Astronomical Data Analysis I
Examples Sheet 2, Model Answers

Q.1 a) $\mathrm{var}(ax) = E\left([ax - E(ax)]^2\right)$

$$E(ax) = \int ax\, p(x)\,dx = a\,E(x)$$

So $\mathrm{var}(ax) = E\left(\left[a\,(x - E(x))\right]^2\right) = a^2\,\mathrm{var}(x)$

b) $\mathrm{cov}(ax, by) = E\left[(ax - a\bar{x})(by - b\bar{y})\right]$

$$= E\left[ab\,(x-\bar{x})(y-\bar{y})\right]$$

$$= ab\,\mathrm{cov}(x, y)$$

c) We can use the general result $\mathrm{cov}(x, y) = E\left[(x-\bar{x})(y-\bar{y})\right]$
$= E(xy) - E(x)E(y)$

Thus $\mathrm{cov}(x, x+y) = E\left(x(x+y)\right) - E(x)E(x+y)$

$$= E(x^2) + E(xy) - E(x)^2 - E(x)E(y)$$

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

d) $\mathrm{cov}(x+y, x-y) = E\left((x+y)(x-y)\right) - E(x+y)\,E(x-y)$

$$= E(x^2 - y^2) - E(x)^2 + E(y)^2$$

$$= E(x^2) - E(x)^2 - \left[E(y^2) - E(y)^2\right]$$

$$= \mathrm{var}(x) - \mathrm{var}(y)$$

(a) + (b) $\Rightarrow$ correlation coefficient is independent of scale.

$$\rho[ax, by] = \frac{\mathrm{cov}(ax, by)}{\sqrt{\mathrm{var}(ax)}\sqrt{\mathrm{var}(by)}} = \frac{ab\,\mathrm{cov}(x,y)}{a\sqrt{\mathrm{var}(x)}\;b\sqrt{\mathrm{var}(y)}} = \rho(x,y)$$

Q.2 Mean number of quasars per square degree $= \dfrac{2340}{1000} = 2.34$

$$p(<5) = p(0) + p(1) + \ldots\; p(4)$$

$$p(r) = \frac{\mu^r e^{-\mu}}{r!} \Rightarrow p(<5) = e^{-2\cdot 34}\left[1 + 2\cdot 34 + \frac{2\cdot 34^2}{2} + \frac{2\cdot 34^3}{6} + \frac{2\cdot 34^4}{24}\right]$$

$$= 91\cdot 1\%$$

We require μ = mean number of quasars per unit area, A, such that

$$p(0) = e^{-\mu} = 0\cdot 01 \quad \text{in area } A$$

i.e.

$$-\mu = \ln 0\cdot 01$$

$$\mu = 4\cdot 605$$

Since the observed rate is $2\cdot 34$ per square degree $\Rightarrow A = \frac{4\cdot 605}{2\cdot 34}$

$= 1\cdot 97$ sq. deg.

Q.3

(a) The Poisson distribution describes the probability of n events occurring, given a (constant) mean occurrence rate, λ, and given that the probability of two or more events occurring simultaneously is zero.

(b) Here the mean rate is 2 fires per day, so the probability of n fires occuring ona particular day is

$$P(n) = e^{-2}\frac{2^n}{n!}.$$

We can therefore write:

$$P(0) = e^{-2} = 0.135$$

$$P(1) = 2e^{-2} = 0.271, \quad \text{so} \quad P(\leq 1) = 0.406$$

$$P(2) = 2e^{-2} = 0.271, \quad \text{so} \quad P(\leq 2) = 0.677$$

$$P(3) = \frac{4}{3}e^{-2} = 0.180, \quad \text{so} \quad P(\leq 3) = 0.857$$

$$P(4) = \frac{2}{3}e^{-2} = 0.090, \quad \text{so} \quad P(\leq 4) = 0.947.$$

There is a 94.7 percent probability of there being 4 or fewer fires per day, so *4* engines are sufficient to attend all the fires of Glasgow $>$ 90 percent of the time.

Q.4

$\text{Prob}(\geqslant 2 \text{ non-binaries}) > 99\% \iff \text{Prob}(n \text{ binaries}) + \text{Prob}(n\text{-}1 \text{ binaries}) < 0.01$

$\iff (0.6)^n + n(0.6)^{n-1}(0.4) < 0.01$

Let $f(n) = 0.6^{n-1}\left[0.6 + 0.4n\right]$

calculation shows that $f(13) = 0.0126$
$f(14) = 0.008$

$\Rightarrow$ we need at least 14 stars in our sample

Q.5 Sample mean, $\hat{r} = \dfrac{0\times 30 + 1\times 62 + 2\times 46 + 3\times 10 + 4\times 2}{150}$

$= 1.28$

Setting $E(r) = \hat{r} = np = 1.28 \Rightarrow \hat{p} = 0.32$ (since $n=4$)

predicted no. of nights with r plates
$= 150 \cdot p(r;\hat{p})$

$\Rightarrow N(0) = 32 \quad N(1) = 60 \quad N(2) = 43 \quad N(3) = 13 \quad N(4) = 2$
(all to nearest integer)

Q.6 Set $E(r) = \hat{r} = \lambda$ i.e. $\lambda = 1.28$ $\qquad p(r;\lambda) = \dfrac{\lambda^r e^{-\lambda}}{r!}$

Again $N(r) = 150\, p(r;\lambda)$

$\Rightarrow N(0) = 42 \quad N(1) = 53 \quad N(2) = 34 \quad N(3) = 15 \quad N(4) = 5$

The binomial model fits the data better.

For a binomial we have $r \leq 4$ (which is true!). For a Poisson model, $r \in \mathbb{N}$

Q.7 We form $\chi^2 = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i}$

(a) Under binomial model, $\chi^2 = \frac{2^2}{32} + \frac{2^2}{60} + \frac{3^2}{43} + \frac{3^2}{13} + \frac{0^2}{2}$

$= 1 \cdot 09$

Under Poisson model, $\chi^2 = \frac{12^2}{42} + \frac{9^2}{53} + \frac{12^2}{34} + \frac{5^2}{15} + \frac{3^2}{5} = 12 \cdot 65$

(b) No. of dof = 3 because there are 5 bins, but we subtract 1 dof because we know $\sum o_i = 150$ and another d.of. because we replace p and μ respectively by their sample estimate.

(c) Binomial – P-value = $0 \cdot 7795$
Poisson – P-value = $0 \cdot 0055$

(d) This confirms that the binomial model is a much better fit, and the Poisson model is a poor fit

Q.8 We form $\chi^2 = \sum \frac{(o_i - e_i)^2}{e_i}$ where $e_i = 100$ for uniformly random digits

$\Rightarrow \chi^2 = \frac{1}{100}\left(8^2 + 9^2 + 8^2 + 1^2 + 15^2 + 12^2 + 9^2 + 1^2 + 14^2 + 11^2\right)$

$= 9 \cdot 78$

We have 9 dof (since $\sum o_i = 1000$), so $p = 0 \cdot 3686$
This suggests the digits are consistent with being uniformly random.