Overview of Sampling Theory: Estimation Methods and Sampling Strategies

Mainak Thakur

31.08.2020

Mean Squared Error and Standard Error

- MSE: MSE of an estimator $\hat{\theta}$, MSE($\hat{\theta}$) = $E(\hat{\theta} - \theta)^2$
- MSE($\hat{\theta}$)= Var($\hat{\theta}$)+ $[B(\hat{\theta})]^2$
- If $\hat{\theta}$ is unbiased for θ, then MSE($\hat{\theta}$)= Var($\hat{\theta}$)
- $\sqrt{Var(\hat{\theta})}$ is known as standard error of the estimator $\hat{\theta}$

Ref: http://mgmt.iisc.ac.in/CM/LectureNotes/statistical_inference_frequentist.pdf
https://bookdown.org/egarpor/inference/est-methods.html
https://people.missouristate.edu/songfengzheng/Teaching/MTH541/Lecture%20notes/MOM.pdf

Class of unbiased estimators

- Suppose a random sample $Y_1, Y_2, \ldots, Y_n$ is taken from a population having mean μ
- $E(Y_i) = \mu, \forall i$
- Y_2 can be an unbiased estimator of μ
- $\dfrac{Y_1 + Y_2 + Y_7}{3}$ is another unbiased estimator of μ

Ref: http://mgmt.iisc.ac.in/CM/LectureNotes/statistical_inference_frequentist.pdf
https://bookdown.org/egarpor/inference/est-methods.html
https://people.missouristate.edu/songfengzheng/Teaching/MTH541/Lecture%20notes/MOM.pdf
Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Estimation Methods: Maximum Likelihood Estimation

- Provided criteria under which a given estimator may be called good, we dont know how to find an estimator yet
- How about learning some automated methods to find good estimators?
- MLE is the most popular method to find estimates
- Can be automated to produce numerical estimates
- Works very well for large samples and have favorable properties

Ref: http://mgmt.iisc.ac.in/CM/LectureNotes/statistical_inference_frequentist.pdf
https://bookdown.org/egarpor/inference/est-methods.html
https://people.missouristate.edu/songfengzheng/Teaching/MTH541/Lecture%20notes/MOM.pdf
Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

MLE: Example

- Think of an experiment: tossing a coin two times independently. Let X be the random variable denoting "number of heads in two tosses" . $X \sim Bin(2, \theta)$,

- θ= P(getting head in a coin toss) and θ is either 0.2 or 0.8.

- The pmf is given by $p(x;\theta) = \binom{2}{x}\theta^x(1-\theta)^x$, x = 0, 1, 2.

- The pmf of X for possible θ values is provided below:

θ	x=0	x=1	x=2
0.2	0.64	0.32	0.04
0.8	0.04	0.32	0.64

Given a sample of two tosses how can you logically estimate θ?

Ref: http://mgmt.iisc.ac.in/CM/LectureNotes/statistical_inference_frequentist.pdf
https://bookdown.org/egarpor/inference/est-methods.html
https://people.missouristate.edu/songfengzheng/Teaching/MTH541/Lecture%20notes/MOM.pdf
Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

MLE: concept

- The concept of the maximum likelihood method was exercised in the last example. The unknown parameter θ is estimated with the value that maximizes the probability of obtaining an observed sample realization. The most probable value of θ is required to be chosen as the estimate according to the available data.

- Let $f(x_1, x_2, \ldots, x_n/\theta)$ be the joint pdf or pmf of the sample observations. For fixed θ, this is a function of sample observations. But when $x_1, x_2, \ldots, x_n$ are observed and θ is unknown, it can be considered as a function of θ, $L(\theta)$ which is known as likelihood function. So in order to be a maximum likelihood estimator of θ, $L(\hat{\theta}) = max(L(\theta))$ should be satisfied and $\hat{\theta}$ is the MLE.

- In general it is convenient to deal with ln $L(\theta)$ rather than $L(\theta)$. ln $L(\theta)$ gets its highest value for the same value of θ as $L(\theta)$ does.

Ref: http://mgmt.iisc.ac.in/CM/LectureNotes/statistical_inference_frequentist.pdf
https://bookdown.org/egarpor/inference/est-methods.html
https://people.missouristate.edu/songfengzheng/Teaching/MTH541/Lecture%20notes/MOM.pdf
Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Example

- Consider $x_1, x_2, \ldots, x_n$ to be a realization of random sample of independent observations from a Poisson distribution with parameter λ. The likelihood function can be derived as:

$$L(\lambda) = \frac{exp[-n\lambda]\lambda^{\sum x_i}}{\prod(x_i!)}$$

$$ln\ L(\lambda) = -n\lambda + \sum x_i\ ln\ \lambda - \sum ln(x_i!)$$

$$\frac{d\ ln\ L(\lambda)}{d\lambda} = -n + (\sum x_i)\frac{1}{\lambda}$$

The maximum likelihood estimate of λ is $\hat{\lambda} = \sum x_i\ /\ n$.

Ref: http://mgmt.iisc.ac.in/CM/LectureNotes/statistical_inference_frequentist.pdf
https://bookdown.org/egarpor/inference/est-methods.html
https://people.missouristate.edu/songfengzheng/Teaching/MTH541/Lecture%20notes/MOM.pdf
Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Method of Moments

- An unknown population X has parameter vector $\theta = [\theta_1, ..., \theta_k]'$. The population moments are functions of the unknown parameters. That is, population moment of order r is given by: $\alpha_r = \alpha_r(\theta_1, ..., \theta_k) = E(X^r)$.

- Given an i.i.d. sample $X_1, X_2, ..., X_n$, let us denote the sample non-central moment of order r by a_r:

$$a_r = \frac{1}{n}\sum X_i^r, \text{ r=1,2,3...}$$

The sample moment(s) are then equated with the theoretical population moment(s) and are then solved for $\theta_1, ..., \theta_k$. If there are k unknowns, one typically considers the first k non-central moments, whichever is convenient.

Ref: http://mgmt.iisc.ac.in/CM/LectureNotes/statistical_inference_frequentist.pdf
https://bookdown.org/egarpor/inference/est-methods.html
https://people.missouristate.edu/songfengzheng/Teaching/MTH541/Lecture%20notes/MOM.pdf
Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

Example

- Population $X \sim N(\mu, \sigma^2)$, i.i.d sample $X_1, X_2, \ldots, X_n$ from the population. Compute the MME of μ and σ^2.
- As there are two parameters, at least two equations are required.

1. $\alpha_1 = \alpha_1(\mu, \sigma^2) = E(X) = \mu$
2. $\alpha_2 = \alpha_2(\mu, \sigma^2) = E(X^2) = Var(X) + [E(X)]^2 = \sigma^2 + \mu^2$

The first two sample moments are given by

1. $a_1 = \bar{X}$
2. $a_2 = \dfrac{1}{n} \sum X_i^2$

Equating a_1 and α_1 gives: $\hat{\mu}_{MME} = \bar{X}$

Equating a_2 and α_2 gives (substituting μ by $\hat{\mu}_{MME} = \bar{X}$): $\sigma^2_{MME} = \dfrac{1}{n} \sum X_i^2 - (\bar{X})^2$

Ref: http://mgmt.iisc.ac.in/CM/LectureNotes/statistical_inference_frequentist.pdf
https://bookdown.org/egarpor/inference/est-methods.html
https://people.missouristate.edu/songfengzheng/Teaching/MTH541/Lecture%20notes/MOM.pdf
Fundamentals of Statistics, Vol 1 by Gun, Gupta, Dasgupta

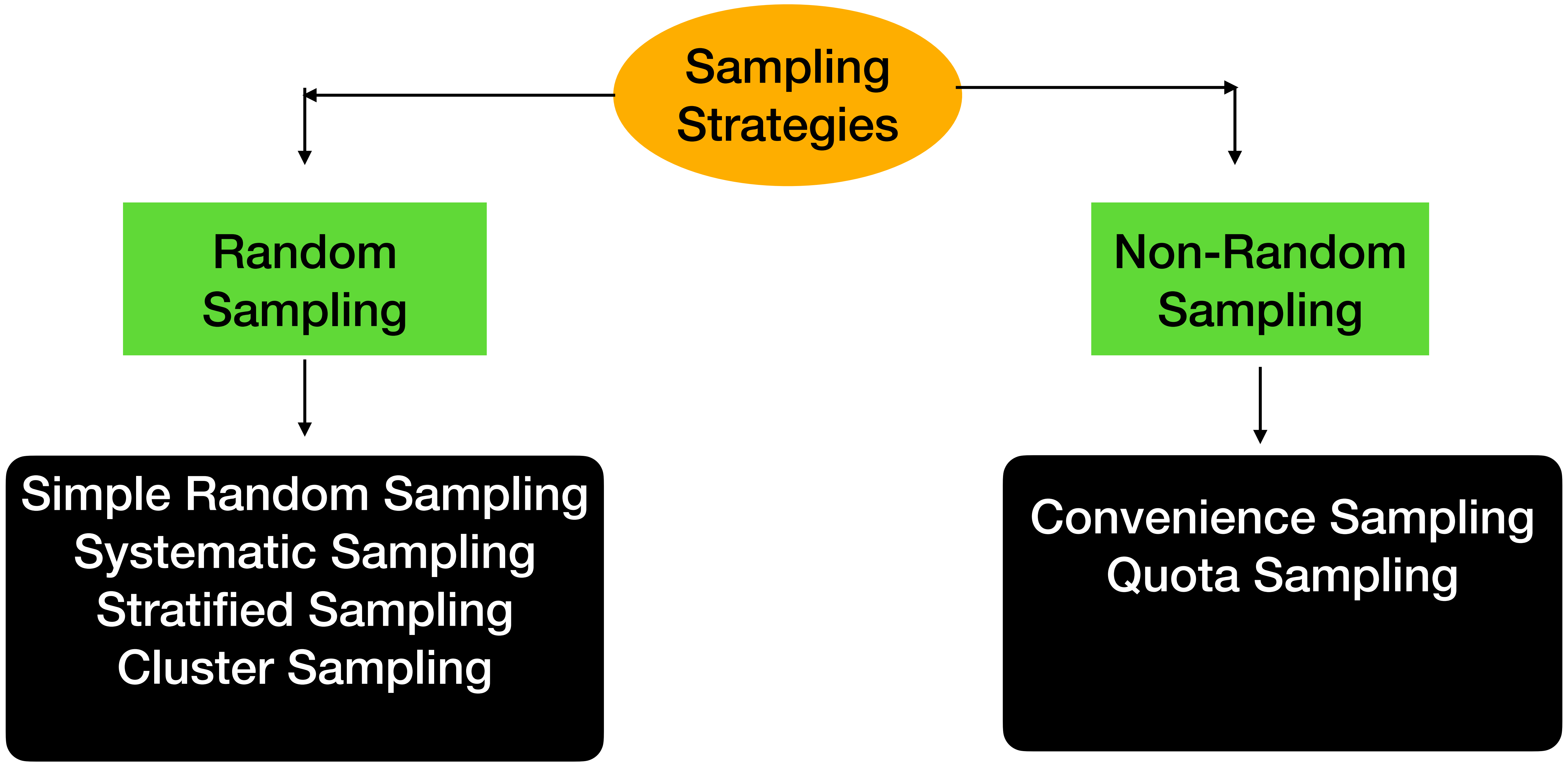
MLE vs MME

- MLE is statistically well established and has often lower variance than other methods
- MLE may be computationally very expensive and slow
- MMEs are easy to derive
- MMEs are often used to provide starting values while computing MLE
- MME may not be unique for a given sample

Ref: http://www.stat.tamu.edu/~hart/611/estimates.pdf
http://www.deduveinstitute.be/~opperd/private/max_likeli.html

Sampling Strategies

- Sampling strategies refer to different ways to choose members from the population. Biases may occur due to non-random selection leading to an unrepresentative sample of the population.
- Sampling strategies are broadly divided in two ways:



Ref: http://www2.hawaii.edu/~cheang/Sampling%20Strategies%20and%20their%20Advantages%20and%20Disadvantages.htm
https://people.richland.edu/james/lecture/m170/ch01-not.html
http://www.statstutor.ac.uk/resources/uploaded/13samplingtechniques.pdf
https://www.khanacademy.org/math/statistics-probability/designing-studies/sampling-methods-stats/a/sampling-methods-review
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.628.7338&rep=rep1&type=pdf
wikipidea