smfsoln1.tex

SMF SOLUTIONS 1. 28.1.2011

Q1. To fit a straight line

$$y = a + bx$$

by least squares through a data set (x_1, y_1), ..., (x_n, y_n), we choose a, b so as to minimise

$$SS := \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - a - bx_i)^2.$$

Taking $\partial SS/\partial a = 0$ and $\partial SS/\partial b = 0$ gives

$$\begin{aligned} \partial SS/\partial a := -2\sum_{i=1}^n e_i &= -2\sum_{i=1}^n (y_i - a - bx_i), \\ \partial SS/\partial b := -2\sum_{i=1}^n x_i e_i &= -2\sum_{i=1}^n x_i(y_i - a - bx_i). \end{aligned}$$

To find the minimum, we equate both these to zero:

$$\sum_{i=1}^n (y_i - a - bx_i) = 0, \quad \text{and} \quad \sum_{i=1}^n x_i(y_i - a - bx_i) = 0.$$

This gives two simultaneous linear equations in the two unknowns a, b, called the *normal equations*. Using the notation

$$\bar{x} := \frac{1}{n}\sum_{i=1}^n x_i,$$

dividing both sides by n and rearranging, the normal equations are

$$a + b\bar{x} = \bar{y}, \quad \text{and} \quad a\bar{x} + b\overline{x^2} = \overline{xy}.$$

Multiply the first by $\bar{x}$ and subtract from the second:

$$b = (\overline{xy} - \bar{x}\bar{y})/(\overline{x^2} - (\bar{x})^2), \quad \text{and then} \quad a = \bar{y} - b\bar{x}.$$

We will use this bar notation systematically. We call $\bar{x} := \frac{1}{n}\sum_{i=1}^n x_i$ the *sample mean*, or average, of $x_1, \ldots, x_n$, and similarly for $\bar{y}$. In this book (though not all others!), the *sample variance* is defined as the average, $\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$, of $(x_i - \bar{x})^2$, written s_x^2 or s_{xx}. Then using linearity of average, or 'bar',

$$s_x^2 = s_{xx} = \overline{(x - \bar{x})^2} = \overline{x^2 - 2x.\bar{x} + \bar{x}^2} = \overline{(x^2)} - 2\bar{x}.\bar{x} + (\bar{x})^2 = \overline{(x^2)} - (\bar{x})^2.$$

Similarly, the *sample covariance* of x and y is defined as the average of $(x-\bar{x})(y-\bar{y})$, written s_{xy}. So

$$s_{xy} = \overline{(x-\bar{x})(y-\bar{y})} = \overline{xy - x.\bar{y} - \bar{x}.y + \bar{x}.\bar{y}} = \overline{(xy)} - \bar{x}.\bar{y} - \bar{x}.\bar{y} + \bar{x}.\bar{y} = \overline{(xy)} - \bar{x}.\bar{y}.$$

Thus the slope b is given by $b = s_{xy}/s_{xx}$, the ratio of the sample covariance to the sample x-variance.

Q2. With two regressors u and v and response variable y, given a sample of size n of points $(Uu_1, v_1, y_1), \ldots, (u_n, v_n, y_n)$ we have to fit a least-squares *plane* – that is, choose parameters a, b, c to minimise the sum of squares

$$SS := \sum_{i=1}^{n} (y_i - c - au_i - bv_i)^2.$$

Taking $\partial SS/\partial c = 0$ gives

$$\sum_{i=1}^{n} (y_i - c - au_i - bv_i) = 0: \qquad c = \bar{y} - a\bar{u} - b\bar{v}.$$

We re-write SS as

$$SS = \sum_{i=1}^{n} [(y_i - \bar{y}) - a(u_i - \bar{u}) - b(v_i - \bar{v})]^2.$$

Then $\partial SS/\partial a = 0$ and $\partial SS/\partial b = 0$ give

$$\sum_{i=1}^{n} (u_i - \bar{u})[(y_i - \bar{y}) - a(u_i - \bar{u}) - b(v_i - \bar{v})],$$

$$\sum_{i=1}^{n} (v_i - \bar{v})[(y_i - \bar{y}) - a(u_i - \bar{u}) - b(v_i - \bar{v})].$$

Multiply out, divide by n to turn the sums into averages, and re-arrange using our earlier notation: these become

$$as_{uu} + bs_{uv} = s_{yu},$$

$$as_{uv} + bs_{vv} = s_{yv}.$$

These are the *normal equations* for a and b. The determinant is

$$s_{uu}s_{vv} - s_{uv}^2 = s_{uu}s_{vv}(1 - r_{uv}^2)$$

(as $r_{uv} := s_{uv}/(s_u.s_v)$), $\neq 0$ iff $r_{uv} \neq \pm 1$, i.e., iff the (u_i, v_i) are not collinear, and this is the condition for the normal equations to have a unique solution.

Q3. (i) For $n = 1$, the mean is 1, because a $\chi^2(1)$ is the square of a standard normal, and a standard normal has mean 0 and variance 1. The variance is 2, because the fourth moment of a standard normal X is 3, and

$$var(X^2) = E[(X^2)^2] - [E(X^2)]^2 = 3 - 1 = 2.$$

For general n, the mean is n because means add, and the variance is $2n$ because variances add over independent summands.
(ii) For X standard normal, the MGF of its square X^2 is

$$M(t) := \int e^{tx^2}\phi(x)dx = \frac{1}{\sqrt{2\pi}}\int_0^\infty e^{tx^2}.e^{-\frac{1}{2}x^2}dx.$$

We see that the integral converges only for $t < \frac{1}{2}$, when it is $1/\sqrt{(1-2t)}$:

$$M(t) = 1/\sqrt{1-2t} \qquad (t < \frac{1}{2}) \qquad \text{for } X\ N(0,1).$$

Now when X, Y are independent, the MGF of their sum is the product of their MGFs. For, e^{tX}, e^{tY} are independent, and the mean of an independent product is the product of the means. Combining these, the MGF of a $\chi^2(n)$ is given by

$$M(t) = 1/(1-2t)^{\frac{1}{2}n} \qquad (t < \frac{1}{2}) \qquad \text{for } X\ \chi^2(n).$$

(iii) First, $f(.)$ is a density, as it is non-negative, and integrates to 1:

$$\begin{aligned}
\int f(x)dx &= \frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}.\int_0^\infty e^{tx}.x^{\frac{1}{2}n-1}\exp(-\frac{1}{2}x)dx \\
&= \frac{1}{\Gamma(\frac{1}{2}n)}.\int_0^\infty u^{\frac{1}{2}n-1}\exp(-u)du \qquad (u := \frac{1}{2}x) \\
&= 1,
\end{aligned}$$

by definition of the Gamma function. Its MGF is

$$M(t) = \frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}.\int_0^\infty e^{tx}.x^{\frac{1}{2}n-1}\exp(-\frac{1}{2}x)dx$$

$$= \frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)}.\int_0^\infty x^{\frac{1}{2}n-1}\exp(-\frac{1}{2}x(1-2t))dx.$$

Substitute $u := x(1-2t)$ in the integral. One obtains

$$M(t) = (1-2t)^{-\frac{1}{2}n} . \frac{1}{2^{\frac{1}{2}n}\Gamma(\frac{1}{2}n)} . \int_0^\infty u^{\frac{1}{2}n-1} e^{-u} du = (1-2t)^{-\frac{1}{2}n},$$

by definition of the Gamma function. //

Q4. (i) $A^T A$ is symmetric, so $P = A(A^T A)^{-1} A^T$ is symmetric. $P^2 = A(A^T A)^{-1} A^T A (A^T A)^{-1} A^T = A(A^T A)^{-1} A^T = P$.
(ii) $(I-P)^2 = I - 2P + P^2 = I - 2P + P = I - P$, so $I - P$ is a (symmetric) projection.
(a) $tr(A+B) = tr(A) + tr(B)$ follows as the trace is additive from its definition.
$tr(AB) = \sum_i (AB)_{ii} = \sum_i \sum_j a_{ij} b_{ji}$, and this is $tr(BA)$ on interchanging the dummy duffices i and j.
(b) $tr(P) = tr(A(A^T A)^{-1} A^T) = tr(A^T A (A^T A)^{-1}) = tr(I_p) = p$, as A is $n \times p$, so $A^T A$ is $p \times p$.
$tr(I-P) = tr(I_n) - tr(P) = n - p$, as $P = A(A^T A)^{-1} A^T$ is $n \times n$.

NHB