ISyE 6416: Computational Statistics
Spring 2017

Lecture 3: Basics of Statistical Inference

Prof. Yao Xie

H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Statistics and statistical thinking

The field of Statistics deals with the collection, presentation, analysis, and use of data to model systems, make decisions, solve problems, and design products and processes

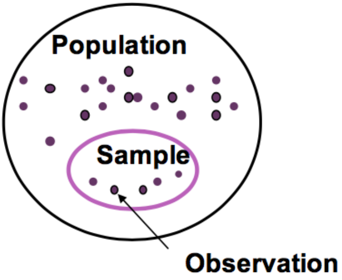
Statistics is the science of data

Examples: Statistics helps us

- Predict the demand of a product / the stock prices
- Select the best supplier with the least lead time (or highest quality)
- Monitor and control a process
- Simulate and model an ER
- Determine the probabilistic distribution of machines life
- Design new products

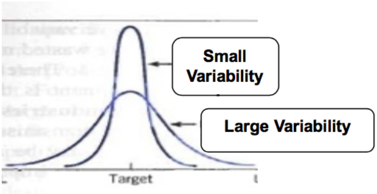
Population vs. sample

- Population: a finite well-defined group of ALL objects which, although possibly large, can be enumerated in theory
- Sample: A sample is a SUBSET of a population (e.g. select 50 out of 1,000 GT students for the survey)



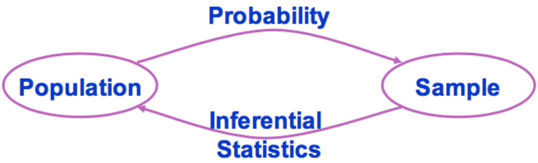
Variability in data

- Samples are random – individual variability
  - noise
  - measurement errors
  - the world has randomness
- Random results in variability:
  successive observations of a system or phenomenon do not produce exactly the same result.
- We capture the randomness by probability models



Probability vs. statistics

- Probability: given the information in the pail, what is in your hand?
- Statistics: given the information in your hand, what is in the pail?



- Commonly used probabilistic models
  - Discrete random variables
    - Bernoulli: $X \in \{0, 1\}$, $\mathbb{P}(X = 1|p) = p$
    - Binomial: $X \in \{0, 1, \ldots, n\}$, $X \sim$BIN(n, p), $\mathbb{P}(X = k|n, p) = \binom{n}{k} p^k (1-p)^{n-k}$
    - Geometric: $\mathbb{P}(X = k|p) = (1-p)^{k-1} p$, $k = 1, 2, \ldots$
  - Continuous random variables
    - Normal (Gaussian) distribution (central limit theorem): $\mathcal{N}(\mu, \sigma^2)$
    - Exponential distribution: $f(x|\lambda) = \lambda e^{-\lambda x}$, $x > 0$, $\lambda > 0$
    - Uniform distribution, Beta distribution, Gamma distribution...
- Fundamental statistical tasks
  - Point estimator
  - Confidence interval
  - Hypothesis testing
  - Regression analysis
  - Variable selection

Point estimator

- Digital thermometer takes measurements
- measurements subject to a random error additive to the true value
- If you take 6 measurements, and obtain a sequence of numbers

$$98.2 \quad 98.6 \quad 97.4 \quad 98.2 \quad 97.9 \quad 98.9$$

  What is the value of the true parameter?
- Methods for constructing point estimators: method-of-moments, **maximum likelihood**

Maximum likelihood estimator

- Maximum likelihood: assume data x following distribution with $f(x|\theta)$ with true parameter value θ
- likelihood function $\ell(\theta|x) = f(x|\theta)$, usually we consider log-likelihood $\log \ell(\theta|x)$
- maximum likelihood estimator

$$\hat{\theta} = \arg\max_{\theta} \ell(\theta|x)$$

  estimator $\hat{\theta}$ is a function of data x, and hence itself is random and has certain distribution.
- Example: maximum likelihood estimator for Gaussian $\mathcal{N}(\mu, \sigma^2)$
- property of an estimator
  - Bias: $|\hat{\theta} - \theta|$
    Unbias estimator has zero-bias.
  - Mean-square-error: $\mathbb{E}[(\hat{\theta} - \theta)^2]$

Bayesian estimator

- Assume the parameter has a prior distribution $\rho(\theta|\tau)$: τ hyper-parameter
- Posterior distribution of the parameter

$$f(\theta|x) = \frac{f(x|\theta)\rho(\theta|\tau)}{f(x)} \quad \text{(Bayes formula)}$$

  usually the marginal distribution $f(x)$ does not matter
- Maximum a-posterior (MAP) estimator

$$\hat{\theta} = \arg\max_{\theta} \log f(\theta|x) = \arg\max_{\theta} \ell(\theta|x)\rho(\theta|\tau)$$

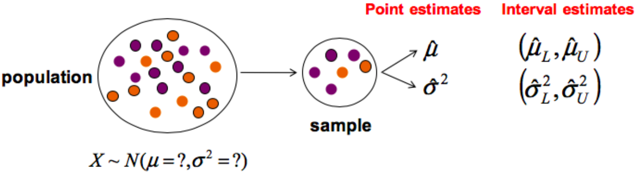
- Homework: compute the posterior distribution for Gaussian mean

Confidence interval

- if instead of asking "what is the most likely true temperature"
- we ask "what is a range [a, b], such that the true temperature is most likely to be within"
- Usually determined such that

$$\mathbb{P}(\theta \in [a, b]) = 1 - \alpha$$

$1 - \alpha$: confidence level, 0.95, 0.99



Example: For Gaussian mean, confidence interval is

$$[\bar{x} - t_{\alpha/2,n-1}\hat{\sigma}/\sqrt{n}, \bar{x} + t_{\alpha/2,n-1}\hat{\sigma}/\sqrt{n}].$$

Since $(\bar{x} - \mu)/(\hat{\sigma}/\sqrt{n})$ follows t_{n-1}-distribution

Cramer-Rao lower bound

Lower bound on the best estimator we can find.

Cramer-Rao lower bound (CRB)

Let $X_1, \ldots, X_n$ be a sample with pdf $f(x|\theta)$, and let $W(\mathbf{X}) = W_{X_1,\ldots,X_n}$ be any estimator satisfying

$$\frac{d}{d\theta} E_\theta W(\mathbf{X}) = \int \frac{d}{d\theta} [W(x) f(x|\theta)] dx$$

and $\mathsf{Var}_\theta W(\mathbf{X}) < \infty$. Then

$$\mathsf{Var}_\theta W(\mathbf{X}) \geq \frac{(\frac{d}{d\theta} E_\theta W(\mathbf{X}))^2}{E_\theta((\frac{d}{d\theta} \log f(\mathbf{X}|\theta))^2)} \quad \text{(CRB)}$$

Fisher information

CRB for i.i.d. samples

$$\text{Var}_\theta W(\mathbf{X}) \geq \frac{(\frac{d}{d\theta} E_\theta W(\mathbf{X}))^2}{n E_\theta((\frac{d}{d\theta} \log f(X|\theta))^2)} \quad \text{(CRB)}$$

- unbiased estimator $E_\theta W(\mathbf{X}) = \theta$
- Example: estimator for parameter of exponential distribution $\text{Var}_\lambda W \geq \frac{\lambda}{n}$ which is met by the sample mean $\frac{1}{n}\sum_{i=1}^{n} X_i$.
- another fact

$$E_\theta((\frac{d}{d\theta} \log f(X|\theta))^2) = -E_\theta(\frac{d^2}{d\theta^2} \log f(X|\theta)) = I(x|\theta)$$

Fisher information

Hypothesis test

- Production line produces a batch of 12 laptops
- Quality test shows that 1 out of 12 laptops's battery life time is shorter than design
- A technician claims the production line is defective.
- The claim is true of false?
- Setup: x_i, $i = 1, \ldots, 12$, $x_i \sim \mathcal{N}(\mu, \sigma^2)$, threshold

$$\begin{aligned} H_0 &: \quad \mu > t \quad \text{(null hypothesis)} \\ H_1 &: \quad \mu < t \quad \text{(alternative hypothesis)} \end{aligned}$$

t: threshold.
Which hypothesis is true?

Construct hypothesis test

- Likelihood ratio test (LRT)
- Consider **simple hypothesis test**: assuming x_i i.i.d. $\sim \mathcal{N}(\mu, \sigma^2)$, $i = 1, \ldots, n$

$$\begin{aligned} H_0 : &\quad \mu = \mu_0 \\ H_1 : &\quad \mu = \mu_1 \end{aligned}$$

- log-likelihood ratio

$$\log \ell(\mu_0, \mu_1) = \sum_{i=1}^{n} \log \frac{f(x_i|\mu_1)}{f(x_i|\mu_0)} \propto$$

- likelihood ratio test (LRT)
  Reject H_0 if $\log \ell(\mu_0, \mu_1) > b$, where b is a threshold.

- Performance metrics:

		truth	
		H_0 is True	H_0 is False
decision	Accept H_0	Correct Decision	Type II Error
	Reject H_0	Type I Error	Correct Decision

  α: type-I error; β: type-II error.
- Power of a test $= 1 - \beta$
- Neyman-Pearson lemma: likelihood ratio test is optimal (it achieves the smallest β for fixed α).
- Generalized likelihood ratio test (GLRT)

$$
\begin{aligned}
H_0 :& \quad \mu \in \Theta_0 \\
H_1 :& \quad \mu \in \Theta_1
\end{aligned}
$$

Reject null when

$$
\frac{\max_{\mu_1 \in \Theta_1} f(x_i|\mu_1)}{\max_{\mu_0 \in \Theta_0} f(x_i|\mu_0)} > b
$$

Neyman-Pearson lemma

The likelihood ratio test is the most poweful test for simple hypothesis.

Neyman-Pearson lemma

Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$, where the pdf corresponding to θ_i is $f(x|\theta_i)$, $i = 0, 1$, using the test that reject the null hypothesis when

$$\frac{f_1(x|\theta_1)}{f_0(x|\theta_0)} > b$$

for some threshold $b \geq 0$, and significant level

$$\alpha = \mathbb{P}_{f_0}(X \text{ leads to rejection}).$$

Such a test is the uniformly most powerful (UMP) test.

UMP Roughly means: has the smallest type-II error for given type-I error.

p-value

- In a nutshell, p-value is the probability that observing something more "extreme" than the data under the null hypothesis
- definition

$$p(x) = \sup_{\theta \in \Theta_0} \mathbb{P}_{f(X|\theta)}(W(X) \geq W(x))$$

$W(x)$ is the value of the test statistic calculated over the data

Relation between hypothesis test and confidence interval

- For instance

$$\begin{aligned} H_0: &\quad \mu = \mu_0 \\ H_1: &\quad \mu \neq \mu_0 \end{aligned}$$

- Construct Confidence Interval (CI) for μ

$$[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$$

- If CI does not include μ_0, then reject H_0

Example

9-47. Medical researchers have developed a new artificial heart constructed primarily of titanium and plastic. The heart will last and operate almost indefinitely once it is implanted in the patient's body, but the battery pack needs to be recharged about every four hours. A random sample of 50 battery packs is selected and subjected to a life test. The average life of these batteries is 4.05 hours. Assume that battery life is normally distributed with standard deviation $\sigma = 0.2$ hour.

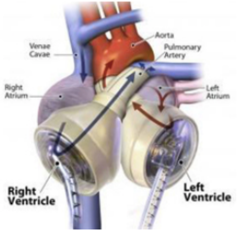
(a) Is there evidence to support the claim that mean battery life exceeds 4 hours? Use $\alpha = 0.05$.

(b) What is the P-value for the test in part (a)?

(c) Compute the power of the test if the true mean battery life is 4.5 hours.

(d) What sample size would be required to detect a true mean battery life of 4.5 hours if we wanted the power of the test to be at least 0.9?

(e) Explain how the question in part (a) could be answered by constructing a one-sided confidence bound on the mean life.



Basics of statistical inference

A quick overview of basic statistical inference problems and classic methods.

- point estimator
- confidence interval
- hypothesis test
- regression
- variable selection

Computational tasks are everywhere.