STAT 339: HOMEWORK 2B (MAXIMUM LIKELIHOOD ESTIMATION)

UPDATED: DUE VIA GITHUB BY THE START OF CLASS ON WEDNESDAY MARCH 11TH

Instructions. Create a directory called `hw2b` in your `stat339` GitHub repo. Your main writeup should be called `hw2b.pdf`.

You may also use any typesetting software to prepare your writeup, but the final document should be a PDF. LaTeXis highly encouraged.

This assignment requires a bit more coding than the last one, but still not that much (especially compared to the first two). Most of it is mathematical derivation.

I will access your work by cloning your repository; make sure that any file path information is written relative to your repo – don't use absolute paths on your machine, or the code won't run for me!

Date: Last Revised: March 1, 2020.

1. **Coding Probability Computations.** Suppose we have a random variable, Y that has a $\mathsf{Poisson}(\lambda)$ distribution. Recall from the previous homework that this means that Y takes non-negative integers as its values, and its PMF is

$$p_Y(y) = \frac{e^{-\lambda}\lambda^y}{y!}$$

   (a) Write a function that takes y and λ as inputs and returns $P(Y = y)$.

   (b) Write a function that takes k and λ as inputs and returns $P(Y \leq k)$.

   (c) Using only a random number generator that generates values uniformly distributed between 0 and 1, write a function that takes inputs λ and n and generates the values of a sequence of n **independent** random variables, each of which has a $\mathsf{Poisson}(\lambda)$ distribution. Hint: for each draw, generate a uniform random number, u, between 0 and 1, checks whether

$$u \leq F_Y(k)$$

   starting with $k = 0$, returns k if so, and if not, increments k (without changing u) and checks again.

   (d) Explain why this algorithm has probability $p_Y(k)$ of returning any particular integer k (Hint: find an expression for the lower and upper bounds on the value of u that will cause the function to return the value k)

   (e) Using the function you wrote in 1c, compute the mean of random samples of size N, starting with $N = 1$, and incrementing N. Plot the sample mean you get for each N. Try this for a few different values of λ. Use your plots to make an educated guess about how $\mathbb{E}[Y]$ depends on λ (don't look up the Poisson distribution!).

   (f) Do the same thing with the variance.

2. **Monte Carlo Integration.** Probability can be useful to get numerical solutions to calculus problems that may have nothing to do with probability. Consider the integral

$$\int_{-\infty}^{\infty} \cos(x)e^{-x^2}\, dx$$

   You will have a hard time finding an analytic solution to the above without some fairly hefty real analysis tools; however, we can get an approximate numerical solution using random variables.

(a) Find a function g and parameters μ and σ^2 such that we can write

$$\mathbb{E}\left[g(X)\right] = \int_{-\infty}^{\infty} \cos(x) e^{-x^2}\, dx$$

for a random variable X which has a $\mathcal{N}(\mu, \sigma^2)$ distribution.

(b) Simulate N independent values of X from this Normal distribution for a fairly large value of N, and use them to estimate

$$\mathbb{E}\left[g(X)\right] \approx \frac{1}{N} \sum_{n=1}^{N} g(x_n)$$

where $x_1, \ldots, x_N$ are the values you get from your simulation **You do not have to write the sampling code yourself**; you may use a pre-written function to do that part. Check that your numerical answer approaches the solution given by an integral solver such as Wolfram Alpha.

3. **Bernoulli MLE.** (Adapted from FCML Ex. 2.9): Assume that a dataset of N binary values, $x_1, ..., x_n$, was produced by sampling independently from a Bernoulli distribution with parameter $\mu := P(X = 1)$.

   (a) **Write out the likelihood function** for μ. Be clear about the domain!

   (b) **Find a formula for the maximum likelihood estimator (MLE)**, $\hat{\mu}$ (Hint: take the natural log of the likelihood first; remember that since log is a strictly increasing transformation, the value of μ that maximizes the log likelihood also maximizes the likelihood.)

4. **Univariate Normal (Gaussian) MLE.** (Adapted from FCML Ex. 2.8) Assume that a dataset of N real-valued observations consists of independent draws from a $\mathcal{N}(\mu, \sigma^2)$ distribution.

   (a) Write down the likelihood function for μ and σ^2 based on the full sample of all N observations.

   (b) Find the maximum likelihood estimates of the mean, μ, and variance, σ^2. (Hint 1: Remember that the product of exponentials is also the exponential of a sum. Hint 2: The parameters that maximize the log likelihood also maximize the likelihood. Hint 3: You will need to differentiate the log likelihood separately with respect to μ and σ^2, and set both derivatives to zero simultaneously. You may need to find the MLE of one parameter first in terms of the other and then substitute.)

5. **MLE for Linear Regression With Non-Constant Noise Variance.** Suppose we have a regression model
$$\mathbf{t} = \mathbf{X}\mathbf{w} + \boldsymbol{\varepsilon}$$
where $\varepsilon \sim \mathcal{N}(0, \mathbf{\Sigma})$, and where $\mathbf{\Sigma}$ is a known diagonal matrix with $\mathbf{\Sigma}_{nn} = \sigma_n^2$. Show that the parameter vector $\mathbf{w}$ that maximizes the (log) likelihood also minimizes the weighted least squares loss
$$\mathcal{L}(\mathbf{w}; \mathbf{x}, \mathbf{t}) = \frac{1}{N} \sum_{n=1}^{N} \alpha_n (t_n - \mathbf{w}^{\mathsf{T}} \mathbf{x}_n)^2$$
for suitable choices of the α_n (and find expressions for the α_n values in terms of the σ_n^2 values).