The Australian National University
Research School of Computer Science

Semester 1, 2021
Theory Assignment

COMP4670: Statistical Machine Learning

Maximum credit. 100

Exercise 1 **Properties of Independent Variables** 10 points

Let X and Y be *independent* continuous random variables over some domain $D \subseteq \mathbb{R}$.

1. Prove that $\mathbb{E}_{X,Y}[x+y] = \mathbb{E}_X[x] + \mathbb{E}_Y[y]$.
2. Prove that $\mathbb{E}_{X,Y}[xy] = \mathbb{E}_X[x]\mathbb{E}_Y[y]$.
3. Prove that $\text{cov}[x,y] = 0$, where

$$\text{cov}[x,y] := \mathbb{E}_{X,Y}[(x - \mathbb{E}_X[x])(y - \mathbb{E}_Y[y])]$$

Exercise 2 **Beta Priors** 30 credits

For $a, b \geq 1$, we define

$$\text{Beta}(x \mid a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$$

which is defined over the domain $x \in [0,1]$, and where B, the beta function, is defined as

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} \, \mathrm{d}x$$

1. Prove that $0 < B(a,b) \leq 1$ for all $a, b \geq 1$.
2. Prove that $\text{Beta}(x \mid a,b)$ is well defined and a valid probability distribution.

We define the gamma function $\Gamma(x)$ as follows

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} \, \mathrm{d}t$$

The Gamma function acts like a continuous form of the factorial function, and is defined for non-integer inputs. It satisfies the following identities for any $z > 0$.

$$\Gamma(z+1) = z\Gamma(z), \Gamma(1) = 1$$

and for any positive integer n, $\Gamma(n) = (n-1)!$.

The gamma function is related to the beta function via the following identity, which you may use without proof.

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

3. Prove the identity

$$B(a+1,b) = \frac{a}{a+b} B(a,b)$$

Exercise 3 — Coin Flips in Generality [1] — 40 credit

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 $\boldsymbol{\theta} = [0, 1], \theta \in \boldsymbol{\theta}$. The two random variables are related by the following conditional probability distribution function of X given Θ.

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

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

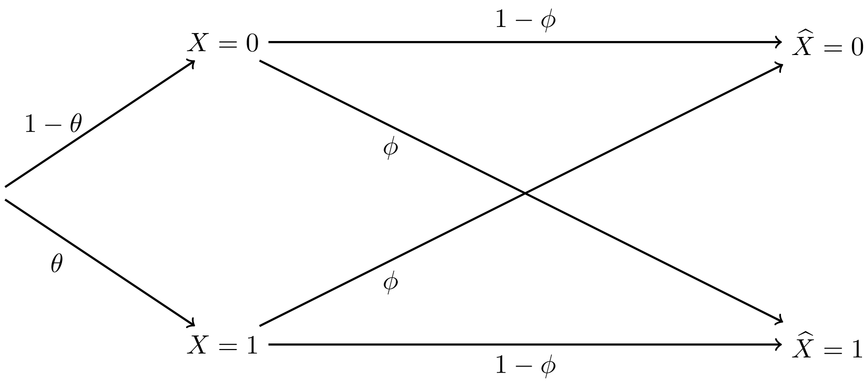
1. Express the uniform prior (θ) in terms of the beta prior.
2. Show that if the prior is chosen to be a beta prior (i.e. that $p(\theta) = \text{Beta}(\theta \mid a, b)$, and if we observe a sequence s of coin flips containing h ones and t zeros, the posterior distribution is $p(\theta \mid s) = \text{Beta}(\theta \mid a + h, b + t)$ for $a, b \geq 1$.
3. Prove that the mean of the distribution $\text{Beta}(\theta \mid a, b)$ is $\frac{a}{a+b}$.
4. Prove that the variance of the distribution $\text{Beta}(\theta \mid a, b)$ is $\frac{ab}{(a+b+1)(a+b)^2}$. (Hint: You might want to use the identity $B(a+1, b) = \frac{a}{a+b} B(a, b)$.)

Consider observing the sequence $X_{1:2N} = 01010101\ldots$ containing N zeros and N ones.

5. What is the posterior distribution after observing this sequence, given a uniform prior? What is the variance?
6. Plot the posterior for $N = 1, 10, 50$. (You might use SciPy.)
7. What's the smallest value of N, such that the probability that θ lies in the interval $[0.49, 0.51]$ is at least 0.99? (You might use SciPy.)

Exercise 4 — Noisy Coin Flips — 20 credit

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. The camera has a probability[2] 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.



[1] You may recall a similar assignment in COMP3670 about flipping coins. This is similar, but in full generality.

[2] 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.

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.

1. Show that $p(\hat{X} = 1 \mid \theta, \phi) = (1 - 2\phi)\theta + \phi$.
2. Let the prior $p(\theta \mid \phi) = \mathrm{Beta}(\theta \mid a, b)$. Prove that the posterior distribution $p(\theta \mid \phi, \hat{X} = 1)$ after having observed a 1 through the camera is given by a convex combination of $\mathrm{Beta}(\theta \mid a, b)$ and $\mathrm{Beta}(\theta \mid a + 1, b)$. That is, prove that for all $a, b \geq 1$, and all $\phi \in [0, 0.5]$, there exists $\alpha \in [0, 1]$ such that
$$p(\theta \mid \phi, \hat{X} = 1) = (1 - \alpha)\mathrm{Beta}(\theta \mid a + 1, b) + \alpha\mathrm{Beta}(\theta \mid a, b)$$
3. For the above equation, find a closed form formula for α in terms of a, b, ϕ. Your answer should contain no integrals, beta functions or gamma functions.
4. Furthermore, show that the expression for α depends only on both ϕ and the ratio b/a.