The Austalian National University
Research School of Computer Science
Liang Zheng

Semester 2, 2020
Assignment 4 Theory Questions

COMP3670: Introduction to Machine Learning

Errata: All corrections are in red.

Note: For the purposes of this assignment, if X is a random variable we let p_X denote the probability density function (pdf) of X, F_X to denote it's cumulative distribution function, and P to denote probabilities. These can all be related as follows:

$$P(X \leq x) = F_X(x) = \int_{-\infty}^{x} p_X(z)dz$$

$$P(a \leq X \leq b) = F_X(b) - F_X(a) = \int_{a}^{b} p_X(z)dz$$

Often, we will simply write p_X as p, where it's clear what random variable the distribution refers to. You should show your derivations, but **you may use a computer algebra system (CAS)** to assist with integration or differentiation. We are not assessing your ability to integrate/differentiate here.[1].

Question 1 Continuous Bayesian Inference 5+5+2+4+4+6+6+5=37 credits

Let X be a random variable representing the outcome of a biased coin with possible outcomes $\mathcal{X} = \{0, 1\}$, $x \in \mathcal{X}$. The bias of the coin is itself controlled by a random variable Θ, with outcomes[2] $\theta \in \boldsymbol{\theta}$, where

$$\boldsymbol{\theta} = \{\theta \in \mathbb{R} : 0 \leq \theta \leq 1\}$$

The two random variables are related by the following conditional probability distribution function of X given Θ.

$$p(X = 1 \mid \Theta = \theta) = \theta$$

$$p(X = 0 \mid \Theta = \theta) = 1 - \theta$$

We can use $p(X = 1 \mid \theta)$ as a shorthand for $p(X = 1 \mid \Theta = \theta)$.

We wish to learn what θ is, based on experiments by flipping the coin.

We flip the coin a number of times.[3] After each coin flip, we update the probability distribution for θ to reflect our new belief of the distribution on θ, based on evidence.

Suppose we flip the coin n times, and obtain the sequence of coin flips [4] $x_{1:n}$.

a) Compute the new PDF for θ after having observed n consecutive **ones** (that is, $x_{1:n}$ is a sequence where $\forall i.x_i = 1$), for an arbitrary prior pdf $p(\theta)$. Simplify your answer as much as possible.

b) Compute the new PDF for θ after having observed n consecutive **zeros**, (that is, $x_{1:n}$ is a sequence where $\forall i.x_i = 0$) for an arbitrary prior pdf $p(\theta)$. Simplify your answer as much as possible.

c) Compute $p(\theta | x_{1:n} = 1^n)$ for the uniform prior $p(\theta) = 1$.

d) Compute the expected value μ_n of θ after observing n consecutive ones, with a uniform prior $p(\theta) = 1$. Provide intuition explaining the behaviour of μ_n as $n \to \infty$.

[1] For example, asserting that $\int_0^1 x^2 \left(x^3 + 2x\right) dx = 2/3$ with no working out is adequate, as you could just plug the integral into Wolfram Alpha using the command `Integrate[x^2(x^3 + 2x),{x,0,1}]`

[2] For example, a value of $\theta = 1$ represents a coin with 1 on both sides. A value of $\theta = 0$ represnts a coin with 0 on both sides, and $\theta = 1/2$ represents a fair, unbaised coin.

[3] The coin flips are independent and identically distributed (i.i.d).

[4] We write $x_{1:n}$ as shorthand for the sequence $x_1 x_2 \dots x_n$.

e) Compute the variance σ_n^2 of the distribution of θ after observing n consecutive ones, with a uniform prior $p(\theta) = 1$. Provide intuition explaining the behaviour of σ_n^2 as $n \to \infty$.

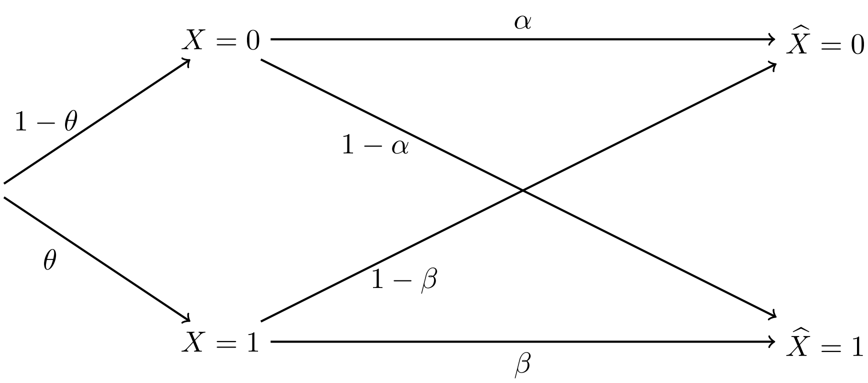
f) Compute the *maximum a posteriori* estimation θ_{MAP_n} of the distribution on θ after observing n consecutive ones, with a uniform prior $p(\theta) = 1$. Provide intuition explaining how θ_{MAP_n} varies with n.

g) Given we have observed n consecutive coin flips of ones in a row, what do you think would be a better choice for the best guess of the true value of θ? μ_n or θ_{MAP}? Justify your answer.

h) Plot the probability distributions $p(\theta|x_{1:n} = 1)$ over the interval $0 \leq \theta \leq 1$ for $n \in \{0, 1, 2, 3, 4\}$ to compare them. Assume $p(\theta) = 1$.

Question 2 Bayesian Inference on Imperfect Information (4+5+8+4+4=25 credits)

We have a Bayesian agent running on a computer, trying to learn information about what the parameter θ could be in the coin flip problem, based on observations through a noisy camera. The noisy camera takes a photo of each coin flip and reports back if the result was a 0 or a 1. Unfortunately, the camera is not perfect, and sometimes reports the wrong value.[5] The probability that the camera makes mistakes is controlled by two parameters α and β, that control the likelihood of correctly reporting a zero, and a one, respectively. Letting X denote the true outcome of the coin, and $\widehat{X}$ denoting what the camera reported back, we can draw the relationship between X and $\widehat{X}$ as shown.



So, we have

$$p(\widehat{X} = 0 \mid X = 0) = \alpha$$
$$p(\widehat{X} = 0 \mid X = 1) = 1 - \beta$$
$$p(\widehat{X} = 1 \mid X = 1) = \beta$$
$$p(\widehat{X} = 1 \mid X = 0) = 1 - \alpha$$

We would now like to investigate what posterior distributions are obtained, as a function of the parameters α and β.

a) (5 credits) Briefly comment about how the camera behaves for $\alpha = \beta = 1$, for $\alpha = \beta = 1/2$, and for $\alpha = \beta = 0$. For each of these cases, how would you expect this would change how the agent updates it's prior to a posterior on θ, given an observation of $\widehat{X}$? (No equations required.)

b) (10 credits) Compute $p(\widehat{X} = x|\theta)$ for all $x \in \{0, 1\}$.

c) (15 credits) The coin is flipped, and the camera reports seeing a one. (i.e. that $\hat{X} = 1$.) Given an arbitrary prior $p(\theta)$, compute the posterior $p(\theta|\hat{X} = 1)$. What does $p(\theta|\hat{X} = 1)$ simplify to when $\alpha = \beta = 1$? When $\alpha = \beta = 1/2$? When $\alpha = \beta = 0$? Explain your observations.

[5] The errors made by the camera are i.i.d, in that past camera outputs do not affect future camera outputs.

d) Evaluate $p(\theta|\hat{X} = 1)$ for the uniform prior $p(\theta) = 1$. Simplify it under the assumption that $\beta := \alpha$.

e) (10 credits) Let $\beta = \alpha$. Plot $p(\theta|\hat{X} = 1)$ as a function of θ, for all $\alpha \in \{0, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, 1\}$ on the same graph to compare them. Comment on how the shape of the distribution changes with α. Explain your observations. (Assume $p(\theta) = 1$.)

Question 3 **Relating Random Variables** (10+7+5+16=38 credits)

A casino offers a new game. Let $X \sim f_X$ be a random variable on $(0, 1]$ with pdf p_X. Let Y be a random variable on $[1, \infty)$ such that $Y = 1/X$. A random number c is sampled from Y, and the player guesses a number $m \in [1, \infty)$. If the player's guess m was lower than c, then the player wins $m - 1$ dollars from the casino (which means higher guesses pay out more money). But if the player guessed too high, $(m \geq c)$, they go bust, and have to pay the casino 1 dollar.

a) Show that the probability density function p_Y for Y is given by

$m < c \quad m - 1$

$$p_Y(y) = \frac{1}{y^2} p_X\left(\frac{1}{y}\right)$$

$m \geq c \quad 1$

b) Hence, or otherwise, compute the expected profit for the player under this game. Your answer will be in terms of m and p_X, and should be as simplified as possible.

c) Suppose the casino chooses a uniform distribution over $(0, 1]$ for X, that is,

$$p_X(x) = \begin{cases} 1 & 0 < x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

What strategy should the player use to maximise their expected profit?

d) Find a pdf $p_X : (0, 1] \to \mathbb{R}$ such that for any $B > 0$, there exists a corresponding player guess m such that the expected profit for the player is at least B. (That is, prove that the expected profit for p_X, as a function of m, is unbounded.)

Make sure that your choice for p_X is a valid pdf, i.e. it should satisfy

$$\int_0^1 p_X(x)dx = 1 \text{ and } p_X(x) \geq 0$$

You should also briefly mention how you came up with your choice for p_X.

Hint: We want X to be extremely biased towards small values, so that Y is likely to be large, and the player can choose higher values of m without going bust.