DS102 - Midterm Review

Monday, 14th October, 2019

1. For each of the following, answer true or false.

   (T / F) The Bayesian viewpoint favors algorithms that work on average over many possible datasets, whereas the Frequentist viewpoint considers the data as fixed.

   > **Solution:** False, the Bayesian viewpoint takes the data as fixed, while the Frequentist favors algorithms that work over many datasets.

   (T / F) The logistic function $f_\theta(x) = \frac{1}{1+e^{-\theta x}}$ is *not* a linear function of x.

   > **Solution:** True, it clearly does not satisfy linearity of scalar multiplication or aditivity in x.

   (T / F) In the LORD Procedure, the longer it has been since the last discovery the higher the amount of wealth you accrue.

   > **Solution:** False, with the LORD procedure, the longer it has been since the last discovery, the less wealth you have.

2. Similar to homework 1 you once again find yourself at the state fair. This time, you play a game that involves picking between two biased coins, C_0 and C_1, where you don't know anything about the bias of the coins. If the coin you pick lands on heads (we will denote a heads by 1 and a tails by 0 from this point) you earn \$5, otherwise you don't get anything. You decide you will play this game 10 times.

   Let $p_0 = \mathbb{P}(C_0 = 1)$ and $p_1 = \mathbb{P}(C_1 = 1)$, let $X_i \in \{0, 1\}$ indicate the number of the coin you pick on the i^{th} game, and let $Y_i \in \{0, 10\}$ be the random variable that indicates the payoff you earn on the i^{th} game where $i \in \{1, \ldots, 10\}$.

   (a) Compute $\mathbb{E}[Y_i|X_i = 0]$ and $\mathbb{E}[Y_i|X_i = 1]$ in terms of p_0 and p_1. *Reminder: you get \$5 when a coin lands heads.*

   > **Solution:**
   >
   > $$\mathbb{E}[Y_i|X_i = 0] = 5\mathbb{P}(Y_i = 5|X_i = 0) + 0\mathbb{P}(Y_i = 0|X_i = 0) = 5p_0$$
   >
   > Similarly
   >
   > $$\mathbb{E}[Y_i|X_i = 1] = 5\mathbb{P}(Y_i = 5|X_i = 1) + 0\mathbb{P}(Y_i = 0|X_i = 1) = 5p_1$$

(b) Assuming you randomly pick a coin on the i^{th} round such that each coin is equally likely $X_i \sim Bern(0.5)$. Compute $\mathbb{E}[Y_i]$ using the law of total expectation (also known as the tower property). *Reminder: the tower property states that* $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]]$.

> **Solution:**
>
> $$\begin{aligned} \mathbb{E}[Y_i] &= \mathbb{E}[\mathbb{E}[Y_i|X_i]] \\ &= \mathbb{P}(X_i = 0)\mathbb{E}[Y_i|X_i = 0] + \mathbb{P}(X_i = 1)\mathbb{E}[Y_i|X_i = 1] \\ &= 0.5(5p_0 + 5p_1) \end{aligned}$$
>
> where we have used the distribution of X_i and the result from Part a in the last equality.

(c) Now let's say that you pick coin C_0 on the first 5 rounds and coin C_1 on the last 5 rounds ($X_1 = 0, X_2 = 0, \ldots, X_5 = 0, X_6 = 1, \ldots X_{10} = 1$). Let c_i denote the value of the coin you observe on the i^{th} round (where a value of 1 indicates a heads and a value of 0 indicates a tail). Write down the log likelihood of p_0 and p_1

$$\log \mathbb{P}(c_1, c_2, \ldots, c_{10}|p_0, p_1)$$

> **Solution:**
>
> $$\begin{aligned} \log \mathbb{P}(c_1, c_2, \ldots, c_{10}|p_0, p_1) &= \sum_{i=1}^{10} \log \mathbb{P}(c_i|p_0, p_1) \\ &= \sum_{i=1}^{5} \log \mathbb{P}(c_i|p_0) + \sum_{i=6}^{10} \log \mathbb{P}(c_i|p_1) \\ &= \sum_{i=1}^{5} \log(p_0^{c_i}(1-p_0)^{(1-c_i)}) + \sum_{i=6}^{10} \log(p_1^{c_i}(1-p_1)^{(1-c_i)}) \\ &= \sum_{i=1}^{5} c_i \log(p_0) + (1-c_i)\log(1-p_0) + \\ &\quad \sum_{i=6}^{10} c_i \log(p_1) + (1-c_i)\log(1-p_1) \end{aligned}$$

(d) Compute the Maximum Likelihood Estimator (MLE) of p_0 and p_1 given the setting in Part c.

Solution: Differentiating the log likelihood wrt p_0 and setting to 0 gives us

$$\sum_{i=1}^{5} \frac{c_i}{\hat{p}_0} - \frac{1-c_i}{1-\hat{p}_0} = 0$$
$$\implies \hat{p}_0 = \frac{1}{5}\sum_{i=1}^{5} c_i.$$

Similarly we have

$$\hat{p}_1 = \frac{1}{5}\sum_{i=1}^{5} c_i.$$

(e) Instead of randomly picking coins, or deciding you'll pick a specific number of coins ahead of time, can you think of a better way to maximize your payout? You don't have to be particularly precise with your idea here.

Solution: We could pick which coin we will flip based on past results. The more heads we observe from a coin the more sure we are that we should pick that specific coin to maximize our payoff.

3. For each of the following likelihood functions show whether the Beta distribution is a conjugate prior. Recall that the Beta distribution with parameters $\alpha > 0$ and $\beta > 0$ has probability mass function:

$$f(p;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

For $0 \leq p \leq 1$, where Γ is the gamma function which does not depend on p.

(a) Geometric Distribution:
$$P(k|p) = (1-p)^{k-1}p$$

Solution: It is a conjugate prior for this likelihood.

$$P(p|k) \propto p^{\alpha}(1-p)^{\beta+k-2}$$

(b) Binomial Distribution:

$$P(p|k,n) = \binom{n}{k}(1-p)^{n-k}p^k$$

Solution: It is a conjugate prior for this likelihood.

$$P(p|k,n) \propto p^{\alpha+k-1}(1-p)^{\beta+n-k-1}$$

4. In this question we will analyze decision making with Gaussians using the Chernoff and Chebyshev bounds we have seen in lecture. Suppose you observe a sample from a Gaussian distribution. Under the null hypothesis, the sample comes from a Gaussian with mean 0 and variance 1. Under the alternative hypothesis the sample comes from a Gaussian with mean $\mu \neq 0$ and variance 1.

   Recall that the probability density function of a Gaussian distribution is given by:

   $$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

   Since you do not want to evaluate the cumulative density function of a normal, you decide to see if you can use the Chebyshev and Chernoff bounds to construct a decision rule. Suppose you collect n data points $X_1, ..., X_n$ and accept the null hypothesis if $|\bar{X}| < c$ and reject otherwise, where $\bar{X}$ is the sample mean:

   $$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

   Recall that for i.i.d samples from a Normal distribution with variance 1 the Chebyshev bound is:

   $$P(|\bar{X}-\mu| \geq c) \leq \frac{1}{nc^2}$$

   And the Chernoff bound is given by:

   $$P(|\bar{X}-\mu| \geq c) \leq e^{-\frac{nc^2}{2}}$$

   (a) Using the Chebyshev bound, what value of c allows you to control the probability of a false discovery below level α?

   Solution:

   $$P(|\bar{X}| \geq c) \leq \frac{1}{nc^2} = \alpha$$

   $$c = \frac{1}{\sqrt{n\alpha}}$$

   (b) Using the Chernoff bound, what value of c allows you to control the probability of a false discovery below level α?

Solution:

$$P(|\bar{X}| \geq c) \leq e^{-\frac{nc^2}{2}} = \alpha$$

$$c = \sqrt{\frac{2}{n}\log(1/\alpha)}$$