Chapter 14 Solutions

Andrew Wu
Wasserman: All of Statistics

August 1, 2025

Problem 14.1. Let a be a vector of length k and let X be a random vector of the same length with mean μ and variance Σ. Prove that $\mathbb{E}(a^T X) = a^T \mu$ and $\mathbb{V}(a^T X) = a^T \Sigma a$. If A is a matrix with k columns, then $\mathbb{E}(AX) = A\mu$ and $\mathbb{V}(AX) = A\Sigma A^T$.

Solution. See the solution to problem 3.20. $\square$

Problem 14.2. Find the Fisher information matrix for the MLE of a Multinomial.

Solution. Let $(X_1, \dots, X_k) = X \sim \text{Multinomial}(n, p) = \text{Multinomial}(n, (p_1, \dots, p_k))$.

First, we'll compute the MLE of a Multinomial. We'll prove that the MLE is $\widehat{p} = \left(\frac{X_1}{n}, \frac{X_2}{n}, \dots, \frac{X_k}{n}\right)$.

The likelihood function is

$$\mathcal{L}(p) = \binom{n}{x_1, \dots, x_k} \prod_{j=1}^{k} p_j^{x_j}$$

so the log-likelihood comes out as

$$\begin{aligned} \ell(p) &= \log\left[\binom{n}{x_1, \dots, x_k} \prod_{j=1}^{k} p_j^{x_j}\right] \\ &= \log\binom{n}{x_1, \dots, x_k} + \sum_{j=1}^{k} x_j \log p_j \end{aligned}$$

so maximizing $\ell(p)$ is equivalent to maximizing $\sum_{j=1}^{k} x_j \log p_j$. Note also the constraint that $\sum_{j=1}^{k} p_j = 1$.

We substitute $p_k = 1 - p_1 - p_2 - \dots - p_{k-1}$, so that

$$\frac{\partial \ell}{\partial p_1} = \frac{x_1}{p_1} + \frac{x_k}{1 - p_1 - p_2 - \dots - p_{k-1}} \cdot (-1)$$

so setting this partial equal to 0, we obtain

$$p_1 x_k = x_1(1 - p_1 - p_2 - \dots - p_{k-1})$$

or equivalently, $\frac{p_1}{p_k} = \frac{x_1}{x_k}$. Of course, this generalizes, so that $\frac{p_i}{p_j} = \frac{x_i}{x_j}$ for all $1 \le i, j \le k$. Thus we get

$$\frac{\sum_{j=1}^{k} p_j}{p_1} = \frac{\sum_{j=1}^{k} x_j}{x_1}$$

so $\widehat{p}_1 = \frac{x_1}{\sum_{j=1}^{k} x_j} = \frac{X_1}{n}$. So, generalizing, we have

$$\widehat{p} = (\widehat{p}_1, \dots, \widehat{p}_k) = \left(\frac{X_1}{n}, \dots, \frac{X_k}{n}\right).$$

Now we compute the Fisher Information matrix. Again, we set $p_k = 1 - p_1 - p_2 - \ldots - p_{k-1}$, so that

$$\ell(p) = \log \binom{n}{x_1, \ldots, x_k} + \sum_{j=1}^{k-1} x_j \log p_j + x_k \log(1 - p_1 - p_2 - \ldots - p_{k-1}).$$

We have, for $1 \leq i, j \leq k-1$,

$$\begin{aligned} H_{ii} &= \frac{\partial^2 \ell}{\partial p_i^2} \\ &= \frac{\partial}{\partial p_i} \left[\frac{x_i}{p_i} - \frac{x_k}{1 - p_1 - p_2 - \ldots - p_{k-1}} \right] \\ &= -\frac{x_i}{p_i^2} - \frac{x_k}{(1 - p_1 - p_2 - \ldots - p_{k-1})^2} \cdot (-1) \cdot \frac{d}{dp_i}(1 - p_1 - p_2 - \ldots - p_{k-1}) \\ &= -\frac{x_i}{p_i^2} - \frac{x_k}{p_k^2}. \end{aligned}$$

Next, we have

$$\begin{aligned} H_{ij} &= \frac{\partial^2 \ell}{\partial p_i \partial p_j} \\ &= \frac{\partial}{\partial p_i} \left[\frac{x_j}{p_j} - \frac{x_k}{1 - p_1 - p_2 - \ldots - p_{k-1}} \right] \\ &= -\frac{x_k}{p_k^2}. \end{aligned}$$

So $\mathbb{E}_p(H_{ii}) = -\frac{n}{p_i} - \frac{n}{p_k}$ and $\mathbb{E}_p(H_{ij}) = -\frac{n}{p_k}$. It follows that

$$I_n(\theta) = \begin{bmatrix} \frac{n}{p_1} + \frac{n}{p_k} & \frac{n}{p_k} & \cdots & \frac{n}{p_k} \\ \frac{n}{p_k} & \frac{n}{p_2} + \frac{n}{p_k} & \cdots & \frac{n}{p_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{n}{p_k} & \frac{n}{p_k} & \cdots & \frac{n}{p_{k-1}} + \frac{n}{p_k} \end{bmatrix}.$$

$\square$