2. Quick Review of Covariance Matrices and the Log-Likelihood Function

Let $\mathbf{X}$ be a random vector of dimension $d \times 1$ with expectation $\mu_{\mathbf{X}}$. Recall from Lecture 10 that the covariance matrix Σ is defined as the following matrix outer product:

$$\Sigma = \mathbb{E}\left[(\mathbf{X}-\mu_{\mathbf{X}})(\mathbf{X}-\mu_{\mathbf{X}})^T\right].$$

It can be shown (similar to the covariance of random variables X, Y in Lecture 10) that

$$\begin{aligned}\Sigma &= \mathbb{E}\left[\mathbf{X}\mathbf{X}^T\right] - \mathbb{E}\left[\mathbf{X}\right]\mathbb{E}[\mathbf{X}]^T \\ &= \mathbb{E}\left[\mathbf{X}\mathbf{X}^T\right] - \mu_{\mathbf{X}}\mu_{\mathbf{X}}^T.\end{aligned}$$

Review of Covariance Matrices

3/3 points (graded)

Consider the following random vector of dimension $d \times 1$: $\mathbf{X} = \left[X^{(1)}, X^{(2)}, \ldots, X^{(d)}\right]^T$ is equally likely to be one of $[1, 0, \ldots, 0]^T$, $[0, 1, \ldots, 0]^T$, ..., $[0, 0, \ldots, 1]^T$. That is, $\mathbf{X}$ is equal to any of the unit vectors along the coordinate axes with probability $\frac{1}{d}$.

Let us compute the entries of the covariance matrix $\Sigma_{ij} = \mathsf{Cov}\left(X^{(i)}, X^{(j)}\right)$.

$\mathsf{Cov}\left(X^{(i)}, X^{(i)}\right) =$ 1/d - 1/d^2 Answer: 1/d - 1/d^2

$\frac{1}{d} - \frac{1}{d^2}$

With $i \neq j$, $\mathsf{Cov}\left(X^{(i)}, X^{(j)}\right) =$ - 1/d^2 Answer: -1/d^2

$-\frac{1}{d^2}$

Is Σ a singluar covariance matrix? **Note:** A matrix Σ is singular if $\det(\Sigma) = 0$.

- (●) Yes
- ( ) No

STANDARD NOTATION

Solution:

For any $i \in \{1, 2, \ldots, d\}$,

$$\begin{aligned}\mathsf{Cov}\left(X^{(i)}, X^{(i)}\right) &= \mathsf{Var}\left(X^{(i)}\right) \\ &= \frac{1}{d} - \frac{1}{d^2},\end{aligned}$$

as each $X^{(i)}$ is equal to 1 with probability $\frac{1}{d}$ and equal to 0 with probability $1 - \frac{1}{d}$.

For any $i \neq j$, $\mathbb{E}\left[X^{(i)}X^{(j)}\right] = 0$ as $X^{(i)}$ and $X^{(j)}$ are never both equal to 1 at the same time. Therefore,

$$\begin{aligned}\text{Cov}\left(X^{(i)}, X^{(j)}\right) &= \mathbb{E}\left[X^{(i)}X^{(j)}\right] - \mathbb{E}\left[X^{(i)}\right]\mathbb{E}\left[X^{(j)}\right] \\ &= -\frac{1}{d^2}.\end{aligned}$$

The covariance matrix looks as follows:

$$\Sigma = \begin{bmatrix} \frac{1}{d} - \frac{1}{d^2} & -\frac{1}{d^2} & \cdots & -\frac{1}{d^2} \\ -\frac{1}{d^2} & \frac{1}{d} - \frac{1}{d^2} & \cdots & -\frac{1}{d^2} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{d^2} & -\frac{1}{d^2} & \cdots & \frac{1}{d} - \frac{1}{d^2} \end{bmatrix}.$$

Adding all the rows and replacing row 1 with the result yields

$$\widehat{\Sigma} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ -\frac{1}{d^2} & \frac{1}{d} - \frac{1}{d^2} & \cdots & -\frac{1}{d^2} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{d^2} & -\frac{1}{d^2} & \cdots & \frac{1}{d} - \frac{1}{d^2} \end{bmatrix}.$$

From the above, we can see that the determinant of $\widehat{\Sigma}$ is equal to 0. This means that Σ, which is row-equivalent to $\widehat{\Sigma}$, is a singular covariance matrix.

提交 你已经尝试了2次（总共可以尝试2次）

Answers are displayed within the problem

Dimensions of Gradient of Log-Likelihood Function

1/2 points (graded)

Let $\left(E, (\mathbf{P}_\theta)_{\theta \in \Theta}\right)$ be a statistical model associated with a random vector $\mathbf{X}$ of dimension $k \times 1$. Let $f_\theta(\mathbf{x})$ be the joint pdf of $\mathbf{X}$ and let $\theta \in \mathbb{R}^d$.

Let the log-likelihood function associated with one observation of $\mathbf{X}$ be denoted $\ell_1(\mathbf{x}, \theta)$. For simplicity, let $\ell_1(\mathbf{x}, \theta)$ be denoted $\ell(\theta)$, where it is assumed that $\mathbf{x}$ is fixed.

Assuming that $\ell(\theta)$ is differentiable with respect to θ for almost all $\mathbf{x}$, what are the dimensions of the gradient $\nabla \ell(\theta)$?

Number of rows in $\nabla \ell(\theta)$: d

d

Answer: d + 0*k

Number of columns in $\nabla \ell(\theta)$: k

k

Answer: 1 + 0*d + 0*k

Solution:

$\ell(\theta)$, at any given $\mathbf{x}$, is a real-valued function of d variables in the parameter $\theta \in \mathbb{R}^d$.

Therefore, the gradient vector $\nabla \ell(\theta)$ is of size $d \times 1$.

提交 你已经尝试了1次（总共可以尝试1次）

Answers are displayed within the problem

Log-Likelihood Function of a Bernoulli-like Random Variable

0/1 point (graded)
Consider the following experiment: You take a coin that lands a head (H) with probability $0 < p < 1$ and you toss it twice. Define X as the following random variable:

$$X = \begin{cases} 1 & \text{if outcome is HH} \\ 0 & \text{otherwise} \end{cases}$$

Let $\ell(p)$ be the log-likelihood function of X when written as a random function, i.e. all of the x in the function written as X. What is $\ell(p)$?

Hint: Write the pmf of X as a one-line formula.

(Enter **X** for X, and **ln(y)** for $\ln(y)$. Do not enter “log".)

$\ell(p) =$ ln(p^(2*X)*(1-p)^(2-2*X)) **Answer:** 2*X*ln(p) + (1-X)*ln(1-p^2)

STANDARD NOTATION

我这个是两次都是T才是0的概率

Solution:

First, X takes on 1 with probability p^2 and 0 with probability $1-p^2$.

Finding the log-likelihood function involves writing down the pmf of X as a one-line equation:

$$(p^2)^x \cdot (1-p^2)^{1-x}, \quad x \in \{0,1\}.$$

Taking logarithm and replacing all x with X yields the desired log-likelihood function written as a random function.

提交 你已经尝试了3次（总共可以尝试3次）

Answers are displayed within the problem

讨论

显示讨论

主题: Unit 3 Methods of Estimation:Lecture 11: Fisher Information, Asymptotic Normality of MLE; Method of Moments / 2. Quick Review of Covariance Matrices and the Log-Likelihood Function