HW 2

Saturday, January 4, 2020 3:44 PM

1. More Probability Review

a. To prove $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ we apply definition of conditional probability.

$P(A|B) = \frac{P(A\cap B)}{P(B)}$ (defn cdtl prob)

$P(A\cap B) = P(B|A)P(A)$ (defn cdtl prob)

$\Rightarrow P(A|B) = \frac{P(B|A)P(A)}{P(B)}$.

b. $P(A,B,C) = P(C\cap B|A)$

$= P(C|B\cap A)P(B|A)P(A)$.

c. $E[X] = \sum_x X_i P(X_i)$

$(1)P(A) + (0)P(\neg A)$

$= P(A)$

$\Rightarrow E[X] = P(A)$

d

i) X is independent of Y iff for all X and all Y knowing the outcome of Y provides no additional information about the outcome of X. Formally, $(\forall i,j)\ P(X_i|Y_j) = P(X_i)$.
X and Y are not independent. When $Y=0$ we know that $P(X=0)$ is more likely than $P(X=1)$ with probability 1/3 and 3/15 respectively.

ii). X is conditionally independent of Y given Z iff given Z knowing the outcome Y provides no additional information regarding the outcome of X. Formally $(\forall i,j,k)\ P(X=x_i|Y=y_j, Z=z_k) = P(X_i|Z_k)$.
X is not conditionally independent of Y given Z

$P(X=0|Y=0\wedge Z=1) = 4/15$

$P(X=0|Z=1) = 30/45$

$\Rightarrow$ X and Y are not conditionally independent given Z.

iii) $P(X=0|X+Y>0)$

$X+Y>0$ for $X=1, Y=0$ and $X=0, Y=1$

$P(X=0|X+Y>0) = \frac{P(X=0, X+Y>0)}{P(X+Y>0)}$

$P(X=0, X+Y>0) = 1/15$

$P(X+Y>0) = P(X=1, Y=0) + P(X=0, Y=1)$.

$= 3/15 + (1/10 + 8/45)$

$= .54\bar{4}$

$1/15 / .54\bar{4}$

$\approx .1224$

2. Maximum likelihood + Maximum a Posteriori Estimates

a. Given that $X_1 \ldots X_n$ are i.i.d from a single Bernoulli distribution the likelihood function, $L(\hat{\theta})$, is $L(X_1 \ldots X_n|\theta) = f(X_1 \ldots X_n|\theta)$

$= \prod_{i=1}^{n} f(X_i|\theta)$ (from defn of independence)

$= \prod_{i=1}^{n} \theta^{X_i}(1-\theta)^{1-X_i}$ pmf for a Bernoulli R.V.

$L(\hat{\theta}) = \prod_{i=1}^{n} \theta^{X_i}(1-\theta)^{1-X}$

No, it is clear by the form of the likelihood function that the function does not depend on the order of the random variables

b. See attachment.

c. $\hat{\theta}_{mle} = \arg\max_{\theta} L(\theta)$.

$\ln\left[\prod_{i=1}^{n} \theta^{X_i}(1-\theta)^{1-X_i}\right]$

$= \sum_{i=1}^{n} X_i \ln\theta + (1-X_i)\ln(1-\theta)$

$= \frac{\partial}{\partial\theta} \ln\theta \sum_{i=1}^{n} X_i + \ln(1-\theta)\sum_{i=1}^{n}(1-X_i)$

$= \frac{\sum_{i=1}^{n} X_i}{\theta} - \frac{\sum_{i=1}^{n}(1-X_i)}{(1-\theta)}$

$= (1-\theta)\sum_{i=1}^{n} X_i = \theta\sum_{i=1}^{n}(1-X_i)$

$= \sum_{i=1}^{n} X_i - \theta X_i = \sum_{i=1}^{n} \theta - \theta X_i$

$= \sum_{i=1}^{n} X_i = \theta\sum_{i=1}^{n} 1$

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

$\hat{\theta}_{mle} = \sum_{i=1}^{n} X_i / n$, i.e the sample mean.

c Yes, according $\hat{\theta}_{mle} = 6/10 = .6$. For $\hat{\theta} = .6$ the data is most likely.

d. For each dataset the maximum likelihood estimate derived above is accurate. This is not all that surprising as the central limit theorem states that R.Vs i.i.d from a distribution will form a gaussian distribution. Thus $\hat{\theta}_{mle}$ as the sample mean is the $\hat{\theta}$ that maximizes the likelihood of the data.
For larger n's the likelihood of the data for $\hat{\theta}_{mle}$ was still relatively low compared to smaller n's.

maximum a posteriori probability estimation

f. See attachment.

g. See attachment for estimate

$\hat{\theta}_{map} = \arg\max_{\theta} P(\theta|D) = \arg\max_{\theta} \frac{P(D|\theta)P(\theta)}{P(D)} = \arg\max P(D|\theta)P(\theta)$

$P(\theta) = \theta^2(1-\theta)^2 / B(3,3) \Rightarrow$ in general the Beta Distribution is $B(\alpha,\beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$

$\arg\max_{\theta}\ \theta^{\alpha-1}(1-\theta)^{\beta-1} \cdot \prod_{i=1}^{n} \theta^{X_i}(1-\theta)^{1-X_i}$

$\frac{\alpha-1}{\theta} - \frac{\beta-1}{(1-\theta)} + \frac{\sum_{i=1}^{n} X_i}{\theta} - \frac{\sum_{i=1}^{n} 1-X_i}{1-\theta} = 0$

$\frac{\alpha-1}{\theta} + \frac{\sum_{i=1}^{n} X_i}{\theta} = \frac{\beta-1}{1-\theta} + \frac{\sum_{i=1}^{n} 1-X_i}{1-\theta}$

$(1-\theta)\left[\alpha-1+\sum_{i=1}^{n} X_i\right] = \theta\left[\beta-1+\sum_{i=1}^{n} 1-X_i\right]$

$= \alpha - 1 + \sum_{i=1}^{n} X_i - \alpha\theta + \theta - \cancel{\theta\sum_{i=1}^{n} X_i} =$

$\beta\theta - \theta + \sum_{i=1}^{n}\theta \cancel{-\theta\sum_{i=1}^{n} X_i}$

$\frac{\alpha - 1 + \sum_{i=1}^{n} X_i}{(\alpha+\beta-2+n)} = \hat{\theta}_{mle}$

If we relate this estimate to the plot we get $\frac{2+6}{4+10} = .5714$. Our estimate for the plot was .57, so yes this closed form agrees w/ the plot.

h). The map estimate takes into account a prior and thus the closed form for the map is not exactly the sample mean like the mle

i). As the # of observed data approaches infinity the map estimate takes on the form of the mle. In other words, in the case of infinite data map + mle produce the same estimate for θ.

map estimate : $\frac{\alpha - 1 + \sum_{i=1}^{n} X_i}{\alpha+\beta-2+n}$ as $n \to$ infinity (the case of infinite data)

the map estimate becomes $\frac{\sum_{i=1}^{n} X_i}{n}$ which is the closed form of the mle