VE401 Recitation Class Note7

Estimation

Chen Siyi
siyi.chen_chicy@sjtu.edu.cn

1 Parameter Estimation

1.1 Definition

Random Sample:
A random sample of size n from the distribution of X is: a collection of n independent random variables X_1, ... , X_n, each with the same distribution as X.

Parameter θ:
A constant value, but always unknown, and we want to estimate its value. Example: mean μ of the variable X.

Statistic:
A **random variable** derived from a random sample of a population. Example: sample mean $\overline{X}$ of a random sample.

Estimator $\hat{\theta}$:
A statistic used to estimate a parameter. It's also a **random sample**.

Point Estimate:
The **specific value** of the statistic.

1.2 Bias and MSE

When estimating parameters, we wish the estimation can be more precise. So we wish the estimator $\hat{\theta}$ is:

1. having small difference from the parameter θ
2. having small variance $Var[\widehat{\theta}]$

Bias:

$$E[\theta] - \hat{\theta}$$

Unbiased:

$$E[\theta] = \hat{\theta}$$

Mean Square Error:

$$\begin{aligned}\mathrm{MSE}(\widehat{\theta}) &= \mathrm{E}\left[(\widehat{\theta} - \mathrm{E}[\widehat{\theta}])^2\right] + (\theta - \mathrm{E}(\widehat{\theta}))^2 \\ &= \mathrm{Var}\,\widehat{\theta} + (\text{ bias })^2\end{aligned}$$

Comment:

1. Bias is only a measurement of our first expectation on θ.
2. MSE measures the overall quality of an estimator.
3. So in general, unbiased estimators are preferred but sometimes biased estimators are used.

1.3 Estimators for μ and σ^2

Sample Mean:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

1. $E[\overline{X}] = \mu$, so $\overline{X}$ is an unbiased estimator of μ
2. $Var[\overline{X}] = \frac{\sigma^2}{n}$
3. MSE($\overline{X}$) $= Var[\overline{X}]$, so MSE can be deduced by choosing a larger sample size.

Sample Variance:

$$S^2 := \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \bar{X}\right)^2$$

1. $E[S^2] = \sigma^2$, so S^2 is an unbiased estimator of σ^2.

Comments:

1. The above properties exist no matter what kind of distributions $\overline{X}$ and S^2 have.
2. For point estimate, not necessary to know their distributions.
3. For interval estimation later, we need to analyze their distribution.

2 Find Estimator for Parameter

2.1 The Method of Moments

Recall Moments: k^{th} moment of X: $E[X^k]$

Theorem:

$$\widehat{\mathrm{E}[X^k]} = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

This is an unbiased estimator for $E[X^k]$. Simply means $E[\frac{1}{n}\sum_{i=1}^{n} x_i^k] = E[X^k]$

General Steps:

1. Express a parameter in terms of moments, such as $Var[X] = E[X^2] - E[X]^2$.
2. Insert the estimators for these moments to obtain an estimator for the parameter.

Comment:
The found estimators may not be unbiased.

2.2 The Method of Maximum Likelihood

Assumption:

1. X_θ is a random variable, whose PDF can be written out with an unknown parameter θ as f_{X_θ}.
2. X_1, X_2, X_3,...,X_n is a random sample.
3. x_1, x_2, x_3,...,x_n are the "yielded" values of the random sample. Which we consider as constant in the below steps.

General Steps:

1. Define the likelihood function L by

$$L(\theta) = \prod_{i=1}^{n} f_{X_\theta}(x_i)$$

2. Find the expression for θ of $(x_1, x_2, ..., x_n)$ that maximizes $L(\theta)$. This can also be done by maximizing $ln(L(\theta))$

3. In the founded expression, replace θ with $\widehat{\theta}$, replace $(x_1, x_2, ..., x_n)$ with $(X_1,\ X_2,\ X_3,...,X_n)$. Then we obtain the maximum likelihood estimator $\widehat{\theta}$.

Comment:

1. Generally we do not need $(x_1, x_2, ..., x_n)$ to be concrete known numbers, but we treat the "yielded" values as constant, in our steps finding the expression for θ.
2. Since what we generally find is an expression, we define the estimator $\widehat{\theta}$(which is a random variable) based on the expression obtained.
3. A point estimate can be obtained by putting the observed values of $(x_1, x_2, ..., x_n)$ into the expression for $\widehat{\theta}$ and calculate.

Question: Find Estimator for exp.

Recall the exponential distribution has PDE: $f_\beta(x) = \begin{cases} \beta e^{-\beta x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$

1. Find an estimator for β using the method of maximum likelihood.
2. Find an estimator for β^2 using the method of maximum likelihood if $f_\beta(x) = \frac{x}{\beta^2} e^{-\frac{x^2}{2\beta^2}}$.

* In real-life application, when exp. distribution is used to describe failure density, we always not wait for all machines to fail. So there's another method. See reference 1 and 2.

Answer: Find Estimator for exp.

1. Assume a random sample yields values $(x_1, x_2, ..., x_n)$.

$$L(\beta) = \beta^n e^{-\beta(x_1+x_2+...+x_n)}$$

$$\frac{d(L\beta)}{d\beta} = \beta^{n-1}e^{-\beta(x_1+x_2+...+x_n)}(n - \beta(x_1 + x_2 + ... + x_n))$$

So $L(\beta)$ gets its maximum when:

$$\beta = \frac{n}{x_1 + x_2 + ... + x_n} = \frac{1}{\overline{x}}$$

Therefore the estimator is:

$$\widehat{\beta} = \frac{1}{\overline{X}}$$

2. Be careful you view β^2 as a whole. You can choose to replace $\beta^2 = s$ and then find $\widehat{s}$.

The process is similar, but you may need to use $ln(L(s))$. The result is:

$$\widehat{\beta^2} = \frac{\overline{X^2}}{2}$$

2.3 Failure of the Methods

Question: German Tank Problem

In world war 2, the Allies want to know the number of the German tanks. Suppose the German mark their tanks with positive integers form 1 to N, if there are totally N German tanks. After a battle, the Allies checked the destroyed 4 German tanks' numbers, finding the numbers are: 5, 7, 2, 19.
Problem: Estimate N.
Try to solve this problem using the method of moments, and the method of most likelihood. What have you noticed?

Answer: German Tank Problem

Both methods will fail.

1. Method of Moments: $N = 2E[X]$, $\widehat{E[X]} = \widehat{\mu} = \overline{X}$. So $\widehat{N} = 2\overline{X}$ Then the point estimate gives: $N = 2\overline{x} = 16.5$.
2. Method of Maximum Likelihood: $L(N) = (\frac{1}{N})^N$, $\frac{d(ln(LN)}{dN} = ln\frac{1}{N} - 1$ So the smaller N, the larger L(N). Notice N $\leq$ max X_i. N = max X_i. Then the point estimate gives: $N = 2\overline{x} = 19$.

Both ridiculous. So these two method might fail together.
How to solve the GTP? See References:
1. https://en.wikipedia.org/wiki/German_tank_problem
2. Reference 3, page 23-25

3 Interval Estimation

3.1 Confidence Interval

Confidence Interval:
Let $0 \leq \alpha \leq 1$. A $100(1-\alpha)\%$ confidence interval for a parameter θ is an interval $[L_1, L_2]$ such that

$$P[L_1 \leq \theta \leq L_2] = 1 - \alpha$$

Centered Confidence Interval:

$$P[\theta < L_1] = P[\theta > L_2] = \frac{\alpha}{2}$$

One-sided Confidence Interval:
A $100(1-\alpha)\%$ Upper Confidence Interval $\theta < L$.
With the upper confidence bound satisfying:

$$P[\theta < L_1] = 1 - \alpha$$

A $100(1-\alpha)\%$ Lower Confidence Interval $\theta > L$.
With the upper confidence bound satisfying:

$$P[\theta > L_1] = 1 - \alpha$$

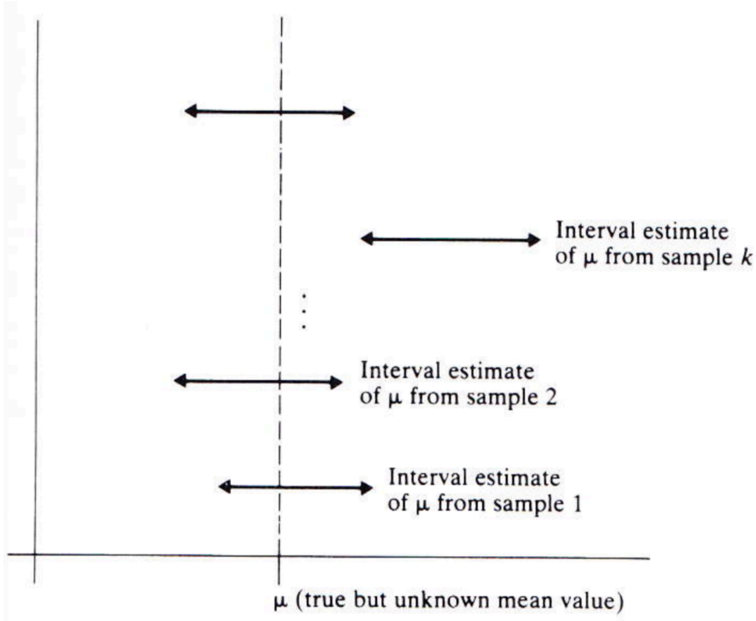

Figure 1: Random Intervals

Question: Random Intervals

What does $100(1-\alpha)\%$ stands for? Is the below statement right?
We have generated a numerical interval L_0 for θ, from a concrete data, there are $100(1-\alpha)\%$ probability that $\theta \in L_0$.

Answer: Random Intervals

Of course wrong. Let's think about an unfair coin...

1. Flow a coin: generate a numerical interval from a concrete data.
2. Coin turn head: The generated interval contains θ.
3. Coin turn tail: The generated interval excludes θ.

At this stage, we do not know the exact value of θ, we just can not know exactly whether a generated interval contains θ.
Like you flow a coin, but only the god sees the result.
The true meaning for $100(1-\alpha)\%$ is:
Of all the intervals constructed for θ by using $[L_1, L_2]$, $100(1-\alpha)\%$ of them contain $\theta \in L_0$.

Introduction:

Generally, to construct confidence intervals for θ, we need to:

Find a statistic whose expression involves θ, and whose probability distribution is known at least approximately.

For example if we know the statistic $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ follows a standard normal distribution, then...

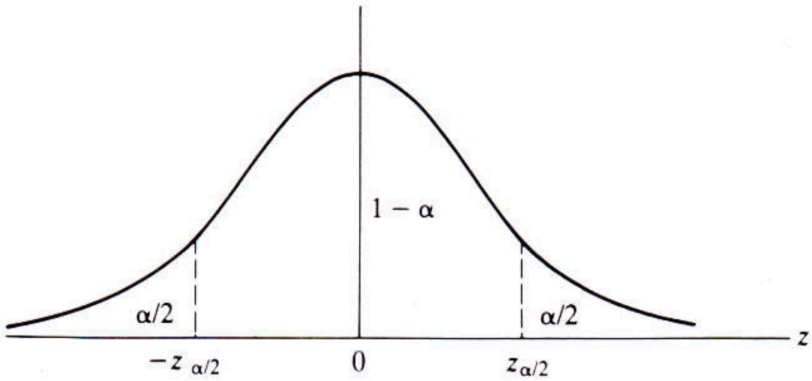

Figure 2: Partition of Z

Let's review estimation for μ and σ^2 as detailed examples.

3.2 Distributions of $\overline{X}$ and S^2

Sample Mean $\overline{X}$:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Theorem:

Let X_1, X_2, X_3,...,X_n be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then $\overline{X}$ is approximately normal with mean μ and variance $\frac{\sigma^2}{n}$.

Central Limit Theorem:

Let X_1, X_2, X_3,...,X_n be a random sample of size n from a distribution with mean μ and variance σ^2. Then for large n, $\overline{X}$ is approximately normal with mean μ and variance $\frac{\sigma^2}{n}$.

One Statistic:

The below random variable is standard normal distributed:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

Comment:

Using the method of transforming random variables, when X does not follow a normal distribution, we can obtain $\overline{X}'s$ actual distribution. But when sample size is large, $\overline{X}'s$ distribution can be approximated as normal distribution.

Sample Variance S^2:

$$S^2 := \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \bar{X}\right)^2$$

Theorem:

Let X_1, X_2, X_3,...,X_n be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then:

1. The below random variable is chi-squared distributed with n − 1 degrees of freedom:

$$\chi^2_{n-1} = \frac{(n-1)S^2}{\sigma^2}$$

2. The sample mean $\overline{X}$ and the sample variance S^2 are independent.

Mix $\overline{X}$ and S^2:

Let X_1, X_2, X_3,...,X_n be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then the below random variable is T distributed with n − 1 degrees of freedom:

$$T_{n-1} = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

3.3 Estimate Mean

Overview:

Distribution of X_i	Sample size n	Variance σ^2	Statistic	$1-\alpha$ confidence interval
$X_i \sim \mathcal{N}(\mu, \sigma)$	any	known	$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$	$\left[\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$
$X_i \sim$ any distribution	large	known	$\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0,1)$	$\left[\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$
$X_i \sim$ any distribution	large	unknown	$\frac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}} \sim \mathcal{N}(0,1)$	$\left[\overline{X} - z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \overline{X} + z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right]$
$X_i \sim \mathcal{N}(\mu, \sigma)$	small	unknown	$\frac{\overline{X}-\mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$	$\left[\overline{X} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \overline{X} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right]$
$X_i \sim$ any distribution	small	known or unknown	Go home!	Go home!

One-sided Confidence Interval:

1. When using the statistic $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$–

   A $100(1-\alpha)\%$ Upper Confidence Interval $\mu < L$:

   $$L = \overline{X} + \frac{z_\alpha \cdot \sigma}{\sqrt{n}}$$

   A $100(1-\alpha)\%$ Lower Confidence Interval $\mu > L$:

   $$L = \overline{X} - \frac{z_\alpha \cdot \sigma}{\sqrt{n}}$$

2. When using the statistic $T_{n-1} = \frac{\bar{X}-\mu}{S/\sqrt{n}}$–

   A $100(1-\alpha)\%$ Upper Confidence Interval $\mu < L$:

   $$L = \overline{X} + \frac{t_{\alpha,n-1} \cdot S}{\sqrt{n}}$$

   A $100(1-\alpha)\%$ Lower Confidence Interval $\mu > L$:

   $$L = \overline{X} - \frac{t_{\alpha,n-1} \cdot S}{\sqrt{n}}$$

3. When using the statistic $Z = \frac{\overline{X}-\mu}{S/\sqrt{n}}$, you can conclude by yourself.

3.4 Estimate Variance

Overview:

Distribution of X_i	Sample size n	Mean μ	Statistic	$1-\alpha$ confidence interval
$X_i \sim \mathcal{N}(\mu, \sigma)$	any	known or unknown	$\frac{s^2(n-1)}{\sigma^2} \sim \chi^2_{n-1}$	$\left[\frac{s^2(n-1)}{\chi^2_2}, \frac{s^2(n-1)}{\chi^2_1}\right]$

One-sided Confidence Interval:

A $100(1-\alpha)\%$ Upper Confidence Interval:

$$\left[0, \frac{(n-1)S^2}{\chi^2_{1-\alpha,n-1}}\right]$$

A $100(1-\alpha)\%$ Lower Confidence Interval:

$$\left[\frac{(n-1)S^2}{\chi^2_{\alpha,n-1}}, \infty\right)$$

3.5 Demo: Central Limit Theorem