Section 4.4. Multiparameter Likelihood Functions

Note. Consider a discrete probability space (Y, p) so that Y is finite or countable and $p : Y \to [0, 1]$ with the property that $\sum_{y \in Y} p(y) = 1$. For a "sample" $y = (y_1, y_2, \ldots, y_n)$, where each $y_i \in Y$, the *likelihood function* is $L(y) = \prod_{i=1}^n p(y_i)$. The *log likelihood function* is

$$\ell(y) = \log(L(y)) = \log\left(\prod_{i=1}^n p(y_i)\right) = \sum_{i=1}^n \log(p(y_i)).$$

Note. Gentle states that "if the distribution is the d-variate normal distribution with mean d-vector μ and $d \times d$ positive definite variance-covariance matrix Σ" (page 163) then the likelihood function is

$$L(\mu, \Sigma; y) = \frac{1}{((2\pi)^{d/2}\text{det}(\Sigma)^{1/2})^n} \exp\left(\frac{-1}{2}\sum_{i=1}^n (y_i - \mu)^T \Sigma^{-1}(y_i - \mu)\right).$$

So this gives the log likelihood function

$$\begin{aligned} \ell(\mu, \Sigma; y) &= \log\left((2\pi)^{-nd/2}\text{det}(\Sigma)^{-n/2} \exp\left(\frac{-1}{2}\sum_{i=1}^n (y_i - \mu)^T \Sigma^{-1}(y_i - \mu)\right)\right) \\ &= \frac{-nd}{2}\log(2\pi) - \frac{n}{2}\log(\text{det}(\Sigma)) - \frac{1}{2}\sum_{i=1}^n (y_i - \mu)^T \Sigma^{-1}(y_i - \mu). \end{aligned}$$

Now a quadratic form $x^T A x$ is a scalar (or a $1 \times t$ matrix), so $x^T A x = \text{tr}(X^T A x)$. By Exercise 3.2.E, $\text{tr}(AB) = \text{tr}(BA)$ so $x^T A x = \text{tr}(x^T A x) = \text{tr} A x x^T)$. This implies

$$(y_i - \mu)^T \Sigma^{-1}(y_i - \mu) = \text{tr}(\Sigma^{01}(y_i - \mu)(y_i - \mu)^T).$$

So we can rewrite the log likelihood function as

$$\ell(\mu, \Sigma; y) = c - \frac{n}{2}\log(\text{det}(\Sigma)) = \frac{1}{2}\text{tr}\left(\Sigma^{-1}\sum_{i=1}^n (y_i - \mu)(y_i \mu)^T\right)$$

where $c = (-nd/2)\log(2\pi)$. As seen in Section 2.3, $\|x\|^2 = \|\overline{x}\|^2 + \|x - \overline{x}\|^2$, so with $x = y - \mu$ we have

$$\begin{aligned}
\|x\|^2 &= \|y-\mu\|^2 = \langle y-\mu, y-\mu\rangle = \sum_{i=1}^{n}(y_i-\mu)(y_i-\mu)^T \\
&= \|\overline{x}\|^2 + \|x-\overline{x}\|^2 = \|\overline{y}-\mu\|^2 + \|(y-\mu)-(\overline{y}-\mu)\|^2 \\
&= \|\overline{y}-\mu\|^2 + \|y-\overline{y}\|^2 \\
&= (\overline{y}-\mu)(\overline{y}-\mu)^T + \sum_{i=1}^{n}(y_i-\overline{y})(y_i-\overline{y})^T.
\end{aligned}$$

So we can also write

$$\begin{aligned}
\ell(\mu,\Sigma;y) &= c - \frac{n}{2}\log(\det(\Sigma)) \\
&\quad -\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\left((\overline{y}-\mu)(\overline{y}-\mu)^T + \sum_{i=1}^{n}(y_i-\overline{y})(y_i-\overline{y})^T\right)\right) \\
&= c - \frac{n}{2}\log(\det(\Sigma)) - \frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\sum_{i=1}^{n}(y_i-\overline{y})(y_i-\overline{y})^T\right) \\
&\quad -\frac{1}{2}\mathrm{tr}(\Sigma^{-1}(\overline{y}-\mu)(\overline{y}-\mu)^T)
\end{aligned}$$

Note. In *maximum likelihood estimates* the likelihood function is maximized (or, equivalently, the log likelihood is maximized) by keeping sample y fixed and manipulating the parameters of the probability distribution (such as μ and Σ in the d-variate normal distribution above). Exercise 4.5 illustrates this.

Revised: 4/25/2018