

W1211 Introduction to Statistics
Fall 2010

Lecture 14 | Nov 4th 2010

CLT

- Theorem:

The Central Limit Theorem (CLT)

Let X_1, X_2, \dots, X_n , be an i.i.d. sequence from a distribution with mean μ and variance σ^2 . Then if n is sufficiently large, the sample mean \bar{X} has approximately a normal distribution with $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}}^2 = \sigma^2/n$; And the sample total has approximately a normal distribution with $\mu_T = n\mu$, $\sigma_T^2 = n\sigma^2$. The larger the value of n , the better the approximation.

- Rule of Thumb: if $n > 30$, the CLT can be used.

Daily examples of statistics

Ex. This is an example of *statistical learning* as well as an example of *classification*.

The following table is summarized from 4601 email messages, in a study to try to predict whether the email was junk email, or “spam.” For all 4601 email messages, the true outcome (email type) *email* or *spam* is available, along with the relative frequencies of 57 of the most commonly occurring words and punctuation marks in the email message.

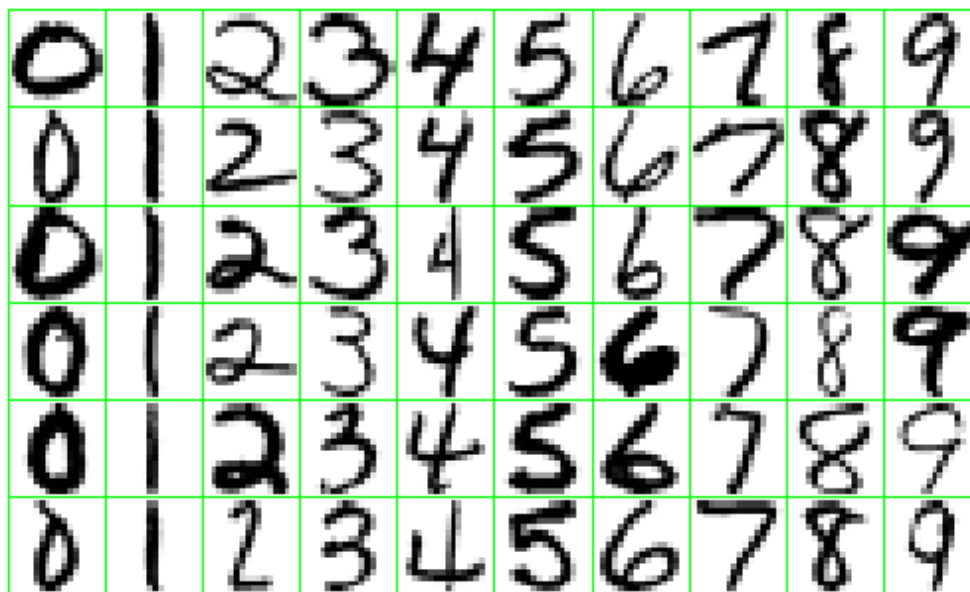
	george	you	your	hp	free	hpl	!	our	re	edu	remove
spam	0.00	2.26	1.38	0.02	0.52	0.01	0.51	0.51	0.13	0.01	0.28
email	1.27	1.27	0.44	0.90	0.07	0.43	0.11	0.18	0.42	0.29	0.01

Rule #1: if (%george < 0.6) & (%you > 1.5) then spam
else email.

Rule #2: if $(0.2 \cdot \%you - 0.3 \cdot \%george) > 0$ then spam
else email.

Daily examples of statistics

Ex. The data from this example come from the handwritten ZIP codes on envelopes from U.S. postal mail. The images are 16×16 eight-bit grayscale maps, with each pixel ranging in intensity from 0 to 255. Some sample images are shown in the following figure, The task is to predict, from the 16×16 matrix of pixel intensities, the identity of each image (0, 1, . . . , 9) quickly and accurately.

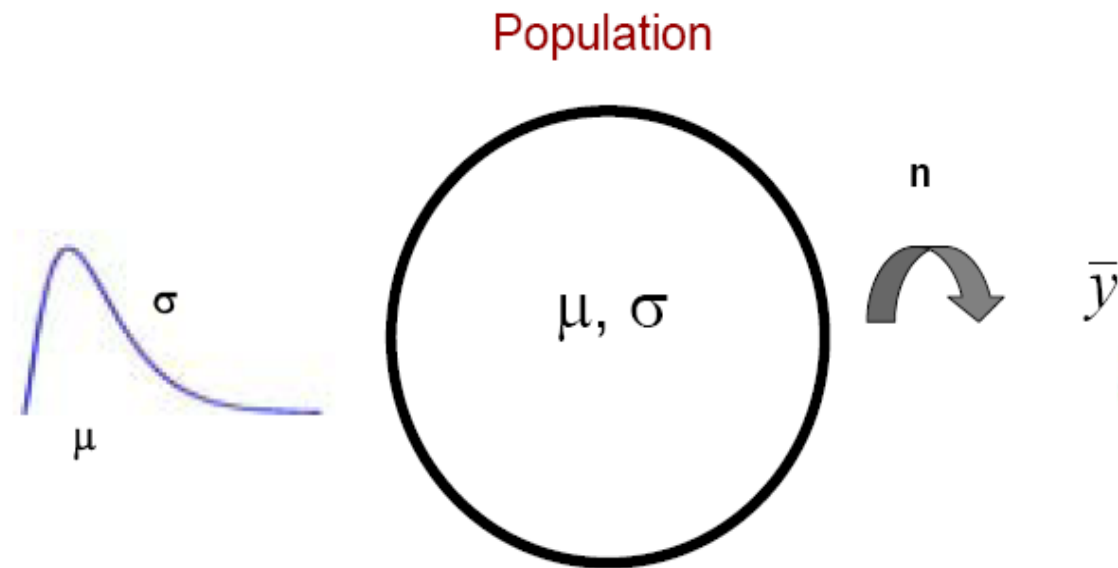


Statistical Inference

- From the previous two examples, we know that quite often, we need to infer the truth (**population**) from some partial information (**sample**).
- Question: why do we need a model?
- **Statistical inference** comprises the use of statistics and random sampling to make inferences concerning some unknown aspect of a population.
- A **point estimate** of a parameter θ is a single number that can be regarded as a sensible value for θ . A point estimate is obtained by selecting a suitable statistic and computing its value from the given sample data. The selected statistic is called the **point estimator** of θ .

Sampling scheme for a Mean

- Usually our problem set up will be as illustrated in the graph.



- The actual sample observations y_1, y_2, \dots, y_n (**realizations**) are assumed to be the result of a random sample Y_1, Y_2, \dots, Y_n (**random variables**) from a certain distribution.

Estimating probability

Ex. A biased coin has probability p of having heads and p is unknown. Suppose we flipped the coin for 100 times and had 73 heads. What is your best guess for p ?

Naturally, people would use estimator $\hat{p} = \frac{\text{number of heads}}{\text{number of flips}} = \frac{73}{100} = 0.73$

In other words, we are using the **sample proportion** to estimate the **population probability**.

Is this a good estimator? Are there any other estimators?

Guidelines on Estimation

1. Decide **what** (population mean/variance, proportion, etc.) to estimate and **how** (sample mean/variance, proportion, etc.) to estimate.
2. Obtain the sample, and calculate the estimator based on “**how**” from 1.
3. Report the estimator as well as its associated **error**.

More estimators

- In most problems, there will be **more than one** reasonable estimators.

Ex. Use `rnorm` to generate 20 normal random variables with mean 3 and standard deviation 1. Let's denote them as x_1, x_2, \dots, x_{20} . Now we pretend we do NOT know anything about the true underlying distribution. Come up with some estimators for the mean parameter.

(mean, trimmed-mean, median, partial mean)

Measure of a good Estimator

- Our estimator $\hat{\theta}$ is in fact a function of the sample x_i 's, therefore, it is also a random variable. For some samples, $\hat{\theta}$ may yield a value larger than θ , whereas for other samples $\hat{\theta}$ may underestimate θ .
- The quantity $\hat{\theta} - \theta$ characterize the error of estimation. A good estimator should result in small estimation errors.
- A commonly used measure of accuracy is the **mean square error**.

$$\text{MSE} = E(\hat{\theta} - \theta)^2$$

- However, since MSE will generally depend on the value of θ , finding an estimator with smallest MSE is typically **NOT** possible.

Unbiased Estimators

- One way to find good estimators, is to restrict our attention just to estimators that have some specified desirable properties and then find the best in this restricted group.
- One popular property is *unbiasedness*.
- A point estimator $\hat{\theta}$ is said to be an *unbiased estimator* of θ if $E(\hat{\theta}) = \theta$ for every possible value of θ . If $\hat{\theta}$ is not unbiased, the difference $E(\hat{\theta}) - \theta$ is called the *bias* of $\hat{\theta}$.

Example

Ex. Recall the unbiased coin example. Is the sample proportion an unbiased estimator of the population probability?

$$\text{estimator } \hat{p} = \frac{\text{number of heads}}{\text{number of flips}} = \frac{73}{100} = 0.73$$

What distribution does “number of heads” follow? What is its expectation?

General Result

- Proposition:

When X is a binomial rv with parameters n and p , the sample proportion $\hat{p} = X/n$ is an unbiased estimator of p .

Example

Ex. Suppose that X , the reaction time to a certain stimulus, has a uniform distribution on the interval from 0 to an unknown upper limit θ . It is desired to estimate θ based on random sample x_1, x_2, \dots, x_n . What is the best guess of θ ? Is your estimator unbiased?

General Result

- Proposition:

Let X_1, X_2, \dots, X_n be an i.i.d. sequence of random samples from a distribution with mean μ and variance σ^2 . Then the estimator

$$\hat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}$$

is an unbiased estimator of σ^2 .

General Result

- Proposition:

Let X_1, X_2, \dots, X_n be an i.i.d. sequence of random samples from a distribution with mean μ . Then the sample mean \bar{X} is an unbiased estimator of μ . If in addition the distribution is continuous and symmetric, then the sample median M and any trimmed mean are also unbiased estimators of μ .

MVUE

- For unbiased estimators, what are their MSE's?

$$E(\hat{\theta} - \theta)^2 = E(\hat{\theta} - E(\hat{\theta}))^2 = \text{Var}(\hat{\theta})$$

- Among all estimators of θ that are unbiased, choose the one that has minimum variance. The resulting $\hat{\theta}$ is called the **minimum variance unbiased estimator (MVUE)** of θ .
- One needs more knowledge to actually identify if some estimator is really MVUE. But in a special case, we have the following theorem.

Let X_1, X_2, \dots, X_n be an i.i.d. sequence of random samples from a **normal distribution** with mean μ and σ . Then the estimator $\hat{\mu} = \bar{X}$ is the **MVUE** for μ .

The Standard Error

- When reporting a point estimator, one also reports the **standard error** associated with it.
- The **standard error** of an estimator $\hat{\theta}$ is its standard deviation $\sigma_{\hat{\theta}} = \sqrt{\text{Var}(\hat{\theta})}$. If the standard error itself involves unknown parameters whose values can be estimated, substitution of these estimates into $\sigma_{\hat{\theta}}$ yields the **estimated standard error** of the estimator, which we denote as $\hat{\sigma}_{\hat{\theta}}$.
- The associated standard error gives us an idea of how good/accurate the estimators are.

Examples

Ex. What is the standard error of the sample proportion?

Ex. What is the standard error of the sample mean? Given that we know the variance.

Ex. What is the standard error of the sample mean? Given that we don't know the variance.

Methods of Point Estimation

- The definition of unbiasedness does not in general indicate how unbiased estimators can be derived.
- There are two commonly used “constructive” methods for obtaining point estimators: the [method of moments](#) and the [method of maximum likelihood](#).
- Although maximum likelihood estimators are generally preferable to moment estimators because of certain efficiency properties, they often require significantly more computation than do moment estimators.
- It is **NOT** guaranteed that these two methods would yield unbiased estimators.