The Austalian National University
Research School of Computer Science
Liang Zhang

Semester 2, 2019
Assignment 5 Theory Questions

COMP3670: Introduction to Machine Learning

- **Question 1**

Let X be a random variable representing a biased coin with possible outcomes $\mathcal{X} = \{0, 1\}$. The bias of the coin is controlled by a parameter $\theta \in [0, 1]$, so

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

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

or, more compactly,

$$p(X = x) = \theta^x (1 - \theta)^{1-x}$$

We wish to learn what θ is, based on experiments by flipping the coin. Before we flip the coin, we have no knowledge about what θ could be, so we let the prior distribution on θ be uniform on the unit interval,

$$p(\theta) = \mathcal{U}([0, 1]) := \mathbb{1}_{[0,1]}(\theta)$$

where $\mathbb{1}_E(x)$ is the indicator function on the set E, defined as

$$\mathbb{1}_E(x) := \begin{cases} 1 & x \in E \\ 0 & x \notin E \end{cases}$$

We flip the coin a number of times.[1] After each coin flip, we update the distribution for θ to reflect what we learned from the coin flip.

Suppose we flip the coin three times, and obtain the sequence [2] $x_{1:3} = 101$. For each subsequence $x_1, x_{1:2}, x_{1:3}$,

a) Compute the posterior distributions;

b) The expectation values μ;

c) The variances σ^2;

d) The *maximum a posteriori* estimation θ_{MAP}.

Present your results in a table like as shown below. The first has been done for you.

Posterior	**PDF**	$\boldsymbol{\mu}$	$\boldsymbol{\sigma^2}$	$\boldsymbol{\theta_{MAP}}$
$p(\theta)$	$\mathbb{1}_{[0,1]}(\theta)$	1/2	1/12	Any θ in $[0, 1]$.
$p(\theta \| x_1 = 1)$	?	?	?	?
$p(\theta \| x_{1:2} = 10)$	?	?	?	?
$p(\theta \| x_{1:3} = 101)$	?	?	?	?

You should show your derivations, but **you may use a computer algebra system to assist with integration or differentiation**.[3].

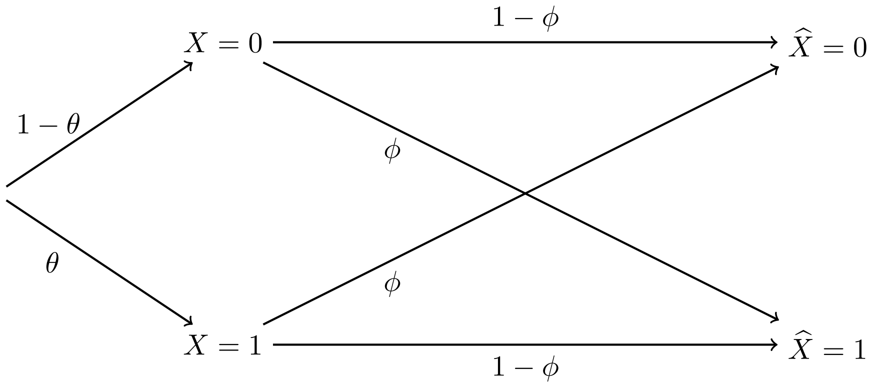
e) Plot each of the probability distributions.

[1]The coin flips are independent and identically distributed (i.i.d).
[2]We write $x_{1:n}$ as shorthand for the sequence $x_1 x_2 \dots x_n$.
[3]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}]`

f) What behaviour would you expect of the posterior distribution $p(\theta|x_{1:n})$ if we updated on a very long sequence of alternating coin flips $x_{1:n} = 10101010\ldots$?

- **Question 2**

  We have a Bayesian agent running on a computer, trying to learn information about what the parameter θ could be in the coil 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. The camera has a probability[4] of returning the wrong answer of $\phi \in [0, 0.5]$, 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

$$p(\widehat{X} \neq X \mid \phi) = \phi$$
$$p(\widehat{X} = X \mid \phi) = 1 - \phi$$

  We would now like to investigate what posterior distributions are obtained, as a function of the parameter ϕ. Let $\widehat{x}_{1:n}$ be a sequence of coin flips as observed by the camera.

  a) Briefly comment about how the camera behaves for $\phi = 0$ and $\phi = 0.5$, and how you expect this would affect Bayesian updating on the observations $\widehat{X}$.
  b) Compute $p(\widehat{X} = x|\theta, \phi)$ for all $x \in \{0, 1\}$.
  c) Given the same choice of prior $p(\theta|\phi) = \mathcal{U}([0, 1])$ as before, compute $p(\theta|\widehat{x}_{1:2} = 10, \phi)$. What term (from Question 1) does $p(\theta|\widehat{x}_{1:2} = 10, \phi)$ simplify to when $\phi = 0$? When $\phi = 0.5$? Explain your observations.
  d) Derive θ_{MAP} for the posterior $p(\theta|\widehat{x}_{1:2} = 10, \phi)$. How does θ_{MAP} depend on the parameter ϕ? Can you explain this behaviour?
  e) Plot $p(\theta|\widehat{x}_{1:2} = 10, \phi)$ as a function of θ for all $\phi \in \{0, \frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{4}{10}, \frac{5}{10}\}$ and comment on how the shape of the distribution changes with ϕ. Explain your observations.
  f) Prove[5] that for every $\phi \in [0, 0.5]$ there exists some $\alpha \in [0, 1]$ such that

$$p(\theta \mid \widehat{X} = 1, \phi) = p(\theta \mid X = 1)\alpha + p(\theta)(1 - \alpha)$$

- **Question 3**

  Consider two random variables x, y, with distribution $p(x, y)$. Show that

$$\mathbb{E}_X[x] = \mathbb{E}_Y[\mathbb{E}_X[x|y]]$$

  Note that $\mathbb{E}_X[x|y]$ is the expectation on X with respect to the conditional distribution $p(x|y)$ instead of the usual $p(x)$.

[4] The reason we don't define ϕ to be any number in the range of $[0, 1]$, is that we can always emulate any number in the range $[0.5, 1]$ by a number in $[0, 0.5]$, and then flipping the result. A camera that always says the wrong answer 100% of the time is still very useful, as you just report back the opposite of what it says.

[5] We are essentially proving that the posterior based on evidence from the noisy camera, is a convex combination of the posterior based on true evidence, and the original prior.

- **Question 4**

  Let X be a random variable, on $[0, \infty)$, with probability density function

  $$p(x) = e^{-x}$$

  Let Y be a random variable on $[0, \infty)$, such that $Y = X^2$. Find the probability density function for Y.