Homework 1
Bayesian Inference *

10-424/10-624 Bayesian Methods in ML
https://www.cs.cmu.edu/~hchai2/courses/10624

OUT: 01/23/25
DUE: 02/06/25

Instructions

- **Collaboration Policy**: Please read the collaboration policy in the syllabus at https://www.cs.cmu.edu/~hchai2/courses/10624/#Syllabus.
- **Late Submission Policy:** See the late submission policy in the syllabus at https://www.cs.cmu.edu/~hchai2/courses/10624/#Syllabus.
- **Submitting your work:** You will use Gradescope to submit answers to all questions.
  - **Written:** We will provide you with an Overleaf template for you to complete the written portion of this homework. You may also use the raw LaTeX source of this assignment (included in the handout .zip) to typeset your answer. **You must use LaTeX to complete this assignment**; we will not grade any submissions that are not completed using LaTeX and you will be asked to resubmit (with some penalty). You will submit your completed homework as a PDF to Gradescope.
  - **Programming:** This assignment is a written-only assignment. However, for some questions you may find that it is easier and/or beneficial to write some code to help complete your solutions; you may use any programming language and any libraries you wish to complete this assignment (provided you abide by the collaboration policy as outlined above). You are not required to submit any code you write to complete this assignment. **You are required to programmatically generate all figures**.

*Compiled on Thursday 23rd January, 2025 at 19:57

1 Bayes' rule (10 points)

Two agents from the Men in Black have contacted Henry to inform him that 10 aliens have infiltrated his 10-301/601: Introduction to Machine Learning course! Unfortunately, they find themselves under-staffed and under-resourced so the best they could do is leave Henry with an old alien scanner. The scanner is imperfect: 95% of all scanned aliens are identified as aliens and 95% of all humans are identified as human. Suppose Henry scans all 266 enrolled and waitlisted students in 10-301/601 this semester.

1.1. (5 points) **Numerical answer:** What is the probability that the first student the scanner identifies as an alien is actually an alien? **You must show your work in the provided space for full credit.**

Answer

Work

1.2. (5 points) **Numerical answer:** What is the expected number of students that the scanner identifies as an alien? **You must show your work in the provided space for full credit.**

Answer

Work

2 Gaussian Conjugacy (20 points)

An interesting property of Gaussians is *self-conjugacy*. Suppose we have a Gaussian likelihood for a (scalar) random variable $x \sim \mathcal{N}(\mu, \sigma^2)$ where the variance, σ^2, is fixed but the mean, μ, is unknown. If we put a Gaussian prior on the unknown mean, $\mu \sim \mathcal{N}(m, s^2)$, then the resulting posterior on μ is still a Gaussian!

2.1. (10 points) Suppose you gather a single observation from the Gaussian likelihood, $x^{(1)}$. Prove that the posterior distribution of μ conditioned on this observation is a Gaussian. Specifically, show that $\mu \mid x^{(1)} \sim \mathcal{N}\left(m_{post}, s^2_{post}\right)$ where

$$m_{post} = \frac{x^{(1)}s^2 + m\sigma^2}{s^2 + \sigma^2} = s^2_{post}\left(\frac{m}{s^2} + \frac{x^{(1)}}{\sigma^2}\right)$$

$$s^2_{post} = \frac{\sigma^2 s^2}{s^2 + \sigma^2} = \left(\frac{1}{s^2} + \frac{1}{\sigma^2}\right)^{-1}$$

2.2. (5 points) Now suppose you gather a sequence of observations: $x^{(1)}, x^{(2)}, \ldots, x^{(t)}$. Using the result from the previous part, compute the posterior mean and variance of μ after conditioning on all t observations.

Hint: One way to do this is to start with the prior $\mathcal{N}(m, s^2)$, compute the posterior conditioned on $x^{(1)}$, then treat that as a new prior and condition on $x^{(2)}$ to arrive at a new posterior and so on until you go through all t observations. This amounts to recursively applying the update implicitly defined in the previous part.

2.3. (5 points) Interpret your answer to the previous part as $t \to \infty$.

3 Confidence Intervals (20 points)

(Scenario quoted from Morey, et al.) A 10-meter-long research submersible with several people on board has lost contact with its surface support vessel. The submersible has a rescue hatch exactly halfway along its length, to which the support vessel will drop a rescue line. Because the rescuers only get one rescue attempt, it is crucial that when the line is dropped to the craft in the deep water that the line be as close as possible to this hatch. The researchers on the support vessel do not know where the submersible is, but they do know that it forms distinctive bubbles. These bubbles could form anywhere along the craft's length, independently, with equal probability, and float straight up to the surface where they can be seen by the support vessel. We wish to perform inference about the location of the rescue hatch given observations of bubbles; call this location ϕ.

A common "trick" when wishing to express absolute prior ignorance of a parameter is to use a so-called *uninformative* prior. In this case, we will consider the uninformative "prior" $p(\phi) = 1$. This prior does not normalize, but we will see that this fact does not lead to major problems.

3.1. (5 points) **Math:** Suppose the researchers observe the locations of exactly two bubbles, x_1 and x_2. Write down an appropriate likelihood for these observations given ϕ

3.2. (5 points) **Math:** Using the likelihood from the previous part and the uninformative prior described above, derive the posterior distribution for the location of the hatch, $p(\phi \mid x_1, x_2)$.

3.3. (10 points) **Short answer:** Using the posterior distribution from the previous part, compute a 50% Bayesian credible interval for θ given (x_1, x_2) that is centered at the midpoint between x_1 and x_2. Plot the width of this interval as a function of $|x_1 - x_2|$ and describe in words the general trend as $|x_1 - x_2|$ increases.

4 Maximum-likelihood estimation (15 points)

Suppose there are n balls in a bin labeled $1, 2, \ldots, n$ for some unknown n. You draw a single ball uniformly at random from the bin and observe that it is labeled m.

4.1. (5 points) **Math:** What is the likelihood model for this scenario, $\Pr(m \mid n)$?

4.2. (5 points) **Math:** What is the maximum likelihood estimator for n i.e., what value of n maximizes the likelihood of your observation, m?

4.3. (5 points) **Math:** What is the expected value of m given n?

5 Weird priors (15 points)

Suppose we observe $x = 90$ "successes" on $n = 100$ independent trials of some experiment or task. Call the unknown probability of success $\theta \in (0, 1)$.

For each of the prior distributions $p(\theta)$ below:

- (2 points each) Plot the prior distribution $p(\theta)$ and the posterior distribution $p(\theta \mid \mathcal{D})$ over the range $0 < \theta < 1$ on the same axes.
- (1 point each) Describe a scenario where you believe the prior would be appropriate and why.
- (2 points each) Compute the posterior mean $\mathbb{E}[\theta \mid \mathcal{D}] = \int \theta \, p(\theta \mid \mathcal{D}) \, \mathrm{d}\theta$.

5.1. (5 points) **Short answer:** $\mathcal{B}(\theta; \alpha = 1, \beta = 100)$

5.2. (5 points) **Short answer:** $\mathcal{B}(\theta; \alpha = 0.5, \beta = 0.5)$

5.3. (5 points) **Short answer:** $p(\theta) = 2$ if $\theta < 1/2$ and 0 otherwise

6 Optimal Price is Right bidding (20 points)

Suppose you have a beta belief about an unknown parameter θ with hyperparameters $\alpha = \beta = 2$. You are asked to give a point estimate $\hat{\theta}$ of θ, but are told that there is a heavy penalty for guessing too high. Specifically, the loss function is

$$\ell(\hat{\theta}, \theta; c) = \begin{cases} \theta - \hat{\theta} & \hat{\theta} < \theta; \\ c & \hat{\theta} \geq \theta \end{cases}$$

where $c \geq 1$ is a constant cost for overestimating.

6.1. (10 points) **Numerical answer:** What is the Bayesian optimal action when $c = 1$? Please show your work. **Hint:** you may need to find the roots of a polynomial; in Python, this can be done using `numpy.roots`. Don't forget that the beta distribution only has support over the domain $(0, 1)$.

6.2. (10 points) **Short answer:** Plot the optimal action as a function of c for $1 \leq c \leq 10$ and describe in words the general trend of the Bayes action as c increases. Justify why this trend is logical given the problem setting.

7 Collaboration Questions (0 points)

After you have completed all other components of this assignment, report your answers to these questions regarding the collaboration policy. Details of the policy can be found in the syllabus.

7.1. Did you collaborate with anyone on this assignment? If so, list their name or Andrew ID and which problems you worked together on.

7.2. Did you find or come across code that implements any part of this assignment? If so, include full details.