Semester 1 2012 Lecturer: Michael Stewart

Tutorial Week 9

1. Suppose an urn contains w white and b black balls (each uniquely numbered, say) and n draws are made *with replacement* such that each possible sequence is equally likely.
   (a) How many possible sequences are there?
   (b) For how many of these are *all* n balls black?
   (c) For how many sequences is the first ball white?
   (d) For each $x = 1, 2, \ldots, n-1$, how many sequences have
      - the first x balls all black *and*
      - the $(x+1)$-th ball white?

   (e) Hence, letting X denote how many black balls are seen *before* the first white ball, write down $P(X = x)$, for $x = 0, 1, 2, \ldots, n$ in terms of $\theta = w/(w+b)$.
   (f) Determine $q_x = \lim_{n\to\infty} P(X = x)$ and moreover verify that $\sum_{x=0}^{\infty} q_x = 1$.

2. Suppose X_1 and X_2 are independent random variables each with $P(X_i = x) = q_x$ (where q_x is defined in the question above). That is, suppose that the underlying sample space is the sequence of pairs:

$$\{(0,0), (0,1), (1,0), (0,2), (1,1), (2,0), \ldots\},$$

   and that each outcome (x_1, x_2) has non-negative weight $q_{x_1} q_{x_2}$.
   (a) Show that the (limiting) sum of all the weights thus defined is 1.
   (b) Determine $P(S = s)$, where $S = X_1 + X_2$, for each $s = 0, 1, 2, \ldots$

3. Suppose $y_1, y_2, \ldots, y_n$ $(= \mathbf{y})$ are modelled as values taken by independent and identically distributed random variables (IID RVs) $Y_1, Y_2, \ldots, Y_n$ with $P(Y_1 = y) = \theta(1-\theta)^{y-1}$, $y = 1, 2, 3, \ldots$. Derive the maxmimum likelihood estimate of θ (**hint:** the likelihood function is the probability of the data:

$$L(\theta; \mathbf{y}) = P(Y_1 = y_y, \ldots, Y_n = y_n) = P(Y_1 = y_1) \cdots P(Y_n = y_n),$$

   and the estimate maximises this function of θ, or equivalently maximises $\log L(\theta, \mathbf{y})$).

4. Suppose that the 100 counts summarised in the frequency table below are modelled as being values taken by IID RVs $Y_1, Y_2, \ldots, Y_{100}$ with $P(Y_1 = y) = \theta(1-\theta)^{y-1}$ (as in the previous question):

Count	1	2	3
Frequency	94	5	1

   (a) Compute the maximum likelihood estimate of θ.
   (b) Compute expected frequencies under the maximum likelihood fit.
   (c) Obtain a standard error for your estimate (remember $Var(Y_1) = (1-\theta)/\theta^2$).
   (d) Obtain standardised residuals and comment on the goodness (or badness) of fit.

More questions over the page…

5. Suppose X is a random variable with a Poisson(θ) distribution, that is

$$P(X = x) = \frac{e^{-\theta}\theta^x}{x!}\,, \quad \text{for } x = 0, 1, 2, \ldots$$

Show that for each $k = 1, 2, 3, \ldots$,

$$E[X(X-1)\cdots(X-k+1)] = \theta^k\,.$$

One way to do this is to:

- write out the sum defining that expectation;
- show that the first k terms are zero;
- change variable inside the sum to $y = x - k$;
- take a factor outside the sum leaving a sum that adds to 1.

6. Suppose data are modelled as values taken by IID RVs $X_1, X_2, \ldots, X_n$ ($= \mathbf{X}$) with $X_1 \sim \mathit{Pois}(\theta)$. According to the previous question, $E(X_1) = \theta$ while $P(X_1 = 0) = e^{-\theta}$. These suggest two different ways to estimate θ. Write $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, let

$$I_i = \begin{cases} 1 & \text{if } X_i = 0, \\ 0 & \text{otherwise} \end{cases}$$

and write $\bar{I} = \frac{1}{n}\sum_{i=1}^{n} I_i$.

- Since $E(\bar{X}) = E(X_1) = \theta$, just use $\hat{\theta}_1(\mathbf{X}) = \bar{X}$;
- Since $E(\bar{I}) = E(I_1) = P(I_1 = 1) = e^{-\theta}$, set $\bar{I} = e^{-\theta}$ and solve, giving $\hat{\theta}_0(\mathbf{X}) = -\log(\bar{I})$.

(a) Write down $\mathrm{MSE}_\theta(\hat{\theta}_1(\mathbf{X})) = E_\theta[(\bar{X} - \theta)^2]$.

(b) Write down $\mathit{Var}(\bar{I})$.

(c) Derive a large-sample approximation to $\mathrm{MSE}_\theta(\hat{\theta}_0(\mathbf{X}))$ (**hint:** write $\hat{\theta}_0(\mathbf{X}) - \theta = g(\bar{I}) - g(\mu_I)$, where $\mu_I = E(\bar{I})$, for some $g(\cdot)$).

(d) Based on your answers to (a) and (c), which estimator would you prefer to use?

7. Suppose $x_1, x_2, \ldots, x_n$ are modelled as values taken by n independent $\mathit{Pois}(\theta)$ random variables. Derive, as a function of the x_i's, the maximum likelihood estimate for θ.

8. The table below summarises 100 counts:

Count	0	1	2
Frequency	90	9	1

Suppose we model them as values taken by 100 $\mathit{Pois}(\theta)$ RVs, for some unknown $\theta > 0$.

(a) Obtain the maximum likelihood estimate of θ.

(b) Obtain a standard error for this estimate.

(c) Compute standardised residuals and comment on the goodness of fit.

9. [*Harder*] Generalise question 2(b) to the case where $S = X_1 + X_2 + \cdots + X_k$, and $X_1, X_2, \ldots, X_k$ are IID RVs with $P(X_1 = x) = q_x$, $x = 0, 1, 2, \ldots$.