6. Maximum Likelihood Estimation for a Multivariate Standard Normal

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n \overset{i.i.d.}{\sim} \mathcal{N}(\mu, \mathbf{1})$, where $\mu \in \mathbb{R}^d$ and $\mathbf{1}$ is the $d \times d$ identity matrix. (The $\mathbf{X}_i$ are random vectors.)

Recall the pdf defining the distribution $\mathcal{N}(\mu, \mathbf{1})$ is

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T \mathbf{1}(\mathbf{x}-\mu)\right)$$

(a)

1.0/1 point (graded)

What is the likelihood function $L(\mathbf{X}_1, \ldots, \mathbf{X}_n, \mu)$ for μ?

(Enter **(Sigma_i(norm(x_i-mu)^2))** for $\sum_{i=1}^{n} \|\mathbf{x}_i - \mu\|^2$.)

$L(\mathbf{X}_1, \ldots, \mathbf{X}_n, \mu) =$ 1/(2*pi)^(d*n/2)*exp(-1/2*(Sigma_i(norm(x_i-mu)^2)))

Answer: (2*pi)^(-n*d/2)*exp(-1/2*Sigma_i(norm(x_i-mu)^2))

Solution:

$$\begin{aligned} L(\mathbf{X}_1, \ldots, \mathbf{X}_n, \mu) &= \prod_{i=1}^{n} \frac{1}{(2\pi)^{d/2}} \exp(-\frac{1}{2}(\mathbf{x}_i-\mu)^T \mathbf{1}(\mathbf{x}_i-\mu) \\ &= \prod_{i=1}^{n} \frac{1}{(2\pi)^{d/2}} \exp(-\frac{1}{2}\|\mathbf{x}_i-\mu\|_2^2) \\ &= (2\pi)^{-nd/2} \exp(-\frac{1}{2}\sum_{i=1}^{n}\|\mathbf{x}_i-\mu\|_2^2) \end{aligned}$$

提交 你已经尝试了1次（总共可以尝试3次）

Answers are displayed within the problem

(b)

1/1 point (graded)

Compute the maximum likelihood estimator $\hat{\mu}_{MLE}$ for μ.

(Enter **barX_n** for the sample average.)

$\hat{\mu}_{MLE} =$ barX_n **Answer:** barX_n

Prove to yourself that the result you obtained above indeed maximizes the likelihood function. Is this step necessary?

STANDARD NOTATION

Solution:

The log likelihood function is

$$\ell(\mu) = \frac{nd}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{n}\|\mathbf{x}_i - \mu\|_2^2$$

The gradient of the log likelihood function is

$$\nabla\ell(\mu) = \sum_{i=1}^{n}(\mathbf{x}_i - \mu)$$

Setting the gradient to zero

$$\begin{aligned}
\sum_{i=1}^{n}(\mathbf{x}_i - \mu) &= 0 \\
\sum_{i=1}^{n}\mathbf{x}_i - n\mu &= 0 \\
\mu &= \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i \\
\hat{\mu}_{MLE} &= \bar{\mathbf{X}}_n
\end{aligned}$$

提交 你已经尝试了1次（总共可以尝试3次）

Answers are displayed within the problem

(c)

1/1 point (graded)

What is the distribution of $\hat{\mu}_{MLE}$?

- (selected) $\hat{\mu}_{MLE} \sim \mathcal{N}(\mu, \frac{1}{n}\mathbf{1})$
- $\hat{\mu}_{MLE} \sim \mathcal{N}(\mu, \mathbf{1})$
- $\hat{\mu}_{MLE} \sim \mathcal{N}(0, \frac{1}{n}\mathbf{1})$
- $\hat{\mu}_{MLE} \sim \mathcal{N}(\mu, \frac{1}{\sqrt{n}}\mathbf{1})$

Solution:

When the distribution of the population is normal, then the distribution of the sample mean is also normal. For a normal population distribution with mean μ and variance σ^2, the distribution of the sample mean is normal, with mean μ and variance $\frac{\sigma^2}{n}$. So in this multivariate case,

$$\hat{\mu}_{MLE} \sim \mathcal{N}(\mu, \frac{1}{n}\mathbf{1})$$

提交 你已经尝试了1次（总共可以尝试2次）

Answers are displayed within the problem

(d)

1.0/1 point (graded)

What is the asymptotic variance of $\mathbf{A}\hat{\mu}_{MLE}$? (here, A is a fixed $m \times d$ matrix)

(If applicable, enter **trans(A)** for the transpose of a matrix A.)

A*trans(A)

Answer: A*trans(A)

STANDARD NOTATION

Solution:

$\mathbf{A}\hat{\mu}_{MLE} \in \mathbb{R}^m$, so its variance is actually a $m \times m$ covariance matrix.

$$\begin{aligned} \text{Cov}\left(\mathbf{A}\hat{\mu}_{MLE}\right) &= \mathbf{A}\text{Cov}\left(\hat{\mu}_{MLE}\right)\mathbf{A}^T \\ &= \mathbf{A}\frac{1}{n}\mathbf{1}\mathbf{A}^T \\ &= \frac{1}{n}\mathbf{A}\mathbf{A}^T \end{aligned}$$

提交 你已经尝试了1次（总共可以尝试3次）

Answers are displayed within the problem

(e)

1.0/1 point (graded)

What is the asymptotic variance of $\|\hat{\mu}_{MLE}\|^2$?

(If applicable, enter **norm(v)** for the norm $\|\mathbf{v}\|$ of a vector $\mathbf{v}$, and **trans(v)** for the transpose $\mathbf{v}^T$ of a vector $\mathbf{v}$.)

4*norm(mu)*norm(tran:

Answer: 4*norm(mu)^2

Solution:

Define the function $g\left(\mathbf{X}\right) = \mathbf{X}^T\mathbf{X}$ (i.e., $g\left(\mathbf{X}\right)$ is the squared norm of a vector $\mathbf{X}$).

$$\begin{aligned} g\left(\mathbf{X}\right) &= \mathbf{X}^T\mathbf{X} \\ \nabla g\left(\mathbf{X}\right) &= 2\mathbf{X} \end{aligned}$$

We know from part(c) that

$$\hat{\mu}_{MLE} \sim \mathcal{N}\left(\mu, \frac{1}{n}\mathbf{1}\right)$$

So

$$\sqrt{n}\left(\hat{\mu}_{MLE} - \mu\right) \sim \mathcal{N}\left(0, \mathbf{1}\right)$$

Note that this is stronger than saying that convergence in distribution.

By multivariate delta method,

$$\begin{aligned} \sqrt{n}\left(g\left(\hat{\mu}_{MLE}\right) - g\left(\mu\right)\right) \xrightarrow[n\to\infty]{(d)} \mathcal{N}\left(0, \nabla g(\mu)^T\mathbf{1}\nabla g\left(\mu\right)\right) &= \mathcal{N}\left(0, \left(2\mu\right)^T\left(2\mu\right)\right) \\ &= \mathcal{N}\left(0, 4\|\mu\|^2\right) \end{aligned}$$

Therefore, the asymptotic variance is $4\|\mu\|^2$

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:Homework 6 Maximum Likelihood Estimation and Method of Moments / 6. Maximum Likelihood Estimation for a Multivariate Standard Normal