# CLT

- Theorem:

**The Central Limit Theorem (CLT)**

Let $X_1$, $X_2$, …, $X_n$, be an i.i.d. sequence from a distribution with mean $\mu$ and variance $\sigma^2$. Then if n is sufficiently large, the sample mean $\bar{X}$ has approximately a normal distribution with $\mu_{\bar{X}} = \mu$ and $\sigma^2_{\bar{X}} = \sigma^2/n$;
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.

# Distribution of a Linear Combination

- Sample mean is a particular case of linear combinations.
- The expectation and variance of a general linear combination

$$a_1 X_1 + a_2 X_2 + \ldots + a_n X_n$$

  is given by the following result.

# A key result ***

Let $X_1, X_2, \ldots, X_n$, have mean values $\mu_1, \mu_2, \ldots, \mu_n$, respectively, and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$, respectively.

- Whether or not the Xi's are independent,

$$\begin{aligned} \mathrm{E}(a_1\mathrm{X}_1 + a_2\mathrm{X}_2 + \cdots + a_n\mathrm{X}_n) &= a_1\mathrm{E}(\mathrm{X}_1) + a_2\mathrm{E}(\mathrm{X}_2) + \cdots + a_n\mathrm{E}(\mathrm{X}_n) \\ &= a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n \end{aligned}$$

- For any $X_1, X_2, \ldots, X_n$,

$$\mathrm{Var}(a_1\mathrm{X}_1 + a_2\mathrm{X}_2 + \cdots + a_n\mathrm{X}_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mathrm{Cov}(\mathrm{X}_i, \mathrm{X}_j)$$

If they are independent, then

$$\begin{aligned} & \mathrm{Var}(a_1\mathrm{X}_1 + a_2\mathrm{X}_2 + \cdots + a_n\mathrm{X}_n) \\ = \ & a_1^2\mathrm{Var}(\mathrm{X}_1) + a_2^2\mathrm{Var}(\mathrm{X}_2) + \cdots + a_n^2\mathrm{Var}(\mathrm{X}_n) \\ = \ & a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2 \end{aligned}$$

# Special Cases

- $E(X+Y) = E(X) + E(Y)$;

- $E(X-Y) = E(X) - E(Y)$;

- $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

- $Var(X-Y) = Var(X) + Var(Y) -2Cov(X, Y)$

- If X and Y are independent, then $Cov(X, Y) = 0$, and
  $Var(X+Y) = Var(X) + Var(Y)$
  $Var(X - Y) = Var(X) + Var(Y)$

# Example

Ex. Show that if X ~ Bin($n$, $p$), then E(X) = $np$, and Var(X) = $np(1 - p)$.

Ex. Show that if X is a negative binomial rv with pmf $nb(x; r, p)$, then E(X) = $r(1-p)/p$,
Var(X) = $r(1 - p)/p^2$.

# Statistical Inference

- From the previous two examples, we know that quite often, we need to infer the truth (population) from some partial information (sample).

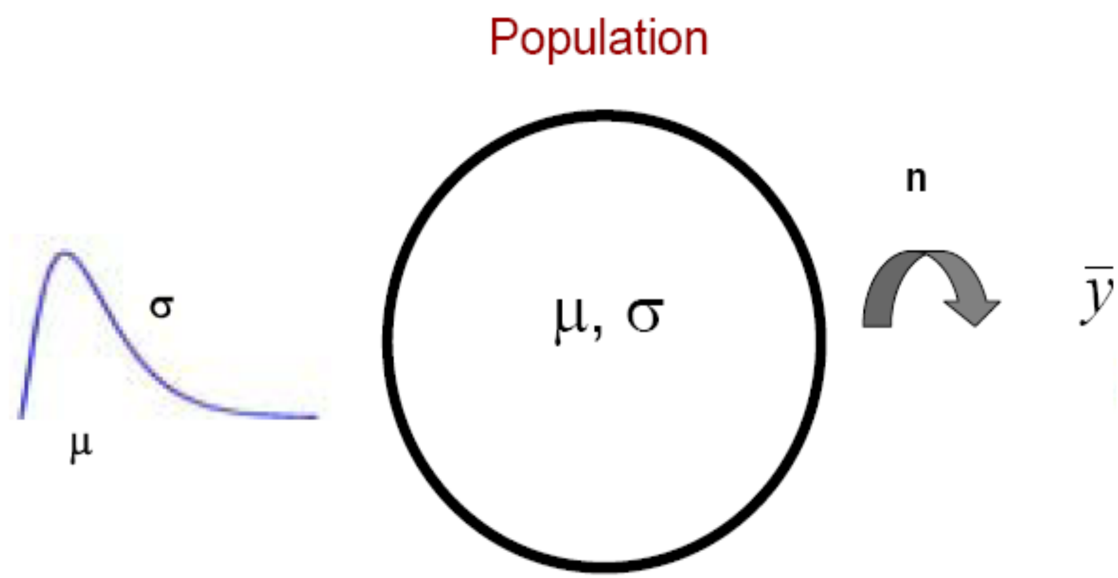
- Question: why do we need a model?

- 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 $\theta$ is a single number that can be regarded as a sensible value for $\theta$. 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 $\theta$.

# Sampling scheme for a Mean

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




- The actual sample observations $y_1$, $y_2$, …, $y_n$ (realizations) are assumed to be the result of a random sample $Y_1$, $Y_2$, …, $Y_n$ (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 $\text{estimator } \hat{p} = \frac{\text{number of heads}}{\text{number of flips}} = \frac{73}{100} = 0.73$

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

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

# 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 $\theta$, whereas for other samples $\hat{\theta}$ may underestimate $\theta$.

- 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.

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

- However, since MSE will generally depend on the value of $\theta$, 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 $\theta$ if $E(\hat{\theta}) = \theta$ for every possible value of $\theta$. If $\hat{\theta}$ is not unbiased, the difference $E(\hat{\theta}) - \theta$ is called the bias of $\hat{\theta}$.

# Example

<u>Ex.</u> 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 $p$.

# General Result

- Proposition:

Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sequence of random samples from a distribution with mean $\mu$ and variance $\sigma^2$. 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 $\sigma^2$.

# General Result

- Proposition:

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

# Estimator, Its Standard Error and Estimated Standard Error

- Now we are trying to estimate the probability of getting heads of a biased coin, so each flip $X_i$ is a Bernoulli RV with parameter $p$, the estimator of parameter $p$ is the sample mean/proportion

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$$

# Estimator, Its Standard Error and Estimated Standard Error

- Now we are trying to estimate the probability of getting heads of a biased coin, so each flip $X_i$ is a Bernoulli RV with parameter $p$, the estimator of parameter $p$ is the sample mean/proportion

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$$

- If we flipped 100 times and observed 75 heads, then our estimate of $p$ is

$$\hat{p} = \frac{75}{100} = 0.75$$

# Estimator, Its Standard Error and Estimated Standard Error

- Now we are trying to estimate the probability of getting heads of a biased coin, so each flip $X_i$ is a Bernoulli RV with parameter $p$, the estimator of parameter $p$ is the sample mean/proportion

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$$

- If we flipped 100 times and observed 75 heads, then our estimate of $p$ is

$$\hat{p} = \frac{75}{100} = 0.75$$

- Also, we need to report how good our estimator is through its Standard Error. This is also related to the Interval Estimation.

# Estimator, Its Standard Error and Estimated Standard Error

- The standard error is $Var(\hat{p}) = \frac{p(1-p)}{n}$, but we cannot report it since we don't know what $p$ is.
- So we can only report the estimated standard error of the estimator $\hat{p}$

$$\widehat{Var(\hat{p})} = \frac{\hat{p}(1-\hat{p})}{n}$$

# Another Example

- Now we have $X_1, X_2, \ldots, X_N$ IID with mean $\mu$ and variance $\sigma^2$, what's the estimator of $\mu$?

# Another Example

- Now we have $X_1, X_2, \ldots, X_N$ IID with mean $\mu$ and variance $\sigma^2$, what's the estimator of $\mu$?
- $$\hat{\mu} = \bar{X}$$

# Another Example

- Now we have $X_1, X_2, \ldots, X_N$ IID with mean $\mu$ and variance $\sigma^2$, what's the estimator of $\mu$?
- $$\hat{\mu} = \bar{X}$$
- The standard error of $\hat{\mu}$ is $\sqrt{Var(\hat{\mu})} = \frac{\sigma}{\sqrt{n}}$. Can we report $\frac{\sigma}{\sqrt{n}}$?

# Another Example

- Now we have $X_1, X_2, \ldots, X_N$ IID with mean $\mu$ and variance $\sigma^2$, what's the estimator of $\mu$?
- $$\hat{\mu} = \bar{X}$$
- The standard error of $\hat{\mu}$ is $\sqrt{Var(\hat{\mu})} = \frac{\sigma}{\sqrt{n}}$. Can we report $\frac{\sigma}{\sqrt{n}}$?
- It really depends on whether or not we know $\sigma$. If we know it, then we can report $\frac{\sigma}{\sqrt{n}}$; otherwise, we can only report $\frac{\hat{\sigma}}{\sqrt{n}}$.

# 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.

# Population Moment and Sample Moment

- Let $X_1, \ldots, X_n$ be a random sample from a pmf or pdf $f(x)$. For $k = 1, 2, \ldots$, the $k$th population moment is $E(X^k)$. The $k$th sample moment is $(1/n)\sum_{i=1}^{n} X_i^k$ .
- The essence of the Methods of Moment is to equate population moments with sample moments and solve the resulting equations.

# Moment Estimators

- Definition:

Let $X_1$, $X_2$, …, $X_n$ be an i.i.d. sample from a pmf or pdf $f(x)$. For $k$ = 1, 2, 3, …, the moment estimator for the $k$th population moment, is the $k$th sample moment, i.e.,

$$\widehat{\mathrm{E}(\mathrm{X}^k)} = \frac{\sum_{i=1}^{n} \mathrm{X}_i^k}{n}$$

# Example

Ex. Show that the sample proportion is the moment estimator of the population probability.

# Example

<u>Ex.</u> Let $X_1$, $X_2$, …, $X_n$ be an i.i.d. normal sample, and assume that the underlying normal distribution is $N(\mu,\sigma^2)$ where $\mu,\sigma^2$ are unknown. How can we construct moment estimators to estimate the two unknown parameters?

As we already know if $X \sim N(\mu,\sigma^2)$, then $E(X) = \mu$, and $E(X^2) = \mu^2+\sigma^2$.

Therefore, we have two equations:

$$\begin{cases} \hat{\mu} = \sum_{i=1}^n \mathrm{X}_i/n \\ \hat{\mu}^2 + \hat{\sigma}^2 = \sum_{i=1}^n \mathrm{X}_i^2/n \end{cases} \longrightarrow \begin{cases} \hat{\mu} = \sum_{i=1}^n \mathrm{X}_i/n \\ \hat{\sigma}^2 = \sum_{i=1}^n \mathrm{X}_i^2/n - \bar{\mathrm{X}}^2 \end{cases}$$

Is the variance estimator unbiased?

# Example

Ex. Let $X_1$, $X_2$, …, $X_n$ be an i.i.d. sample from exponential distribution with parameter $\lambda$ which is unknown. How do we estimate $\lambda$ using moment estimator?

As we already know if X ~ Exp($\lambda$), then E(X) = $1/\lambda$.

Thus, we have equation $1/\hat{\lambda} = \bar{\mathrm{X}} \rightarrow \hat{\lambda} = 1/\bar{\mathrm{X}}$.

Is this estimator unbiased?

# Maximum Likelihood Est.

- The method of maximum likelihood was first introduced by R.A. Fisher, a geneticist and statistician, in the 1920s. It is by far the most commonly used method to obtain estimators.

- Likelihood function is just another way of looking at the *joint pmf or the pdf*. In particular, let $X_1, X_2, \ldots, X_n$ (not necessarily i.i.d.) have joint pmf or pdf

  $$f(x_1, x_2, \ldots, x_n; \theta_1, \ldots, \theta_m)$$

  where $\theta_1, \ldots, \theta_m$ are parameters whose values are unknown. When $x_1, x_2, \ldots, x_n$ are the observed sample values and $f(.)$ is then regarded as a function of $\theta_1, \ldots, \theta_m$, it is called the likelihood function.

# Example

Ex. A biased coin has been flipped for 10 times. Let $X_1$, $X_2$, …, $X_{10}$ denote the outcomes of the coin flips. Assume the probability of having a head is *p* (parameter of interest), and the sample we observed is {0,1,1,0,0,0,1,0,0,0}. Write down the likelihood function for *p*.

$$f(x_1, x_2, \ldots, x_n; p) = f(x_1; p)\, f(x_2; p) \ldots f(x_n; p) = (1-p)\, p\, p\, (1-p) \ldots (1-p) = p^3(1-p)^7$$

Idea of Maximum Likelihood: can we find a *p* that can maximize the above function?

# MLE

- The maximum likelihood estimates (mle's) $\hat{\theta}_1, \ldots, \hat{\theta}_m$ are those values of $\theta_i$'s that maximize the likelihood function, so that

$$f(x_1, \ldots, x_n; \hat{\theta}_1, \ldots, \hat{\theta}_m) \geq f(x_1, \ldots, x_n; \theta_1, \ldots, \theta_m) \quad \text{for all } \theta_1, \ldots, \theta_m$$

  when the $X_i$'s are substituted in place of the $x_i$'s.

- Remark: the likelihood function tells us how likely the observed sample is as a function of the possible parameter values. Maximizing the likelihood gives the parameter values for which the observed sample is most likely to have been generated – that is, the parameter values that "agree most closely" with the observed data.

- In practice, in stead of maximizing the likelihood itself, people usually choose to maximize the log-likelihood function.

# Example

Ex. Let $X_1$, $X_2$, …, $X_n$ be an i.i.d. sample from exponential distribution with parameter $\lambda$ which is unknown. Write down the likelihood function for $\lambda$. What is the MLE of $\lambda$? Is the MLE unbiased?

Since we have an i.i.d. sample, it is easy to see that the likelihood function is a product of the individual pdf's:

$$f(x_1, \ldots, x_n; \lambda) = (\lambda e^{-\lambda x_1}) \cdot \cdots \cdot (\lambda e^{-\lambda x_n}) = \lambda^n e^{-\lambda \sum x_i}$$

$$\Downarrow$$

$$\log[f(x_1, \ldots, x_n; \lambda)] = n \log(\lambda) - \lambda \sum x_i$$

$$\Downarrow$$

$$\hat{\lambda} = n / \sum \mathrm{X}_i$$

# Example with Normal

- Let $X_1, X_2, \ldots, X_n$ be an IID sample from normal distribution with mean $\mu$ and variance $\sigma^2$, what is the likelihood function?

# Example with Normal

- Let $X_1, X_2, \ldots, X_n$ be an IID sample from normal distribution with mean $\mu$ and variance $\sigma^2$, what is the likelihood function?
- $$f(x_1, x_2, \ldots, x_n; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

  or in logarithm

  $$-\frac{n}{2}\log(2\pi\sigma^2) + \sum_{i=1}^{n}[-(x_i - \mu)^2/\sigma^2]$$

# Example with Normal

- Let $X_1, X_2, \ldots, X_n$ be an IID sample from normal distribution with mean $\mu$ and variance $\sigma^2$, what is the likelihood function?
- $$f(x_1, x_2, \ldots, x_n; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

  or in logarithm

  $$-\frac{n}{2}\log(2\pi\sigma^2) + \sum_{i=1}^{n}[-(x_i - \mu)^2/\sigma^2]$$
- Take derivative with respect to $\mu$ and $\sigma^2$ and solve the resulting equations

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

# Some complications

- The following is an example that MLE's can't be calculated analytically.

Ex. Let $X_1$, $X_2$, …, $X_n$ be an i.i.d. sample from Weibull distribution with parameters *α* and *β* and pdf

$$f(x;\alpha,\beta) = \begin{cases} \frac{\alpha}{\beta^\alpha} \cdot x^{\alpha-1} \cdot e^{-(x/\beta)^\alpha} & x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

by solving equations $\dfrac{\partial \log(f)}{\partial \alpha} = 0 \quad \dfrac{\partial \log(f)}{\partial \beta} = 0$

$$\downarrow$$

$$\hat{\alpha} = \left[\frac{\sum x_i^{\hat{\alpha}} \cdot \log(x_i)}{\sum x_i^{\hat{\alpha}}} - \frac{\sum \log(x_i)}{n}\right]^{-1} \quad \hat{\beta} = \left(\frac{\sum x_i^{\hat{\alpha}}}{n}\right)^{1/\hat{\alpha}}$$

# Some Complications

- Also, sometimes we cannot use calculus to get the MLE, such as when the density is not differentiable.
- Read Example 6.22 on textbook P.262.

# The Invariance Principle

- One of the nice features of MLE's is that, the MLE of a function of parameters, is the function of the MLE's of the parameters.

- More specifically, we have

> Let $\hat{\theta}_1, \ldots, \hat{\theta}_m$ be the MLE's of the parameters $\theta_1, \ldots, \theta_m$. Then the MLE of any function $h(\theta_1, \ldots, \theta_m)$ of these parameters is $h(\hat{\theta}_1, \ldots, \hat{\theta}_m)$.

Ex. In the normal example, what is the MLE of $\sigma$?

# Large Sample Behavior

- The following proposition says, for large samples, it is "optimal" to use MLE's, because it is asymptotically unbiased and has the minimal variance among all unbiased estimators.

- Proposition:

  Under very general conditions on the joint distribution of the sample,
  When the sample size n is large, the maximum likelihood estimator is
  Approximately the MVUE of the parameter.

# Confidence Intervals

- A point estimate, because it is a single number, by itself provides no information about the precision and reliability of estimation (the reason why we need standard error).

- An alternative to reporting a single sensible value for the parameter being estimated is to calculate and report an entire interval of plausible values – an *interval estimate* or *confidence interval* (*CI*).

- A confidence interval is always calculated by first selecting a *confidence level*, which is a measure of the degree of reliability of the interval.

- Construct a confidence interval for a standard normal random variable.

# Illustration

- Let's first consider a simple, somewhat unrealistic problem situation.
    1. We are interested in the population mean parameter $\mu$.
    2. The population distribution is normal.
    3. The value of the population standard deviation $\sigma$ is known. (unlikely!)

- Suppose we have a random sample $X_1, X_2, \ldots, X_n$ from a normal distribution with mean value $\mu$ and standard deviation $\sigma$. As we know, $\bar{X}$ also follows a normal distribution with mean value $\mu$ and standard deviation $\sigma/\sqrt{n}$. Thus, we could get a standard normal distribution by normalizing $\bar{X}$.

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

# Construction

- The smallest interval that contains 95% of the possible outcomes of Z is (-1.96, 1.96).

$$-1.96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1.96$$

$$\Downarrow$$

$$-1.96 \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Downarrow$$

$$\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

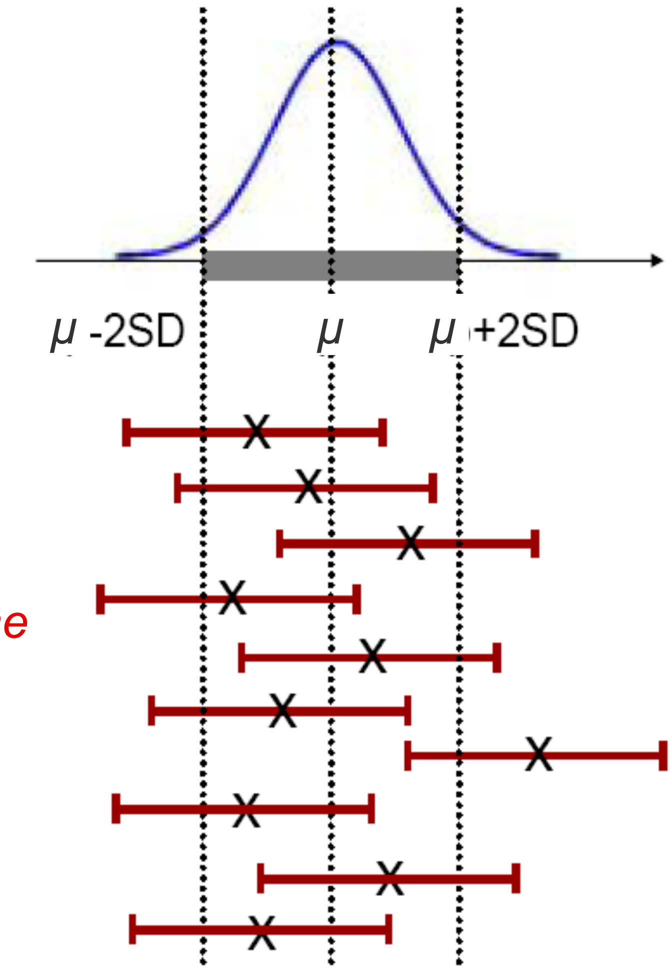
# Interpretation

- Thus we have $\mathrm{P}\left(\bar{\mathrm{X}} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{\mathrm{X}} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0.95$.
- Some people interpreted this as: the true parameter *μ* has 95% chance of falling in the interval of $(\bar{\mathrm{X}} - 1.96 \cdot \sigma/\sqrt{n}, \bar{\mathrm{X}} + 1.96 \cdot \sigma/\sqrt{n})$. Is it right?
- In fact, the two boundaries of the interval given above are random! Thus every time we sample n observations from the same population, we will get a different confidence interval!

# Random Interval

- By constructing a confidence interval like this, we never be sure whether *μ* actually lies in our confidence interval. However, we know that about 95 out of 100 times intervals constructed using this method will capture the true parameter.

- Interpreted as: "*the probability is .95 that the random interval includes or covers the true value of μ*."



# Confidence Interval

- Definition:

> A 100(1-$\alpha$)% confidence interval for the mean $\mu$ of a normal population when the value of $\sigma$ is known is given by
>
> $$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

- $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of a standard normal distribution, i.e., $P(Z>z_{\alpha/2})=\alpha/2$.

# Remarks

- When constructing a confidence interval, *confidence level*, *precision*, and *sample size* are closely related. Is there a finite 100% confidence interval?

- The precision, or the width of the confidence interval when $\sigma$ is known is, $2z_{\alpha/2}\sigma/\sqrt{n}$. Thus we can see, the confidence level of the interval is inversely related to its precision.

- The precision is also inversely related to the sample size.

- An appealing strategy is to specify both the desired confidence level and interval width and then determine the necessary sample size.

# Sample Size Calculation

- The general formula for the sample size *n* necessary to ensure an interval width *w* is obtained from $w = 2 \cdot z_{\alpha/2} \cdot \sigma/\sqrt{n}$.

$$n = \left(2 \cdot z_{\alpha/2} \cdot \frac{\sigma}{w}\right)^2$$

Ex. A new operating system has been installed, and we wish to estimate the true average response time $\mu$ to a particular editing command. Assuming that response times are normally distributed with $\sigma$=25 millisec. How many tests should we do to ensure that the resulting 95% CI has a width of at most 10?

# Non-normal and Unknown Variance

- Previously we constructed a confidence interval for normal population mean with known variance. The next question would then be, what if we don't have normality and what if we don't know the underlying variance?

- If we have large enough sample size, the celebrated CLT can help us construct a confidence interval for the mean parameter of a population with unknown distribution and unknown variance. Consider the following quantity

$$\begin{tikzcd}
\dfrac{\bar{X}-\mu}{\hat{\sigma}/\sqrt{n}} = & \boxed{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}} \arrow[d, "\text{CLT}"'] & \cdot & \boxed{\dfrac{\sigma}{\hat{\sigma}}} \arrow[d, "\text{LLN}"] \\
& \boxed{N(0,1)} & & \boxed{1}
\end{tikzcd}$$

# General Results

- Proposition:

  A 100(1-$\alpha$)% confidence interval for the mean $\mu$ of any population when the value of $\sigma$ is unknown and sample size n is sufficiently large is given by

  $$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\hat{\sigma}}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\hat{\sigma}}{\sqrt{n}}\right)$$

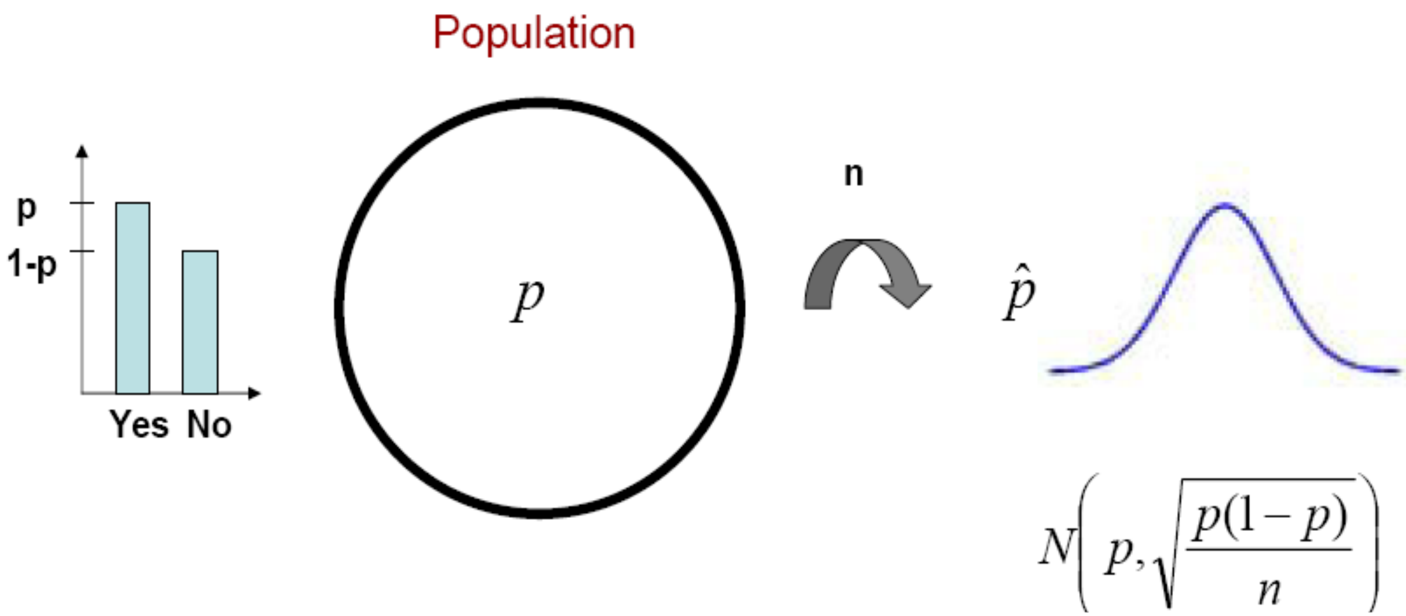
- Rule of Thumb: generally speaking, n>40 will be sufficient to justify the use of this interval. This is somewhat more conservative than the rule of thumb for the CLT, because of the additional randomness coming from $\hat{\sigma}$ .

- One can also derive a similar sample size calculation formula in this case

  $$n = \left(2 \cdot z_{\alpha/2} \cdot \frac{\hat{\sigma}}{w}\right)^2$$

# Proportions

- A special case of non-normal population is Bernoulli population. And the parameter of interest is the population proportion *p*.



# Large Sample CI

- One can directly apply the proposition from the large sample case to construct the CI for the population proportion *p*.

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\hat{\sigma}}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\hat{\sigma}}{\sqrt{n}}\right)$$

- In this case $\bar{x} = \hat{p}$ , $\hat{\sigma}^2 = \hat{p}(1-\hat{p})$.

- If we set $q$=1-$p$, then the large sample confidence interval for $p$ should be

$$\left(\hat{p} - z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}, \hat{p} + z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}\right)$$

- To calculate sample size: $n = \left(2 \cdot z_{\alpha/2} \cdot \frac{\sqrt{\hat{p}\hat{q}}}{w}\right)^2$

# Another way

- The large sample confidence interval works fine if we have enough data. But for finite samples we can construct a better CI.

- Since in this case, we only have 1 parameter *p*, by CLT, we have

$$\mathrm{P}\left(-z_{\alpha/2} < \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} < z_{\alpha/2}\right) \approx 1-\alpha$$

- If we solve the resulting quadratic function, we'll have a new confidence interval for *p*.

$$\left(\frac{\hat{p}+\frac{z_{\alpha/2}^2}{2n}-z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}+\frac{z_{\alpha/2}^2}{4n^2}}}{1+z_{\alpha/2}^2/n}, \frac{\hat{p}+\frac{z_{\alpha/2}^2}{2n}+z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}+\frac{z_{\alpha/2}^2}{4n^2}}}{1+z_{\alpha/2}^2/n}\right)$$

# Remarks

- The latter confidence interval looks complicated, but it “can be recommended for use with nearly all sample sizes and parameter values”. Therefore we don’t have to check for large sample conditions.

- In the latter case, we can also derive a new sample size calculation formula

$$n = \frac{2z_{\alpha/2}^2\hat{p}\hat{q} - z_{\alpha/2}^2 w^2 \pm \sqrt{4z_{\alpha/2}^4\hat{p}\hat{q}(\hat{p}\hat{q} - w^2) + w^2 z_{\alpha/2}^4}}{w^2}$$

  “+” sign is used!

- When sample size is large, the confidence interval we just constructed and the sample size calculation formula will be equivalent to

  $\left(\hat{p} - z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}, \hat{p} + z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}\right)$ and $n = \left(2 \cdot z_{\alpha/2} \cdot \frac{\sqrt{\hat{p}\hat{q}}}{w}\right)^2$

# One-sided CI

- In some situations, an investigator will want only one upper bound or one lower bound for the parameter.

- Follow a similar argument as in the two-sided case, we have the following result

A large sample 100(1-$\alpha$)% confidence upper bound for the mean $\mu$ is

$$\mu < \bar{x} + z_\alpha \cdot \frac{\hat{\sigma}}{\sqrt{n}}$$

and a lower bound is

$$\mu > \bar{x} - z_\alpha \cdot \frac{\hat{\sigma}}{\sqrt{n}}$$

A one-sided confidence bound for $p$ results from replacing $z_{\alpha/2}$ by $z_\alpha$.

# Constructing a CI

- The previous examples show the general procedure of constructing confidence intervals. Suppose $X_1$, $X_2$, …, $X_n$ are the sample on which the CI for a parameter $\theta$ is to be based. Then we construct a so-called "pivotal" quantity whose distribution does not depend on parameters.

- In other words, the pivotal quantity is a function of both samples and parameters, i.e., $h(X_1, X_2, \ldots, X_n, \theta)$, and the distribution of $h(\cdot)$ does not depend on $\theta$ or any other unknowns.

- Then one can find *a* and *b* to satisfy $\mathrm{P}(a < h(\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n; \theta) < b) = 1 - \alpha$ , by the pivotal property, *a* and *b* do not depend on $\theta$. Then the inequality can be manipulated to isolate $\theta$, giving the equivalent probability statement
$$\mathrm{P}(l(\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n) < \theta < u(\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_n)) = 1 - \alpha$$