Exercises Sheet 4

Probability Distributions and Applications

1. The Table shows the distribution of *X*=No. Bacteria/square obtained for counts of the root nodule bacteria (RT) in a Petroff-Hauser counting chamber. Complete the table for Expected number, assuming a Poisson. Comment on the observed distributional form.

No. Bacteria /Square (*X*)	**Number of Squares Observed**	**Expected**
0	34	?
1	68	?
2	112	?
3	94	?
4	55	?
5	21	?
6	12	?
7+	4	?
	400	

2. The hospital period, in days, for patients following treatment for a certain type of kidney disorder is a random variable *Y=X+4*, where X has the density function

$$f(x) = \begin{cases} \dfrac{32}{(x+4)^3}, & x > 0 \\ 0, & elsewhere \end{cases}$$

(i) Obtain p.d.f. of *Y*
(ii) Using the density function of *Y*, find the probability that the hospital period for a patient following the treatment will exceed 8 days.

3. Assume that the average number of crossovers is *m* in a genome segment, flanked by loci *A* and *B* with crossovers treated as Poisson events. Given that recombinant classes are observed only when an odd number of crossovers occur in the interval, obtain the expected recombination fraction, (defined as the probability of recombinant genotypes in the progeny) in terms of the expected number of crossovers (map distance).

For ℓ loci on a genome segment, and with r_ℓ, r_i, the recombinant fractions between two genes or genetic markers flanking the whole segment and between two markers flanking a sub-segment respectively, obtain the form of Haldane's mapping function, by arguing from the simple cases.

4. Suppose ages at onset of a given disease distribute Normally, mean = 11.5 and variance =9. For a child, who has just contracted the disease, what is the probability he/she is
(i) between ages 8.5 and 145
(ii) over 10
(iii) under 12

5. Find the distribution function *F(x,* median and mode for the following p.d.f.'s. [Note: range of *x* for part iii]

$i.\ \ f(x) = \dfrac{1}{2\sqrt{x}}, \quad 0 \le x \le 1$ $\qquad\qquad iii.\ \ f(x) = 1 - |1 - x|, \quad 0 \le x \le 2$

ii. $f(x)=\frac{1}{4}\left(\frac{3}{4}\right)^{x-1}, \quad x=1,2,...$ *iv.* $f(x)=6x(1-x), \quad 0\le x\le 1$

6. If X has the rectangular distribution over (0,1),

$$f(x)=\begin{cases}\frac{1}{b-a}, & a\le x\le b\\ 0 & otherwise\end{cases}$$

Obtain the m.g.f. of

$$Y=-\ell n(X)$$

Identify the distribution of Y.

7. Given the prior distribution for the proportion p of people with a given condition is:

p	0.1	0.2
f(p)	0.6	0.4

Find the Bayes estimate for the proportion of people with the condition, if a random sample of size 2 gives 1 with the disease.

[Hint: Clearly X = No. with the disease, ~ Binomial for given p and basic probability rules mean that $f(x,p)=f(x/p)\,f(p)$]

8. For the following data set, the counts in the four categories can be treated as multinomial variables, with parameters c_1, c_2, c_3, c_4, with pdf

$$f(c_1,c_2,c_3,c_4)=\frac{n!}{f_1!,f_2!,f_3!,f_4!}c_1^{f_1}c_2^{f_2}c_3^{f_3}c_4^{f_4}$$

where $c_1+c_2+c_3+c_4=1$ and $f_1+f_2+f_3+f_4=n$

Complete the table, **give** the mgf, the first moments for the four variables and the moment estimates for the four multinomial proportions. By setting the moment estimates equal to their expectation, *p(P)* in the table, and from the constraint on the *c's*, **obtain** moment estimates of the two parameters.

Table of Expected progeny frequencies and observed counts for a marker and a fusiform rust resistance gene (F) in an experiment, assuming that the frequency of the virulence gene that overcame the resistance of the host gene F in the pathogen is t. Purpose of the experiment is to estimate the recombination fraction between the marker and resistance gene F with the frequency of the virulence gene in the pathogen in the model.

Genetic model for host	Pathogen	
__| ____θ_______F(resistance gene)	Gene Virulent (a)	Freq. t
Marker	Avirulent (A)	Freq $1-t$

Expected genotypic freq.

+F	*0.5(1-*θ*)*	*F - a* = susceptible (*r*)
+f	*0.5*θ	*F - A* = resistance (*R*)
-F	*0.5*θ	*f – a* = susceptible (*r*)
-f	*0.5(1-*θ*)*	*f – A* = susceptible (*r*)

G	**p(G)**	**+R**	**p(P/G)** **+r**	**-R**	**-r**
+F	*0.5(1-*θ*)*	1 – t	t	0	0
+f	*0.5*θ	0	1	0	0
-F	*0.5*θ	0	0	1 - t	t
-f	*0.5(1-*θ*)*	0	0	0	1

p(P)	$0.5(1-\theta)(1-t)$	$0.5\theta(t - t\theta + \theta)$	$0.5(1-t)\theta$	$0.5\theta(1+t)$
p(R/P)	0	?	1	?
Count	9	12	2	30