# Homework 12: Bayesian Statistical Inference

## Problem 1

The number of minutes between successive bus arrivals at Alvin's bus stop is exponentially distributed with parameter $\Theta$, and Alvin's prior PDF of $\Theta$ is:

$$f_\Theta(\theta) = \begin{cases} 10\theta & \text{if } \theta \in [0, 1/\sqrt{5}], \\ 0 & \text{otherwise.} \end{cases}$$

(a) Alvin arrives on Monday at the bus stop and has to wait 30 minutes for the bus to arrive. What is the posterior PDF, the MAP estimate, and the conditional expectation estimate of $\Theta$?

(b) Following his Monday experience, Alvin decides to estimate $\Theta$ more accurately and records his waiting times for five days. These are 30, 25, 15, 40, and 20 minutes, and Alvin assumes that his observations are independent. What is the posterior PDF, the MAP estimate, and the conditional expectation estimate of $\Theta$ given the five-day data?

## Solution 1

(a) Let $X$ denote the random wait time. We observe $X = 30$. Using Bayes' rule, the posterior PDF is:

$$f_{\Theta|X}(\theta|30) = \frac{f_\Theta(\theta) f_{X|\Theta}(30|\theta)}{\int f_\Theta(\theta') f_{X|\Theta}(30|\theta') d\theta'}.$$

Using the given prior $f_\Theta(\theta) = 10\theta$ and likelihood $f_{X|\Theta}(30|\theta) = \theta e^{-30\theta}$ for $\theta \in [0, 1/\sqrt{5}]$:

$$f_{\Theta|X}(\theta|30) = \begin{cases} \frac{\theta^2 e^{-30\theta}}{\int_0^{1/\sqrt{5}} (\theta')^2 e^{-30\theta'} d\theta'} & \text{if } \theta \in [0, 1/\sqrt{5}], \\ 0 & \text{otherwise.} \end{cases}$$

The MAP rule selects $\hat{\theta}$ that maximizes the numerator. Setting the derivative to 0:

$$\frac{d}{d\theta}(\theta^2 e^{-30\theta}) = 2\theta e^{-30\theta} - 30\theta^2 e^{-30\theta} = (2 - 30\theta)\theta e^{-30\theta} = 0.$$

Therefore, $\hat{\theta}_{MAP} = \frac{2}{30} = \frac{1}{15}$.

The conditional expectation estimator is:

$$E[\Theta|X = 30] = \frac{\int_0^{1/\sqrt{5}} \theta^3 e^{-30\theta} d\theta}{\int_0^{1/\sqrt{5}} (\theta')^2 e^{-30\theta'} d\theta'}.$$

(b) Let $X_i$ denote the random wait time for the $i$-th day. We observe vector $x = (30, 25, 15, 40, 20)$. Due to independence:

$$f_{X|\Theta}(x|\theta) = \prod_{i=1}^{5} \theta e^{-x_i\theta} = \theta^5 e^{-(30+25+15+40+20)\theta} = \theta^5 e^{-130\theta}.$$

Using the prior, the posterior is:

$$f_{\Theta|X}(\theta|x) = \begin{cases} \frac{\theta^6 e^{-130\theta}}{\int_0^{1/\sqrt{5}} (\theta')^6 e^{-130\theta'} d\theta'} & \text{if } \theta \in [0, 1/\sqrt{5}], \\ 0 & \text{otherwise.} \end{cases}$$

For MAP, set the derivative of the numerator to 0:

$$\frac{d}{d\theta}(\theta^6 e^{-130\theta}) = 6\theta^5 e^{-130\theta} - 130\theta^6 e^{-130\theta} = (6 - 130\theta)\theta^5 e^{-130\theta} = 0.$$

Therefore, $\hat{\theta}_{MAP} = \frac{6}{130} = \frac{3}{65}$.

The conditional expectation estimator is:

$$E[\Theta|X = x] = \frac{\int_0^{1/\sqrt{5}} \theta^7 e^{-130\theta} d\theta}{\int_0^{1/\sqrt{5}} (\theta')^6 e^{-130\theta'} d\theta'}.$$

---

## Problem 2

Students in a probability class take a multiple-choice test with 10 questions and 3 choices per question. A student who knows the answer to a question will answer it correctly, while a student that does not will guess the answer with probability of success 1/3. Each student is equally likely to belong to one of three categories $i = 1, 2, 3$: those who know the answer to each question with corresponding probabilities $\theta_i$, where $\theta_1 = 0.3$, $\theta_2 = 0.7$, and $\theta_3 = 0.95$ (independent of other questions). Suppose that a randomly chosen student answers $k$ questions correctly.

(a) For each possible value of $k$, derive the MAP estimate of the category that this student belongs to.

(b) Let $M$ be the number of questions that the student knows how to answer. Derive the posterior PMF, and the MAP and LMS estimates of $M$ given that the student answered correctly 5 questions.

## Solution 2

(a) Let $X$ be the number of questions answered correctly. For each category $i$, the probability of answering correctly is:

$$p_i = \theta_i + (1 - \theta_i)\frac{1}{3} = \frac{2\theta_i + 1}{3}.$$

Calculating $p_i$: $p_1 = \frac{1.6}{3} \approx 0.533$, $p_2 = \frac{2.4}{3} = 0.8$, $p_3 = \frac{2.9}{3} \approx 0.967$. The MAP rule selects category $i$ maximizing the binomial probability $\binom{10}{k} p_i^k (1 - p_i)^{10-k}$.

(b) The posterior PMF of $M$ (number of questions known) given $X = 5$ is:

$$p_{M|X}(m|5) = \sum_{i=1}^{3} p_{\Theta|X}(\theta_i|5) P(M = m|X = 5, \Theta = \theta_i).$$

First, calculate $p_{\Theta|X}(\theta_i|5)$. The priors are uniform (1/3).

$$p_{\Theta|X}(\theta_i|5) \propto p_i^5(1 - p_i)^5.$$

Using the values calculated: $p_{\Theta|X}(\theta_1|5) \approx 0.9010$, $p_{\Theta|X}(\theta_2|5) \approx 0.0989$, $p_{\Theta|X}(\theta_3|5) \approx 0.0001$.

The probability a student knows a question given they answered correctly is $q_i = \frac{\theta_i}{p_i}$. $P(M = m|X = k, \Theta = \theta_i) = \binom{k}{m} q_i^m (1 - q_i)^{k-m}$.

The resulting posterior PMF values for $m = 0, \ldots, 5$ are approximately: $0.0145, 0.0929, 0.2402, 0.3173, 0.2335, 0.1015$.

The MAP estimate is $\hat{m} = 3$. The Conditional Expectation (LMS) is $E[M|X = 5] \approx 3$.

---

## Problem 3

Consider a variation of the biased coin problem, and assume the probability of heads, $\Theta$, is distributed over $[0, 1]$ according to the PDF:

$$f_\Theta(\theta) = 2 - 4\left|\frac{1}{2} - \theta\right|, \quad \theta \in [0, 1]$$

Find the MAP estimate of $\Theta$, assuming that $n$ independent coin tosses resulted in $k$ heads and $n - k$ tails.

## Solution 3

We maximize the posterior, which is proportional to $f_\Theta(\theta)p_{X|\Theta}(k|\theta)$:

$$(2 - 4|1/2 - \theta|)\binom{n}{k}\theta^k(1 - \theta)^{n-k}.$$

Ignoring constants, we maximize $(2 - 4|1/2 - \theta|)\theta^k(1 - \theta)^{n-k}$.

- For $\theta < 1/2$, $f_\Theta(\theta) = 4\theta$. We maximize $\theta^{k+1}(1 - \theta)^{n-k}$. Derivative is 0 at $\hat{\theta} = \frac{k+1}{n+1}$.
- For $\theta > 1/2$, $f_\Theta(\theta) = 4(1 - \theta)$. We maximize $\theta^k(1 - \theta)^{n-k+1}$. Derivative is 0 at $\hat{\theta} = \frac{k}{n+1}$.

The MAP estimate is:

$$\hat{\theta}_{MAP} = \begin{cases} \frac{k+1}{n+1} & \text{if } \frac{k+1}{n+1} < \frac{1}{2}, \\ \frac{1}{2} & \text{if } \frac{k}{n+1} \leq \frac{1}{2} \leq \frac{k+1}{n+1}, \\ \frac{k}{n+1} & \text{if } \frac{k}{n+1} > \frac{1}{2}. \end{cases}$$

---

## Problem 4

Professor May B. Hard, who has a tendency to give difficult problems in probability quizzes, is concerned about one of the problems she has prepared for an upcoming quiz. She therefore asks her TA to solve the problem and record the solution time. May's prior probability that the problem is difficult is 0.3, and she knows from experience that the conditional PDF of her TA's solution time $X$, in minutes, is:

$$f_{X|\Theta}(x|\Theta = 1) = \begin{cases} c_1 e^{-0.04x} & \text{if } 5 \leq x \leq 60, \\ 0 & \text{otherwise,} \end{cases}$$

if $\Theta = 1$ (problem is difficult), and is:

$$f_{X|\Theta}(x|\Theta = 2) = \begin{cases} c_2 e^{-0.16x} & \text{if } 5 \leq x \leq 60, \\ 0 & \text{otherwise,} \end{cases}$$

if $\Theta = 2$ (problem is not difficult), where $c_1$ and $c_2$ are normalizing constants. She uses the MAP rule to decide whether the problem is difficult.

(a) Given that the TA's solution time was 20 minutes, which hypothesis will she accept and what will be the probability of error?

(b) Not satisfied with the reliability of her decision, May asks four more TAs to solve the problem. The TAs' solution times are conditionally independent and identically distributed with the solution time of the first TA. The recorded solution times are 10, 25, 15, and 35 minutes. On the basis of the five observations, which hypothesis will she now accept, and what will be the probability of error?

## Solution 4

(a) Calculating constants: $c_1 \approx 0.0549$, $c_2 \approx 0.3561$. Posterior for Difficult ($\Theta = 1$) given $T = 20$:

$$p_{\Theta|T}(1|20) = \frac{0.3c_1 e^{-0.04(20)}}{0.3c_1 e^{-0.04(20)} + 0.7c_2 e^{-0.16(20)}} \approx 0.4214.$$

Posterior for Not Difficult ($\Theta = 2$) is $1 - 0.4214 = 0.5786$. She accepts $\Theta = 2$ (Not Difficult). Probability of error $p_e = 0.4214$.

(b) With 5 observations $(20, 10, 25, 15, 35)$:

$$p_{\Theta|T}(1|\text{data}) = \frac{0.3(c_1)^5 e^{-0.04(\sum x_i)}}{0.3(c_1)^5 e^{-0.04(\sum x_i)} + 0.7(c_2)^5 e^{-0.16(\sum x_i)}} \approx 0.9171.$$

She accepts $\Theta = 1$ (Difficult). Probability of error $p_e = 1 - 0.9171 = 0.0829$.

## Problem 5

We have two boxes, each containing three balls: one black and two white in box 1; two black and one white in box 2. We choose one of the boxes at random, where the probability of choosing box 1 is equal to some given $p$, and then draw a ball.

(a) Describe the MAP rule for deciding the identity of the box based on whether the drawn ball is black or white.

(b) Assuming that $p = 1/2$, find the probability of an incorrect decision and compare it with the probability of error if no ball had been drawn.

## Solution 5

(a) Let $X = 1$ (White), $X = 2$ (Black). Prior $P(\Theta_1) = p$, $P(\Theta_2) = 1 - p$. Likelihoods: $P(X = 1|\Theta_1) = 2/3$, $P(X = 1|\Theta_2) = 1/3$. Posterior for $X = 1$: $P(\Theta_1|1) = \frac{2p/3}{2p/3+(1-p)/3} = \frac{2p}{1+p}$. MAP Rule:

- If White drawn: Select Box 1 if $\frac{2p}{1+p} > \frac{1-p}{1+p} \implies p > 1/3$.
- If Black drawn: Select Box 1 if $p > 2/3$.

(b) If $p = 1/2$, we select Box 1 if White ($X = 1$) and Box 2 if Black ($X = 2$). Error given Box 1 chosen: $P(X = 2|\Theta_1) = 1/3$. Error given Box 2 chosen: $P(X = 1|\Theta_2) = 1/3$. Total Prob Error $= (1/2)(1/3)+(1/2)(1/3) = 1/3$. Without data, error is $1/2$. Data reduced error to $1/3$.

---

## Problem 6

The probability of heads of a given coin is known to be either $q_0$ (hypothesis $H_0$) or $q_1$ (hypothesis $H_1$). We toss the coin repeatedly and independently, and record the number of heads before a tail is observed for the first time. We assume that $0 < q_0 < q_1 < 1$, and that we are given prior probabilities $P(H_0)$ and $P(H_1)$. For parts (a) and (b), we also assume that $P(H_0) = P(H_1) = 1/2$.

(a) Calculate the probability that hypothesis $H_1$ is true, given that there were exactly $k$ heads before the first tail.

(b) Consider the decision rule that decides in favor of hypothesis $H_1$ if $k \geq k^*$, where $k^*$ is some nonnegative integer, and decides in favor of hypothesis $H_0$ otherwise. Give a formula for the probability of error in terms of $k^*$, $q_0$, and $q_1$. For what value of $k^*$ is the probability of error minimized? Is there another type of decision rule that would lead to an even lower probability of error?

(c) Assume that $q_0 = 0.3$, $q_1 = 0.7$, and $P(H_1) > 0.7$. How does the optimal choice of $k^*$ (the one that minimizes the probability of error) change as $P(H_1)$ increases from 0.7 to 1.0?

## Solution 6

(a) Event $K = k$ is $k$ heads then tail. $p_{K|H_i}(k) = (1 - q_i)q_i^k$.

$$P(H_1|K = k) = \frac{\frac{1}{2}(1 - q_1)q_1^k}{\frac{1}{2}(1 - q_1)q_1^k + \frac{1}{2}(1 - q_0)q_0^k} = \frac{(1 - q_1)q_1^k}{(1 - q_1)q_1^k + (1 - q_0)q_0^k}.$$

(b) Probability of error:

$$p_e = \frac{1}{2}(1 + q_0^{k^*} - q_1^{k^*}).$$

To minimize, treating $k^*$ as continuous, we differentiate. The optimal real value is:

$$\bar{k} = \frac{\ln(|\ln q_0|) - \ln(|\ln q_1|)}{|\ln q_0| - |\ln q_1|}.$$

The integer $k^*$ is either $\lfloor \bar{k} \rfloor$ or $\lceil \bar{k} \rceil$. This is the MAP rule, so no other rule yields lower error.

(c) With $q_0 = 0.3, q_1 = 0.7$: $\bar{k} \approx 0.43$. Optimal $k^*$ is 0 or 1 (error is 0.3 for both). As $P(H_1)$ increases, $\bar{k}$ decreases. $k^* = 0$ remains optimal. We always decide $H_1$.

---

## Problem 7

Suppose that the signal $X \sim N(0, \sigma_X^2)$ is transmitted over a communication channel. Assume that the received signal is given by:

$$Y = aX + bW, \qquad W \sim N(0, \sigma_W^2),$$

and $W$ is independent of $X$. (a) Find the ML estimate of $X$, given $Y = y$. (b) Find the MAP estimate of $X$, given $Y = y$.

## Solution 7

(a) The received signal is

$$Y = aX + bW.$$

Conditional on $X = x$, this is a linear transformation of $W$. Using the MGF of a Gaussian random variable, one obtains:

$$Y \mid X = x \sim N(ax,\, b^2\sigma_W^2).$$

The log-likelihood is

$$\log f_{Y|X}(y \mid x) = -\frac{1}{2}\log(2\pi b^2\sigma_W^2) - \frac{(y - ax)^2}{2b^2\sigma_W^2}.$$

Ignoring constants, maximizing the likelihood is equivalent to minimizing $(y - ax)^2$.

Differentiate and set to zero:

$$\frac{d}{dx}(y - ax)^2 = -2a(y - ax) = 0 \quad \Rightarrow \quad x = \frac{y}{a}.$$

Thus, the ML estimate is:

$$\hat{x}_{ML} = \frac{y}{a}.$$

(b) Using Bayes' theorem:

$$f_{X|Y}(x \mid y) \propto f_{Y|X}(y \mid x) f_X(x),$$

where:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{x^2}{2\sigma_X^2}\right),$$

$$f_{Y|X}(y \mid x) \propto \exp\left(-\frac{(y-ax)^2}{2b^2\sigma_W^2}\right).$$

Thus,

$$f_{X|Y}(x \mid y) \propto \exp\left(-\frac{(y-ax)^2}{2b^2\sigma_W^2} - \frac{x^2}{2\sigma_X^2}\right).$$

Taking the log:

$$\log f_{X|Y}(x \mid y) = -\frac{a^2}{2b^2\sigma_W^2}x^2 + \frac{ay}{b^2\sigma_W^2}x - \frac{y^2}{2b^2\sigma_W^2} - \frac{1}{2\sigma_X^2}x^2 + \text{constant}.$$

Grouping the quadratic terms:

$$\log f_{X|Y}(x \mid y) = -\frac{1}{2}\left(\frac{a^2}{b^2\sigma_W^2} + \frac{1}{\sigma_X^2}\right)x^2 + \frac{ay}{b^2\sigma_W^2}x + \text{constant}.$$

A quadratic is maximized at:

$$x_{MAP} = \frac{B}{2A},$$

where

$$A = \frac{1}{2}\left(\frac{a^2}{b^2\sigma_W^2} + \frac{1}{\sigma_X^2}\right), \qquad B = \frac{ay}{b^2\sigma_W^2}.$$

Thus,

$$x_{MAP} = \frac{ay\sigma_X^2}{a^2\sigma_X^2 + b^2\sigma_W^2}.$$

Therefore, the MAP estimate is:

$$\hat{x}_{MAP} = \frac{ay\sigma_X^2}{a^2\sigma_X^2 + b^2\sigma_W^2}.$$

---

## Problem 8

Let $\Theta$ be a continuous random variable with pdf

$$f_\Theta(\theta) = \frac{1}{6}, \qquad 4 \le \theta \le 10,$$

and 0 elsewhere. Suppose that $X = \Theta + U$, with $U \sim \text{Uniform}[-1, 1]$ independent of $\Theta$. Find the conditional expectation estimator $E[\Theta \mid X = x]$.

## Solution 8

1. Since $\Theta \sim \text{Uniform}[4, 10]$,

$$f_\Theta(\theta) = \frac{1}{6}, \qquad 4 \leq \theta \leq 10.$$

2. Because $X = \Theta + U$ and $U \sim \text{Uniform}[-1, 1]$,

$$f_{X|\Theta}(x \mid \theta) = \begin{cases} \frac{1}{2}, & \theta - 1 \leq x \leq \theta + 1, \\ 0, & \text{otherwise.} \end{cases}$$

Thus the joint pdf is:

$$f_{X,\Theta}(x, \theta) = f_\Theta(\theta) f_{X|\Theta}(x \mid \theta) = \frac{1}{12},$$

for $4 \leq \theta \leq 10$ and $\theta - 1 \leq x \leq \theta + 1$.

(b) To obtain $f_X(x)$:

$$f_X(x) = \int f_{X,\Theta}(x, \theta)\, d\theta = \int_{\max(4,x-1)}^{\min(10,x+1)} \frac{1}{12}\, d\theta.$$

Evaluating gives:

$$f_X(x) = \begin{cases} \frac{x-3}{12}, & 3 \leq x < 5, \\ \frac{1}{6}, & 5 \leq x \leq 9, \\ \frac{11-x}{12}, & 9 < x \leq 11, \\ 0, & \text{otherwise.} \end{cases}$$

(c) The conditional pdf is:

$$f_{\Theta|X}(\theta \mid x) = \frac{f_{X,\Theta}(x, \theta)}{f_X(x)} = \frac{\frac{1}{12}}{f_X(x)}, \qquad \theta \in [\max(4, x - 1),\, \min(10, x + 1)].$$

(d) The conditional expectation is:

$$E[\Theta \mid X = x] = \frac{1}{f_X(x)} \int_{\max(4,x-1)}^{\min(10,x+1)} \frac{\theta}{12}\, d\theta.$$

Compute:

$$\int \frac{\theta}{12}\, d\theta = \frac{\theta^2}{24}.$$

Thus,

$$E[\Theta \mid X = x] = \frac{1}{f_X(x)} \left[ \frac{\min(10, x + 1)^2}{24} - \frac{\max(4, x - 1)^2}{24} \right].$$

## Problem 9

Show that the conditional expectation estimator is the best guess in the sense of minimizing mean squared error, i.e.,

$$E[(\theta - E[\theta|X])^2 \mid X] \leq E[(\theta - g(X))^2 \mid X].$$

What is the estimated value of the conditional expectation?

## Solution 9

First consider the simpler case with no conditioning. We wish to find $c$ minimizing $E[(\theta - c)^2]$:

$$E[(\theta - c)^2] = E[\theta^2] + c^2 - 2cE[\theta].$$

Differentiating with respect to $c$ and setting to zero gives:

$$c = E[\theta].$$

Thus,

$$E[(\theta - E[\theta])^2] \leq E[(\theta - c)^2].$$

Now we enter a conditional world and seek $c$ minimizing

$$E[(\theta - c)^2 \mid X = x].$$

Repeating the same argument shows the minimizer is:

$$c = E[\theta \mid X = x].$$

Thus the conditional expectation is the MMSE estimator. Using the law of iterated expectations:

$$E[X_{MM}] = E[E[X \mid Y]] = E[X].$$

To better understand the reasoning, refer to this.