Stat261 - Spring 2018
Assignment #2 - due Thursday, January 18 in class
Neatly hand write your solutions - marks will be assigned for presentation

1. Suppose that diseased trees are distributed randomly and uniformly throughout a large forest with an average of λ per acre. Let X denote the number of diseased trees in a randomly chosen one-acre plot with range, $\mathscr{X} = \{0, 1, 2, ...\}$.

   (a) What distribution can we use to model X? Write down its probability mass function (p.m.f.).

   (b) Suppose that we observe the number of diseased trees on n randomly chosen one-acre parcels, $X_1, X_2, ..., X_n$. The random variables $X_1, X_2, ..., X_n$ can be assumed to be independent. Write down the JOINT probability mass function for $X_1, X_2, ..., X_n$. Simplify this expression which is a function of λ and the $X's$.

   (c) We are going to use the Method of Maximum Likelihood to estimate λ. Write down the Likelihood function $L(\lambda)$.

   (d) Write down the Log-likelihood function $\ell(\lambda)$.

   (e) Write down the Score Function $S(\lambda)$.

   (f) Derive the maximum likelihood estimate of λ.

   (g) Write down the Information Function $I(\lambda)$.

   (h) Use the second derivative test to show that you have found a maximum.

   (i) Suppose that the numbers of diseased trees observed in eight randomly chosen one-acre parcels were: 5, 8, 9, 2, 10, 7, 6, 10. Compute the maximum likelihood estimate of λ using this data.

   (j) Suppose that the unit of measure was a five-acre plot, i.e. we found the same number of diseased trees in eight randomly chosen five-acre plots, but λ is still the mean number per one acre. What is the maximum likelihood estimate of λ now?

2. According to genetic theory, there are three blood types MM, NM and NN which should occur in a very large population with probabilities, θ^2, $2\theta(1-\theta)$ and $(1-\theta)^2$, where θ is the (unknown) gene frequency, $0 \leq \theta \leq 1$.

   (a) Suppose that in a random sample of size $n = 10$ from the population, there are f_1, f_2, and f_3 of types MM, NM and NN respectively. What distribution can we assume for the frequencies, f_1, f_2, and f_3? Write down its probability mass function.

Blood Type	MM	NM	NN	Total
Observed frequency	f_1	f_2	f_3	$n = 100$
Probability	$p_1 = \theta^2$	$p_2 = 2\theta(1-\theta)$	$p_3 = (1-\theta)^2$	1

   (b) What are the assumptions required for the distribution you assume in part (a)?

(c) Write down the Log-likelihood function as a function of θ, $\ell(\theta)$. Write down the Score function, $S(\theta)$.

(d) Find an expression for the maximum likelihood estimate of θ.

(e) Suppose that $f_1 = 3$, $f_2 = 4$, and $f_3 = 3$. Compute the maximum likelihood estimate of θ.

(f) Compute the estimated probabilities for p_1, p_2 and p_3 under this model. i.e. $p_i(\hat{\theta})$, $i = 1, 2, 3$.

(g) Compute the estimated expected frequencies under this model $np_i(\hat{\theta})$ and compare them with the observed frequencies.