35.

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, $\mathscr{X} = \{0, 1, 2, ...\}$.

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

   2 (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 $X's$.

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

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

17 2 (e) Write down the Score Function $S(\lambda)$.

   2 (f) Derive the maximum likelihood estimate of λ.

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

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

   2 (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.

   3 (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, $0 \leq \theta \leq 1$.

   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

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

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

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

16. 2 (e) Suppose that $f_1 = 3$, $f_2 = 4$, and $f_3 = 3$. Compute the maximum likelihood estimate of θ.

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

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

33 + a presentation.

Stat 261 – Spring 2018 – Assign # 2

1. (a) $X \sim \text{Poisson}(\lambda)$ since diseased trees occur randomly over space (one acre)

$f(x;\lambda) = \dfrac{\lambda^x e^{-\lambda}}{x!}$ $x = 0, 1, \ldots$ $\lambda > 0$, λ = mean # diseased trees in 1 acre.

(b) $X_1, X_2, \ldots, X_n$ iid Poisson (λ) $x_i = 0, 1, 2, \ldots$

$$f(x_1, x_2, \ldots, x_n; \lambda) = f(x_1;\lambda) \times f(x_2;\lambda) \times \cdots \times f(x_n;\lambda)$$
$$= \prod_{i=1}^{n} f(x_i;\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$
$$= \frac{\lambda^{\sum_1^n x_i} e^{-n\lambda}}{x_1!\, x_2! \cdots x_n!} \longleftarrow \text{not depend on } \lambda$$

(c) $L(\lambda) = \lambda^{\sum_1^n x_i} e^{-n\lambda}$.

(d) $\ell(\lambda) = \left(\sum_1^n x_i\right) \ln \lambda - n\lambda$

(e) $S(\lambda) = \ell'(\lambda) = \dfrac{\sum_1^n x_i}{\lambda} - n$

(f) $S(\hat\lambda) = 0 = \dfrac{\sum_1^n x_i}{\lambda} - n \Rightarrow \hat\lambda = \dfrac{\sum_1^n x_i}{n}$ Sample mean.

(g) $I(\lambda) = -\dfrac{d^2 \ell}{d\lambda^2} = \dfrac{\sum_1^n x_i}{\lambda^2} > 0$. when not all $x_i = 0$

(h) $I(\hat\lambda) > 0$, therefore we have a maximum.

(i) $\sum_1^n x_i = 57$ $\therefore$ $\hat\lambda = 57/8 = 7.128$

(j) $X_i \sim \text{Poisson}(5\lambda)$.

$$f(x_1, x_2, \ldots, x_n; \lambda) = \prod_{i=1}^{n} \frac{(5\lambda)^{x_i} e^{-5\lambda}}{x_i!} = \frac{\lambda^{\sum_1^n x_i} e^{-5n\lambda}}{x_1! \cdots x_n!} \times \left(5^{\sum x_i}\right).$$

$L(\lambda) = \lambda^{\sum x_i} e^{-5n\lambda}$

$\ell(\lambda) = \left(\sum_1^n x_i\right) \ln\lambda - 5n\lambda$

$$S(\lambda) = \frac{\sum_{1}^{n} x_i}{\lambda} - 5n \quad \Rightarrow \quad \hat{\lambda} = \frac{\sum_{1}^{n} x_i}{5n}$$

$$\therefore \hat{\lambda} = \frac{57}{40} = 1.425.$$

2/ (a) $$P(f_1, f_2, f_3; \theta) = \underbrace{\binom{10}{f_1 f_2 f_3}}_{\text{constant}} p_1(\theta)^{f_1} \, p_2(\theta)^{f_2} \, p_3(\theta)^{f_3}$$

$$(f_1 \, f_2 \, f_3) \sim \text{Multinomial}\,(n = 10, p_1(\theta), p_2(\theta), p_3(\theta))$$

(b) - n subjects are independent
- randomly chosen subjects have constant prob. of blood types mm, Nm, NN.

(c) $$L(\theta) = p_1(\theta)^{f_1} \, p_2(\theta)^{f_2} \, p_3(\theta)^{f_3}.$$
$$= [\theta^2]^{f_1} \, [2\theta(1-\theta)]^{f_2} \, [(1-\theta)^2]^{f_3}.$$
$$= c \, \theta^{2f_1 + f_2} (1-\theta)^{f_2 + 2f_3} \qquad c = 2^{f_2}$$

$$l(\theta) = (2f_1 + f_2) \ln \theta + (f_2 + 2f_3) \ln(1-\theta)$$
$$S(\theta) = \frac{2f_1 + f_2}{\theta} - \frac{(f_2 + 2f_3)}{1-\theta}.$$

(d) $S(\hat{\theta}) = 0$ $\quad \hat{\theta} = \dfrac{2f_1 + f_2}{2n}$ $\quad n = f_1 + f_2 + f_3.$

(e) $$\hat{\theta} = \frac{2 \times 3 + 4}{20} = \frac{10}{20} = .5$$

(f) (g)

	mm	NM	NN
obs freq.	3	4.	3.
$p(\hat{\theta})$	.25	.5.	.25
Est exp freq.	2.5	5.	2.5.

The observed are reasonably close to the estimated expected frequencies.