Astronomy A345H

Astronomical Data Analysis I: Example Sheet 2

1. Let x and y be random variables. Using the definitions of variance and covariance given in the lectures find expressions for

   (a) $\mathrm{var}(ax)$, where a is a constant

   (b) $\mathrm{cov}(ax, by)$

   (c) $\mathrm{cov}(x, x + y)$

   (d) $\mathrm{cov}(x + y, x - y)$

   Given the definition of the correlation coefficient, ρ, in equation (3.3) of your notes, what useful property of the correlation coefficient is suggested by your answers to parts (a) and (b)?

2. In a quasar survey covering an area of 1000 square degrees on the sky, the total number of quasars found is 23,400. Assuming that the projected distribution of quasars can be modelled as a Poisson distribution, what is the probability of observing less than 5 quasars ina given square degree of the sky? What area of sky could one expect to survey before the probability of finding *no* quasars was less than 1%?

3. (a) Explain the circumstances in which the Poisson distribution

   $$p(r) = \frac{\mu^r e^{-\mu}}{r!}$$

   correctly describes the probability of r events (e.g. arrival of photons) occurring, and identify the meaning of the parameter μ in this equation.

   (b) There is a constant probability that a fire will occur at any time in Glasgow, and there are two fires per day on average. Write down the probability , $p(n)$, of n fires occurring on any one day.

   (c) To deal with any fire requires the presence of a fire engine for a whole day. What is the minimum number of fire engines required in Glasgow so that the probability that all fires on a given day are attended to is better than 99%?

4. Let p denote the probability that a particular outcome will happen in any single observation (we call this the probability of a *success*). The probability, $p(r)$, of exactly r successes in n observations is given by the *binomial* distribution:

   $$p(r) = \frac{n!}{r!(n-r)!} p^r (1-p)^{n-r}, \qquad r = 0, 1, ..., n$$

   It can be shown that $p(r)$ has mean $\mu = np$ and variance $\sigma^2 = np(1-p)$.

   (a) Suppose it is given that 60% of the stars in the Hubble Space Telescope guide star catalogue are binaries, use a binomial distribution model to calculate the probability that a random sample of 5 stars from the guide star catalogue contains (i) 0, (ii) 1, (iii) 2, (iv) 3, (v) 4, (vi) 5 binary stars.

(b) How large a sample should be chosen in order that the probability of the sample containing at least two *non*–binary stars is greater than 99%?

5. In a meteor search program 4 CCD images were taken on each observing night and examined for meteor trails. Over a one year period, 150 nights of data were accumulated with the following results.

No of images with trails	0	1	2	3	4
No of nights	30	62	46	10	2

The number, r, of images recording meteor trails on any given night is assumed to follow a binomial distribution. By equating the sample mean value of r for the above observations with the expected value for a binomial distribution, estimate the parameter, p, the probability of a single CCD image recording a meteor trail. Hence determine the *predicted* number of nights on which r CCD images record trails under the binomial model ($r = 1, ..., 5$).

6. Repeat Q.5 but this time fit the data to a Poisson distribution:-

$$p(r) = \frac{\lambda^r e^{-\lambda}}{r!}$$

recalling that $E(r) = \lambda$ for a Poisson variable, r. Hence determine the predicted number of nights on which r images containing trails are observed, under the Poisson model. Which model – Poisson or binomial – appears to give a better fit to the data? Why might a Poisson model not be appropriate?

7. (a) By comparing the expected and observed number of nights on which r images containing trails are observed under the binomial and Poisson model respectively, construct a χ^2 goodness of fit statistic for the data in Q.5 and Q.6.

   (b) Explain carefully why the appropriate number of degrees of freedom, for assessing the significance of the values of your χ^2 statistic, is 3.

   (c) Use the *online Chi-square calculator* at

   `http://statpages.org/pdfs.html`

   to compute $P-$values for the goodness of fit to the Poisson and binomial models.

   (d) Comment on your answer, in relation to your answer to Q.6.

8. A list of 1000 (supposedly) uniformly random digits, i.e. $0 - 9$, are generated by a pseudo-random number generator. Suppose the integers appear in the list with the following frequencies:-

r	0	1	2	3	4	5	6	7	8	9
o_r	92	109	108	99	85	112	91	101	114	89

Form the appropriate χ^2 statistic and use the online Chi-Square calculator to compute a $P-$value for the these data, under the assumption that they do indeed represent uniformly random digits.