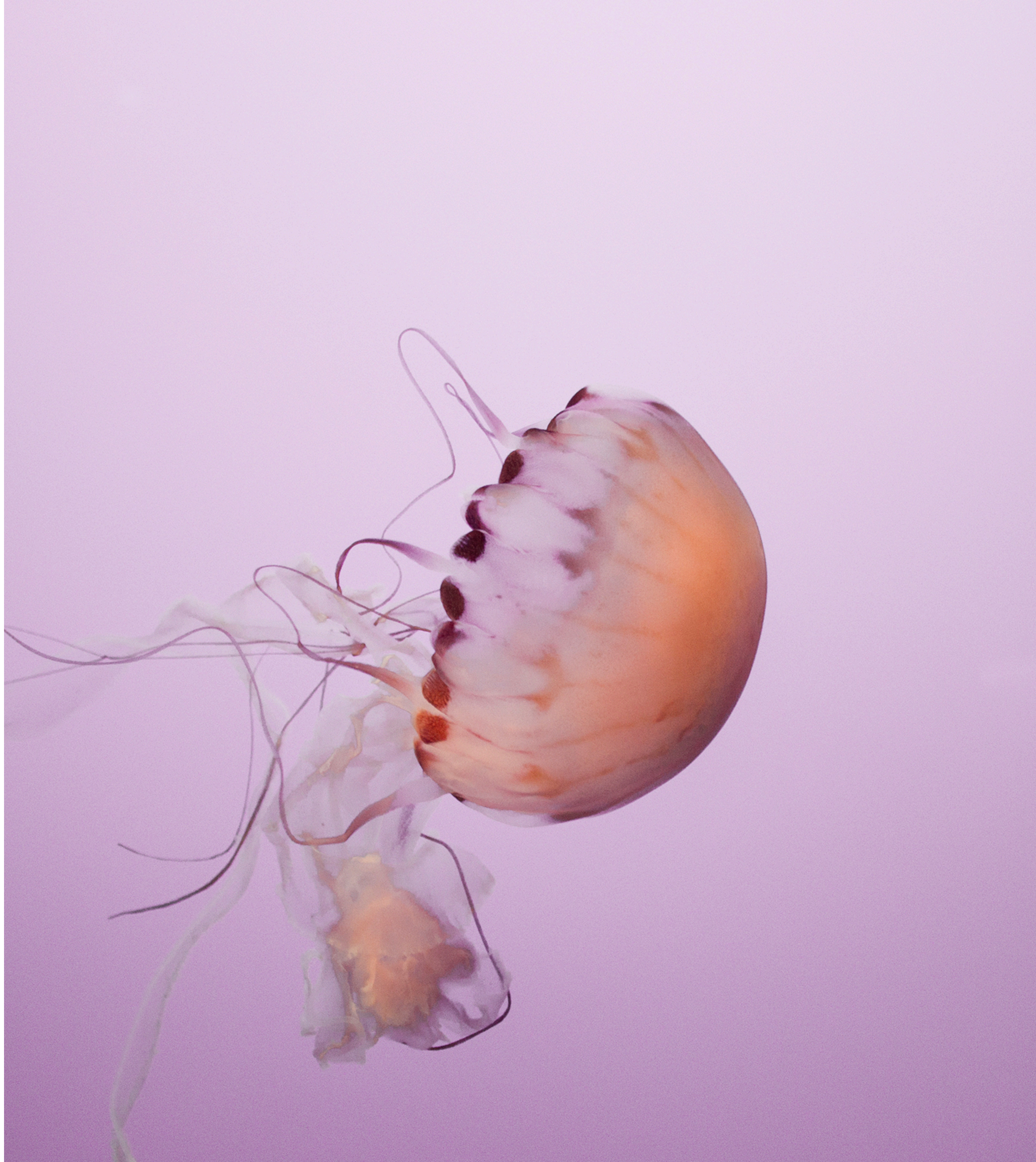
STAT3602
Statistical Inference

Example Class 9
2020-2021 Semester 1

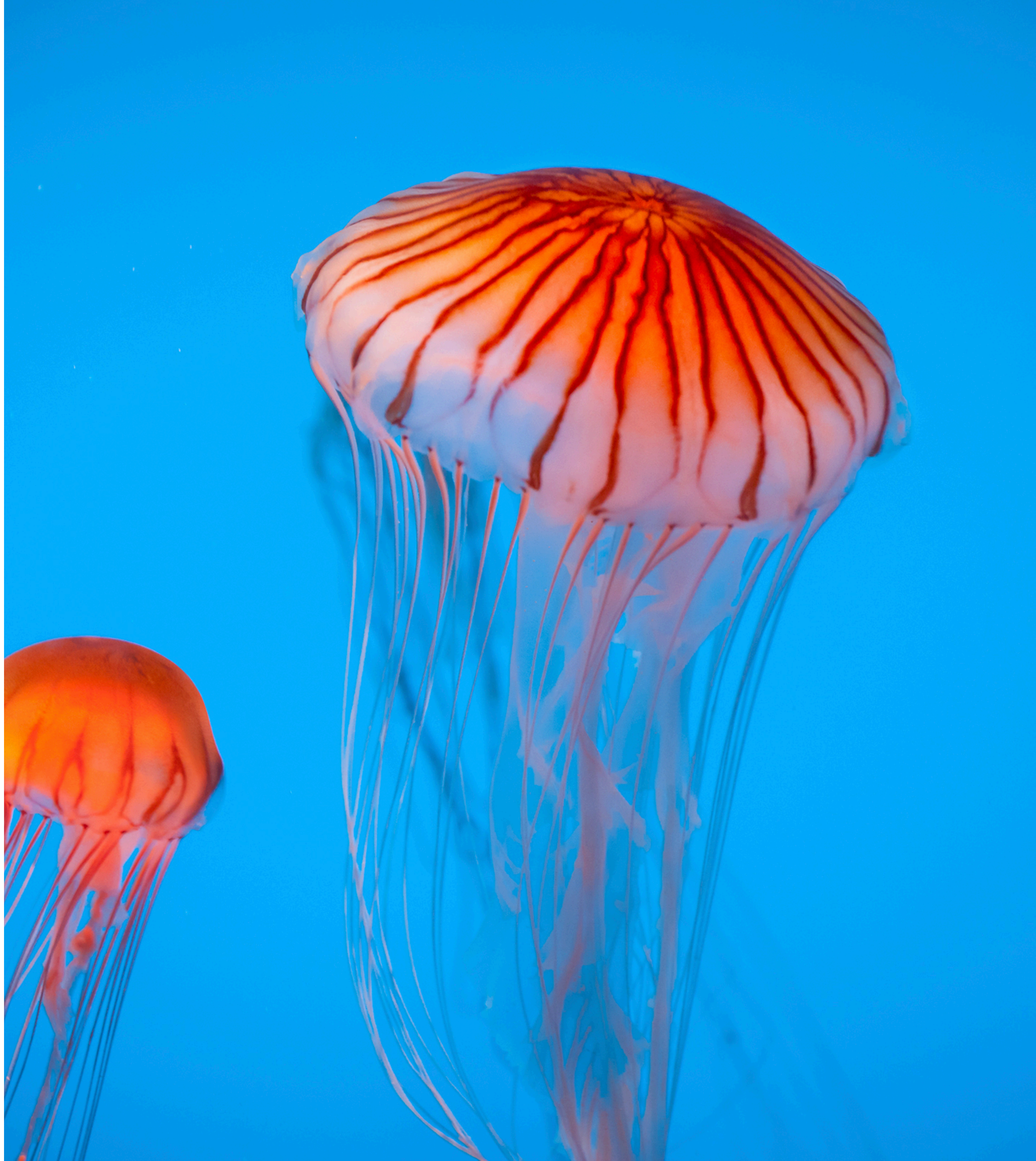
LIU CHEN
RM205, RRS
Nov, 18, 2020

Outline
Chapter 5 (Part II)

- Review:
  - Information Inequality
  - Maximum Likelihood Estimator
- Exercise
  - Opinions of interviewers
  - Maximum Likelihood Estimator and the Exponential Distribution



Information Inequality



Information Inequality:

!!Regularity Assumptions!!

1. $X \sim f(\cdot \mid \theta), \theta = \left(\theta_1, \ldots, \theta_k\right)^\top \in \Theta \subset \mathbb{R}^k$, where Θ contains an open rectangle,

2. Sample Space $\{\boldsymbol{x} : f(\boldsymbol{x} \mid \theta) > 0\} = \mathcal{S}$ is common to all $\theta \in \Theta$,

3. $x \in \mathcal{S}, \theta \in \Theta$ and $i = 1, \ldots, k$, $\dfrac{\partial f(\boldsymbol{x} \mid \theta)}{\partial \theta_i}$ exists and is finite.

Information Inequality:

Definitions: under regularity assumptions

The score function

$$\boldsymbol{U}(\theta) = \left[U_1(\theta), \ldots, U_k(\theta)\right]^\top$$

$$= \left[\frac{\partial \ln f(\boldsymbol{X}|\theta)}{\partial \theta_1}, \ldots, \frac{\partial \ln f(\boldsymbol{X}|\theta)}{\partial \theta_k}\right]^\top,$$

Fisher information matrix

$$I(\theta) = \mathbb{E}_\theta\left[\boldsymbol{U}(\theta)\boldsymbol{U}(\theta)^\top\right],$$

$$I_{ij}(\theta) = \mathbb{E}_\theta\left[U_i(\theta)U_j(\theta)\right]$$

$$= \mathbb{E}_\theta\left[\left(\frac{\partial \ln f(\boldsymbol{X}|\theta)}{\partial \theta_i}\right)\left(\frac{\partial \ln f(\boldsymbol{X}|\theta)}{\partial \theta_j}\right)\right]$$

$$= -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f(\boldsymbol{X}|\theta)\right].$$

Information Inequality:

!! Theorem 5.4.4 !!

1. $T = T(\mathbf{X})$ be a statistic with $\mathbb{E}_\theta\left[T^2\right] < \infty$.
2. Assume that

$$\boldsymbol{\alpha}(\theta) = \left[\frac{\partial}{\partial\theta_1}\mathbb{E}_\theta[T], \ldots, \frac{\partial}{\partial\theta_k}\mathbb{E}_\theta[T]\right]^\top$$ exists,

   and can obtained by differentiating under the integral sign.
3. Assuming that $I(\theta)$ is positive definite,
4. Then, under regularity assumptions

We have

$$\mathrm{Var}_\theta(T) \geq \boldsymbol{\alpha}(\theta)^\top I(\theta)^{-1}\boldsymbol{\alpha}(\theta).$$

Information Inequality:

!! Theorem 5.4.4 !!

1. $T = T(\mathbf{X})$ be a statistic with $\mathbb{E}_\theta\left[T^2\right] < \infty$.
2. Assume that
$$\boldsymbol{\alpha}(\theta) = \left[\frac{\partial}{\partial\theta_1}\mathbb{E}_\theta[T], \ldots, \frac{\partial}{\partial\theta_k}\mathbb{E}_\theta[T]\right]^\top$$
exists, and can obtained by differentiating under the integral sign.
3. Assuming that $I(\theta)$ is positive definite,
4. Then, under regularity assumptions

Let $\psi(\theta)$ be a differentiable function of θ, and T be an unbiased estimator of $\psi(\theta)$.

We have,

$$\mathrm{Var}_\theta(T) \geq \left[\frac{\partial\psi(\theta)}{\partial\theta_1}, \ldots, \frac{\partial\psi(\theta)}{\partial\theta_k}\right] I(\theta)^{-1} \left[\frac{\partial\psi(\theta)}{\partial\theta_1}, \ldots, \frac{\partial\psi(\theta)}{\partial\theta_k}\right]^\top .$$

Maximum Likelihood Estimator

Definition. Suppose $\hat{\theta}$ maximises $\ell_{\boldsymbol{X}}(\theta)$, or equivalently, $S_{\boldsymbol{X}}(\theta)$. We say $\hat{\theta}$ is the ***maximum likelihood estimator*** (mle) of θ .

Suppose $\theta \in \mathbb{R}^k$. Usually $\hat{\theta}$ can be obtained by solving

likelihood equations: $U(\theta) = \mathbf{0}$,

i.e. $\dfrac{\partial}{\partial \theta_j} S_{\boldsymbol{X}}(\theta) = 0, \quad j = 1, \ldots, k.$

Convergence

Definition 1.1 A sequence of real numbers $a_1, a_2, \ldots$ has limit equal to the real number a if for every $\epsilon > 0$, there exists N such that

$$|a_n - a| < \epsilon \quad \text{for all } n > N.$$

In this case, we write $a_n \to a$ as $n \to \infty$ or $\lim_{n\to\infty} a_n = a$ and we could say that "a_n converges to a".

Modes of Convergence

1. Convergence in Probability

Definition 2.1 Let $\{X_n\}_{n\geq 1}$ and X be defined on the same probability space. We say that X_n converges in probability to X, written $X_n \xrightarrow{P} X$, if for any $\epsilon > 0$,

$$P(|X_n - X| < \epsilon) \to 1 \text{ as } n \to \infty. \tag{2.1}$$

Modes of Convergence

1. Convergence in Probability
2. **Convergence in Distribution**

Suppose that X has distribution function $F(x)$ and that X_n has distribution function $F_n(x)$ for each n. Then we say X_n converges in distribution to X, written , $X_n \xrightarrow{d} X$, if $F_n(x) \to F(x)$ for all x at which $F(x)$ is continuous.

Modes of Convergence

1. Convergence in Probability $X_n \xrightarrow{P} X$
2. Convergence in Distribution $X_n \xrightarrow{d} X$
3. Theory

$$\text{If } X_n \xrightarrow{P} X, \text{ then } X_n \xrightarrow{d} X$$

Modes of Convergence

1. **Convergence in Probability**
2. **Convergence in Distribution**
3. **Theory** If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{d} X$
4. **Convergence in Mean**

Definition 2.15 Let a be a positive constant. We say that X_n converges in ath mean to X, written $X_n \xrightarrow{a} X$, if

$$\mathrm{E}\,|X_n - X|^a \to 0 \text{ as } n \to \infty. \tag{2.11}$$

Maximum Likelihood Estimator

Theorem: Subject to regularity conditions on $p_1(\cdot \mid \theta), \ldots, p_n(\cdot \mid \theta)$, we have

(i) $\hat{\theta}_n$ *converges in probability* to θ_0,

(ii) $n^{1/2}(\hat{\theta}_n - \theta_0)$ *converges in distribution* to $N\big(\mathbf{0}, \mathscr{I}(\theta_0)^{-1}\big)$,

(iii) $n^{-1/2}\boldsymbol{U}(\theta_0)$ converges in distribution to $N\big(\mathbf{0}, \mathscr{I}(\theta_0)\big)$,

Exercise 1

A team of 8 interviewers were recruited to conduct a survey in which each interviewer asked for the opinion (positive, negative, neutral) of 10 respondents about a political issue. The following table summarizes the counts of the three response categories found from the survey:

Interviewer	1	2	3	4	5	6	7	8	Total
Positive	2	3	2	1	0	4	2	2	16
Negative	5	7	6	6	4	6	7	7	48
Neutral	3	0	2	3	6	0	1	1	16

Exercise 1

Interviewer	1	2	3	4	5	6	7	8	Total
Positive	2	3	2	1	0	4	2	2	16
Negative	5	7	6	6	4	6	7	7	48
Neutral	3	0	2	3	6	0	1	1	16

Let X_i and Y_i be the counts of the positive and negative responses obtained by the i-th interviewer respectively, for i = 1, 2, ..., 8 . Assume that (X_i, Y_i) follows a multinomial distribution with mass function

$$f(x, y|p, q) = \begin{cases} \frac{10!}{x!y!(10-x-y)!} p^x q^y (1-p-q)^{10-x-y}, & x, y \in (0, 1, ..., 10), x + y \leq 10 \\ 0, & otherwise \end{cases}$$

for unknown parameters p and q which can be interpreted as the probabilities that a respondent gives positive and negative responses respectively.

Exercise 1: Problems

(a) Give the loglikelihood function of p and q based on the observed data.

(b) Find the maximum likelihood estimator of (p, q).

(c) Calculate the expected Fisher information matrix for p and q.

(d) Quoting the large-sample properties of MLE, describe the asymptotic distribution of $(\hat{p}, \hat{q})$

Exercise 1 (a): Solutions

for a multinomial distribution with parameters $(N, p, q, 1-p-q)$, the log-likelihood function is

$$
\begin{aligned}
&\log(l(p, q \mid N, x_1, \dots, x_n, y_1, \dots, y_n)) \\
=&\log\left(\prod_{i=1}^{n} \frac{N!}{x_i!y_i!(N-x_i-y_i)!} p^{x_i} q^{y_i} (1-p-q)^{10-x_i-y_i}, \quad x_i, y_i \in (0, 1, \dots, N), x_i + y_i \leq N\right) \\
=&\log p \sum x_i + \log q \sum y_i + \log(1-p-q) \sum (N - x_i - y_i) + C(x_1, \dots, x_n, y_1, \dots, y_n)
\end{aligned}
$$

Exercise 1 (b: MLP of p and q): Solutions

(a) Log-likelihood function

$$
\begin{aligned}
&\log(l(p, q \mid N, x_1, \ldots, x_n, y_1, \ldots, y_n)) \\
=&\log\left(\prod_{i=1}^{n} \frac{N!}{x_i! y_i! (N - x_i - y_i)!} p^{x_i} q^{y_i} (1-p-q)^{10-x_i-y_i}, \quad x_i, y_i \in (0, 1, \ldots, N), x_i + y_i \leq N\right) \\
=&\log p \sum x_i + \log q \sum y_i + \log(1-p-q) \sum (N - x_i - y_i) + C(x_1, \ldots, x_n, y_1, \ldots, y_n)
\end{aligned}
$$

(b)

$$
\frac{\partial}{\partial p} \log(l(p, q \mid N, x_1, \ldots, x_n, y_1, \ldots, y_n)) = \frac{\sum x_i}{p} - \frac{\sum (N - x_i - y_i)}{1 - p - q}
$$

$$
\frac{\partial}{\partial q} \log(l(p, q \mid N, x_1, \ldots, x_n, y_1, \ldots, y_n)) = \frac{\sum y_i}{q} - \frac{\sum (N - x_i - y_i)}{1 - p - q}
$$

Solving $\frac{\partial}{\partial p} \log\left(l\left(p, q \mid N, x_1, \ldots, x_n, y_1, \ldots, y_n\right)\right) = 0$ and $\frac{\partial}{\partial q} \log\left(l\left(p, q \mid N, x_1, \ldots, x_n, y_1, \ldots, y_n\right)\right) = 0$

, we have

$$
\hat{p} = \frac{\sum x_i}{Nn} = 0.2; \hat{q} = \frac{\sum y_i}{Nn} = 0.6
$$

Exercise 1 (c): Solutions

(c) Calculate the expected Fisher information matrix for p and q.

$$I(\theta) = -E_\theta\left[\frac{\partial^2}{\partial\theta\partial\theta^T}\log(f(\mathbf{X} \mid \theta))\right]$$

$$-\frac{\partial^2}{\partial p^2}\log(l(p,q \mid x,y)) = \frac{x}{p^2} + \frac{N-x-y}{(1-p-q)^2}$$

$$-\frac{\partial^2}{\partial p \partial q}\log(l(p,q \mid x,y)) = \frac{N-x-y}{(1-p-q)^2}$$

$$-\frac{\partial^2}{\partial q^2}\log(l(p,q \mid x,y)) = \frac{y}{q^2} + \frac{N-x-y}{(1-p-q)^2}$$

$$I(\theta) = n\begin{bmatrix} N/p + N/(1-p-q) & N/(1-p-q) \\ N/(1-p-q) & N/q + N/(1-p-q) \end{bmatrix}$$

Exercise 1 (d): Solutions

(d) Quoting the large-sample properties of MLE, describe the asymptotic distribution of $(\hat{p}, \hat{q})$

$$\mathcal{I}(\theta)^{-1} = \frac{1}{N}\begin{bmatrix} p(1-p) & -pq \\ -pq & q(1-q) \end{bmatrix}$$

$$\sqrt{n}\begin{pmatrix} \hat{p} - p \\ \hat{q} - q \end{pmatrix} \to N\left(0, \mathcal{I}(\theta)^{-1}\right) \text{ in distribution.}$$

$$\sqrt{n}\begin{pmatrix} \hat{p} - p \\ \hat{q} - q \end{pmatrix} \sim N\left(0, \mathcal{I}(\theta)^{-1}\right) \text{ approximately, for large n.}$$

Exercise 2: Problems

Let $X = (X_1, \ldots, X_n)$ be a random sample of size $n \geq 3$ from the exponential distributio

density $f(x) = \theta e^{-\theta x}$ (mean=$1/\theta$ and variance=$1/\theta^2$). Obtain the maximum likelihood es

tor $\hat{\theta}_n$ based on the sample of size n for θ.

Find an unbiased estimator which is a function of $\hat{\theta}_n$.

Calculate the Cramer-Rao Lower Bound for the variance of the unbiased estimator. Is the l

attained?

(Hint: Sum of n i.i.d $Exp(\theta)$ distributed random variables follows Gamma(n, θ). If

Gamma(k, θ), then $Y^{-1} \sim$ InverseGamma(k, θ) with mean $= \theta/(k-1)$ and variance $= \theta^2$/

$1)^2(k-2)])$

Exercise 2: Solutions

$$
\begin{aligned}
f(x \mid \theta) &= \theta^n e^{-\theta \sum x_i} \\
\ln f(x \mid \theta) &= n \ln \theta - \theta \sum x_i \\
\frac{\partial}{\partial \theta} \ln f(x \mid \theta) &= \frac{n}{\theta} - \sum x_i \\
I(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial \theta^2} \ln f(x \mid \theta)\right] &= \frac{n}{\theta^2}
\end{aligned}
$$

MLE of θ: $\hat{\theta} = \dfrac{n}{\sum X_i}$

The expected value of $\hat{\theta}$:

$$\mathbb{E}\left(\frac{n}{\sum X_i}\right) = n\mathbb{E}\left(Y^{-1}\right) = \frac{n}{n-1}\theta, \text{ where } Y^{-1} \sim \text{InverseGamma}(n, \theta).$$

Exercise 2: Solutions

MLE of θ: $\hat{\theta} = \frac{n}{\sum X_i}$

The expected value of $\hat{\theta}$: $\mathbb{E}\left(\frac{n}{\sum X_i}\right) = \frac{n}{n-1}\theta,$

$$\boxed{\text{Thus, The unbiased estimator of } \theta \text{ is } \frac{n-1}{\sum X_i}}$$

The variance of the unbiased estimator is

$$\begin{aligned}\mathrm{Var}\left(\frac{n-1}{\sum X_i}\right) &= (n-1)^2 \mathrm{Var}\left(Y^{-1}\right) \\ &= (n-1)^2 \frac{\theta^2}{(n-1)^2(n-2)} = \frac{\theta^2}{n-2}\end{aligned}.$$

Since $\frac{\theta}{n-2} > I^{-1}(\theta) = \frac{\theta}{n}$, The CRLB is not attained. In fact, as $\sum X_i$ is a complete sufficient statistic, the CRLB cannot be attained by any unbiased estimator.