This simplifies to:

$x = \{0, 5, 10\}$

$$E_\pi\left[(\theta-\hat{\theta})^2|x\right] = E_\pi\left[(\theta-E[\theta|x])^2|x\right] + (E[\theta|x]-\hat{\theta})^2.$$

$E(\hat{\theta}) \to \hat{\theta}$

$E(x) = \text{mean} = 5$

$E(\theta|x) - \hat{\theta} = 0$

$\hat{\theta} = E(\theta|x)$ #

$= (E(\theta|x) - \hat{\theta}]^2$

$E_\pi((\theta - \hat{\theta})|x = 0 \quad \therefore E(\theta|x) - \hat{\theta} = 0$

$\hat{\theta} = E(\theta|x)$ #

Step 4: Minimizing the Risk Function

Since the first term is independent of $\hat{\theta}$, the only term affecting the choice of $\hat{\theta}$ is $(E[\theta|x] - \hat{\theta})^2$, which is minimized when:

$$\hat{\theta} = E[\theta|x].$$

Thus, the posterior mean:

$$\hat{\theta}_{\text{Bayes}} = E[\theta|x]$$

minimizes the expected squared error loss.

Since the squared error loss function penalizes deviations quadratically, the **optimal estimator is the posterior mean**. Hence, the **Bayesian estimator under squared error loss is the posterior mean**.

$$\hat{\theta}_{\text{Bayes}} = E[\theta|x].$$

(c) **Compute Bayesian estimator** Given:

$$\theta \sim N(6, 2), \quad x|\theta \sim N(12, 3),$$

$\theta|x \sim N(\mu_{post}, \sigma^2_{post})$

we compute the posterior mean:

$$\mu_{\text{post}} = \frac{\frac{6}{2} + \frac{12}{3}}{\frac{1}{2} + \frac{1}{3}}.$$

Computing step by step:

$$\frac{6}{2} = 3, \quad \frac{12}{3} = 4, \quad \frac{1}{2} = 0.5, \quad \frac{1}{3} = 0.3333.$$

$$\mu_{\text{post}} = \frac{3 + 4}{0.5 + 0.3333} = \frac{7}{0.8333} = 8.4.$$