Probability and Statistics
for Final Year Engineering Students

By Yoni Nazarathy, Last Updated: May 1, 2011.

Lecture 3:
The Basics of Statistical Inference:
Point Estimation, Confidence Intervals and Hypothesis Testing

Point Estimation:

The idea of "point estimation" is to find a statistic (a function of the random sample) which can serve as a "good" estimator for some population parameter. Examples are $\bar{X}$ estimating the population mean, S^2 estimating the population variance and $\hat{p} = \frac{\sum_{i=1}^{n} I_i}{n}$ estimating $P(I_i = 1)$. In these cases, "design" of the statistic/estimator is sensible and simple, yet there are other more advanced cases in which design of a "good" estimator is not immediately trivial.

Much of the mathematical theory of statistical inference deals with properties of estimators. We only briefly touch upon this subject here.

Let $\hat{\theta}$ represent some estimator (statistic) for some parameter of the population denoted by θ. (E.g. $\hat{\theta} = S^2$ and $\theta = \sigma^2$, the variance of elements of the population). Remember that $\hat{\theta}$ is a function of the sample of size n, so we sometimes denote $\hat{\theta}_n$. Here are three types of properties which are of interest:

- We say $\hat{\theta}$ is **unbiased** if $E[\hat{\theta}] = \theta$.
- We say $\hat{\theta}$ is **asymptotically unbiased** if $\lim_{n \to \infty} E[\hat{\theta}_n] = \theta$. Note that any unbiased estimator is asymptotically unbiased.
- We say $\hat{\theta}$ is **consistent** if $\lim_{n \to \infty} \hat{\theta}_n = \theta$. Note: In this course we really didn't define a limit of random variables, yet in case of an unbiased estimator it is consistent if $\lim_{n \to \infty} Var(\hat{\theta}_n) = 0$.

In tutorial 2 we have shown that $\bar{X}$ and $\hat{p}$ are unbiased and consistent estimators.

Here is the calculation for $\hat{p}$ again (denote $P(I_i = 1) = p$):

Unbiased:

$$E[\hat{p}] = E\left[\frac{\sum_{i=1}^{n} I_i}{n}\right] = \frac{1}{n} E\left[\sum_{i=1}^{n} I_i\right] = \frac{1}{n} \sum_{i=1}^{n} E[I_i] = \frac{1}{n} \sum_{i=1}^{n} p = \frac{1}{n} np = p.$$

Consistent:

$$Var(\hat{p}) = Var\left(\frac{\sum_{i=1}^{n} I_i}{n}\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} I_i\right) = \frac{1}{n^2} \sum_{i=1}^{n} Var(I_i) = \frac{1}{n^2} \sum_{i=1}^{n} p(1-p) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}.$$

Thus: $\lim_{n \to \infty} Var(\hat{p}_n) = 0$.

We shall now investigate the estimator S^2 to see if it is unbiased (this will explain why there is $n-1$ in the denominator). First lets present the "working formula" for S^2:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i{}^2 - 2X_i\bar{X} + \bar{X}^2\right) = \frac{\sum_{i=1}^{n}X_i{}^2 - 2n\bar{X}\bar{X} + n\bar{X}^2}{n-1} = \frac{\sum_{i=1}^{n}X_i{}^2 - n\bar{X}^2}{n-1}.$$

Now,

$$\frac{(n-1)}{n}E[S^2] = \frac{1}{n}E\left[\sum_{i=1}^{n}X_i{}^2 - n\bar{X}^2\right] = E[X_i{}^2] - E[\bar{X}^2] = E[X_i{}^2] - E[X_i]^2 + E[X_i]^2 - E[\bar{X}^2]$$

$$= Var(X) - (E[\bar{X}^2] - E[\bar{X}]^2) = Var(X) - \frac{Var(X)}{n} = Var(X)(1 - \frac{1}{n}) = \frac{(n-1)}{n}Var(X).$$

So, $E[S^2] = Var(X)$.

Note that if we use $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ as estimator for the variance then it is biased yet asymptotically unbiased:

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \frac{(n-1)}{n}Var(X).$$

Percentiles (used in confidence intervals and hypothesis testing):

Denote by z_x the x'th **percentile** (also called **quantile)** of the normal distribution. This can be written in several ways:

$$P(Z \leq z_x) = x,$$

or

$$\int_{-\infty}^{z_x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = x,$$

or,

$$\phi(z_x) = x,$$

or,

$$z_x = \phi^{-1}(x).$$

You can "read-off" z_x from the normal distribution table. Here are a few typically used values:

$$z_{.95} = 1.645$$

$$z_{.975} = 1.96$$

$$z_{.995} = 2.576$$

$z_{.9995} = 3.291.$

Confidence Intervals:

For the rest of this lecture we will concentrate on the estimation of a proportion. Our statistic is $\hat{p} = \frac{\sum_{i=1}^{n} I_i}{n}$ and we are interested in estimating p (the population proportion).

We can use the Binomial distribution to get the exact distribution of $\hat{p}$, yet we shall use the central limit theorem to approximate this distribution by a Normal(p, $\left(\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right)^2$) distribution.

Denote now, $Z = \frac{\hat{p}-p}{\sqrt{p(1-p)}/\sqrt{n}}$. The random variable Z is approximately a standard Normal random variable.

Let us pick now a "small" α. Say $\alpha = 0.01$ and go back to the standard normal random variable Z, we have:

$$P(z_{\alpha/2} \le Z \le z_{1-\alpha/2}) = 1 - \alpha.$$

Note also that since the normal distribution is symmetric about 0, $z_x = -z_{1-x}$, so:

$$P(-z_{1-\alpha/2} \le Z \le z_{1-\alpha/2}) = 1 - \alpha.$$

We can now plug in $Z = \frac{\hat{p}-p}{\sqrt{p(1-p)}/\sqrt{n}}$,

$$P\left(-z_{1-\alpha/2} \le \frac{\hat{p}-p}{\sqrt{p(1-p)}/\sqrt{n}} \le z_{1-\alpha/2}\right) = 1 - \alpha.$$

And rearrange the inequality to obtain

$$P\left(\hat{p} - z_{1-\frac{\alpha}{2}}\frac{\sqrt{p(1-p)}}{\sqrt{n}} \le p \le \hat{p} + z_{1-\frac{\alpha}{2}}\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) = 1 - \alpha.$$

Note that the above equation is only approximately true (we have used the CLT to assume a standard normal distribution for Z). We now make a second approximation: we replace the p's in the right and left hand side of the inequality by their estimators $\hat{p}$, to obtain:

$$P\left(\hat{p} - z_{1-\frac{\alpha}{2}}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} \le p \le \hat{p} + z_{1-\frac{\alpha}{2}}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right) = 1 - \alpha.$$

The above is called a **confidence interval** and can be written as:

$$\boxed{p \text{ is in the range } \hat{p} \pm z_{1-\frac{\alpha}{2}} \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} \text{ with probability } 1-\alpha.}$$

The value $1-\alpha$ indicates the **confidence level** that we have with respect to the fact that p *is in the range* $\hat{p} \pm z_{1-\frac{\alpha}{2}} \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$. That is, we know that there is (approximately) a $1-\alpha$ chance that this is indeed the "correct range".

Note: The approximation is due to the CLT and due to the fact we replaced p with $\hat{p}$ in the inequality.

Example: In the robotic arm case assume that we sample n=100 suction pick-up attempts and get $\sum_{i=1}^{n} I_i = 67$. Our estimate for the success probability is 0.67. If we choose $\alpha = 0.05$, thus having a confidence level of 0.95 we have a confidence interval:

$$0.67 \pm 1.96 \frac{\sqrt{0.67(1-0.67)}}{\sqrt{100}}, \text{ or } 0.67 \pm 0.092.$$

We can thus say **with a 95% confidence** that the suction success proportion is somewhere in the range [0.578, 0.762]. Note that there still remains a 5% chance that we were "off the mark" and the actual proportion is outside of this range.

Note: The concept of confidence intervals is applied in all kinds of estimation problems (not just the sample proportion). We do not cover further details of this in this course, yet note that the concepts are similar and you should be able to understand how to use "confidence interval formulas" having understood the concepts outlined here. For illustration we present the following table scanned from page 439 of [LeGar08] (this is not part of the material):

Summary of confidence intervals for Gaussian and non-Gaussian random variables.

Case	Confidence Interval
Gaussian random variable, σ^2 known	$\left[\overline{X}_n - z_{\alpha/2}\sigma/\sqrt{n}, \overline{X}_n + z_{\alpha/2}\sigma/\sqrt{n}\right]$
Non-Gaussian random variable, n large, σ^2 known	$\left[\overline{X}_n - z_{\alpha/2}\sigma/\sqrt{n}, \overline{X}_n + z_{\alpha/2}\sigma/\sqrt{n}\right]$
Gaussian random variable, σ^2 unknown	$\left[\overline{X}_n - t_{\alpha/2,n-1}\hat{\sigma}_n/\sqrt{n}, \overline{X}_n + t_{\alpha/2,n-1}\hat{\sigma}_n/\sqrt{n}\right]$
Non-Gaussian random variable, σ^2 unknown, batch means	$\left[\overline{X}_n - t_{\alpha/2,n-1}\hat{\sigma}_n/\sqrt{n}, \overline{X}_n + t_{\alpha/2,n-1}\hat{\sigma}_n/\sqrt{n}\right]$
Gaussian random variable, μ unknown	$\left[\dfrac{(n-1)\hat{\sigma}_n^2}{\chi_{\alpha/2,n-1}^2}, \dfrac{(n-1)\hat{\sigma}_n^2}{\chi_{1-\alpha/2,n-1}^2}\right]$

Note that in the above table, t_x and χ_x^2 represent percentiles from "student-t" and "chi-squared" distributions. These distributions are related to the normal distribution and often appear in basic statistical inference. We do not discuss these further.

The following is useful: http://demonstrations.wolfram.com/ConfidenceIntervalsForAMean/

Sample Size Planning:

The confidence interval for the proportion ($\hat{p} \pm z_{1-\frac{\alpha}{2}} \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$) shows the relation between sample size, confidence level and the width of the confidence interval. As the confidence level increases (α decreases), $z_{1-\frac{\alpha}{2}}$ increases and thus the confidence interval becomes wider. As the sample size increases, the confidence interval shortens.

Note that we do not know the term $\hat{p}(1-\hat{p})$, until samples have been taken but also note that,

$$\hat{p}(1-\hat{p}) \leq \tfrac{1}{4}.$$

(show this). We now denote $\varepsilon = z_{1-\frac{\alpha}{2}} \frac{\sqrt{1/4}}{\sqrt{n}}$. Here we were "conservative" by using the upper bound on $\hat{p}(1-\hat{p})$. Solving for ε we have:

$$n^* = \left\lceil \frac{\left(z_{1-\frac{\alpha}{2}}\right)^2}{4\varepsilon^2} \right\rceil \quad \text{(these brackets denote "rounding up").}$$

We can interpret the above as: **If $\widehat{p}$ is used to estimate p, we can be at least 1-α confident that the error be less than ε if we choose a sample size n^*.**

Example: Assume that we are using Monte-Carlo simulations to simulate the structural properties of our robotic arm after 10 years of operation. A Monte-Carlo simulation of the robotic arm needs to run for several hours on a super-computer, generating a "0" or a "1" indicating if the robotic arm's metal structure is "still usable" or "requires some replacement" after 10 years. We want to see what proportion of robotic arms will need to be replaced after 10 years. Assume we would like to be accurate up to a level of $\varepsilon = 0.1$ with a confidence level of 1-α = 99%. How many simulation runs do we need?

Answer, we have $z_{.995} = 2.576$, plugging in the above formula we have

$$n^* = \left\lceil \frac{(2.576)^2}{40.1^2} \right\rceil = \lceil 64.4 \rceil = 65.$$

Hypothesis Testing Basics:

One of the purposes of statistical analysis of data is to verify scientific and/or engineering **hypothesis**. Here are some examples:

- Hypothesis: The *mean* weight of items in the warehouse is greater than 20Kg.
- Hypothesis: The *proportion* of successful robotic-arm pick-ups using suction is bigger than that using grabbing.
- Hypothesis: The *variance* of the grades of students of group A is smaller than that of group B.
- Hypothesis: The *distribution* of measurement errors is uniform.
- Hypothesis: Times between arrivals of customers to a super-market are *independent*.

The process of **hypothesis testing** is a statistical procedure for using data (from the sample) in-order to check the validity of hypothesis. Our purpose here is to briefly the basic concepts only.

The ingredients of a hypothesis testing procedure are the following:

- The **parameter space** Θ is the set of values which the parameter(s) under investigation can take.
- The parameter space is partitioned into the **Null Hypothesis** (often denoted H_0) and the **Alternative Hypothesis** (often denoted H_1). It is often the case that H_0 denotes values of the parameter which are known prior to "some new treatment" and H_1 are values of the parameter which are scientifically assumed and are now to be demonstrated by means of the hypothesis testing procedure:

  Example: Assume that it is known that the chance of a heart-attack (say during the ages of 50-70) is 0.23. Now a medical company wants to show that some medicine reduces the chance for heart attack, then $\Theta = [0,1]$ $H_0 = [0.23,1]$ and $H_1 = [0,0.23)$.

- A **test statistic** is a random value calculated from the sample.
- A **decision rule** is a rule for deciding if to select H_0 or H_1 based on the **test statistic**. I.e. it is a rule for choosing a random outcome (random since the statistic is a function of the random sample). Selecting H_1 is often called **rejecting H_0.** This is often viewed as the "desired effect" of the Hypothesis testing procedure. Selecting H_0 is often denoted as **not rejecting H_0** (or sometimes **failing to reject).** The decision rule is typically in the form of a **critical value** or **critical region**: When the test statistic is great/smaller than the critical value one rejects H_0 otherwise, one does not reject H_0.
- The probability of **type-I error**, denoted $\boldsymbol{\alpha}$ is the probability of rejecting H_0 when H_0 actually holds (false-positive). This is also sometimes called the **significance level** of the test. This probability is typically set to one of the values 0.1, 0.05, 0.01, 0.001.

  Example: Continuing the anti-heart-attack drug example from above, it is most sensible to use the sample proportion as the test statistic. The decision rule may be something of the sort: *Reject H_0 if $\hat{p} < 0.19$ otherwise do not reject.* Now the type-I error would be the probability of

getting $\hat{p} < 0.19$ even when the drug has no-effect (i.e. p=0.23). In clinical testing a very small type-I error is often used (e.g. 10^-4) why?

- The probability of **type-II error**, denoted β is the probability of not rejecting H_0 when H_1 actually holds (false negative). The complement 1- β is called the **power of the test**. One would like to have the power as close to 1 as possible, but tests are typically done by selecting a significance level (α), leaving the power to be a function of the sample size (an increasing function).

The best way to understand the mathematics of hypothesis testing is by looking at a **simple hypothesis** test. A test is called simple if the number of elements in the parameter space, Θ, is 2. Such cases situations are often less realistic, but are useful for both understanding the basics of hypothesis testing and for developing further mathematical theory (not done here).

Example (Simple hypothesis – good for theory – not so practical): Assume that certain types of screws are manufactured by two manufacturers. The screws made by both manufacturers are indistinguishable in appearance, weight and other external physical properties. Nevertheless, screws made by manufacturer 0 have a chance of $p_0 = 0.5$ of breaking under a certain load test while screws made by manufacturer 1 have a chance $p_1 = 0.3$ of breaking. We are given a container with tens of thousands of screws all made by a single manufacturer (we don't know if 0 or 1). We set out to determine if the container is from manufacturer 0 or 1 by estimating the proportion of screws breaking under the load. We sample n=40 screws, perform the load test on each screw and obtain $\hat{p}$.

Assume that we set this as an hypothesis test with $H_0 = p_0$ and $H_1 = p_1$. Assume further that we make the following decision rule: If $\hat{p} < 0.38$ reject H_0, otherwise do not reject H_0. Using the CLT to approximate the binomial distribution, this test can be illustrated as follows:
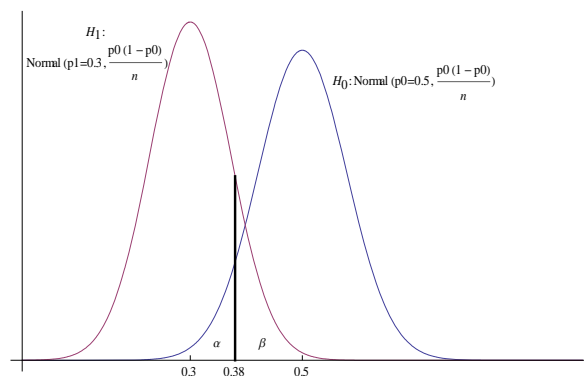
Notice the tradeoff between α and β. We can now calculate α and β and the power of the test. What happens when we increase the sample size n?

Notice that in this setting the null hypothesis and alternative hypothesis appear to be treated in the same manner: the location of the critical value determines our probabilities for false-positive and false-

negative. As opposed to that, we can take the following approach: *Our manufacturer is typically manufacturer 0, so that is why we have this is as the null hypothesis.* We now wish to set $\alpha = 0.05$. The critical value and the power of the test will follow as a result. What would be the critical value? What is the power of the test?

Hypothesis Testing For Proportions:

Moving onto non-simple hypothesis we now present the test for the population proportion. Here is a motivating example: Consider the robotic arm example (we are beta-testing a new robotic arm which picks items up). For illustration, assume that the market already has a robotic arm in use (an older generation) which is known to succeed in picking up items on the first attempt with probability p=0.55. We believe that our robotic arm is better and can get $0.55 < p$.

We formulate this as an hypothesis test (called the **one-sided hypothesis test for a proportion**):

$$H_0 : p \leq 0.55$$

$$H_1 : p > 0.55$$

The test statistic is as in the confidence interval $Z = \frac{\hat{p}-0.55}{\sqrt{p(1-p)}/\sqrt{n}}$. Note that if H_0 is true, Z is a standard

normal random variable (approximately due to the CLT). The decision rule is thus to reject H_0 if Z is "big". How Big? We choose $z_{1-\alpha}$ as a **critical value** and reject H_0 if Z>$z_{1-\alpha}$. This means that the significance level (type I error probability) of the test is α.

Note that an alternative **one-sided** test is of the form:

$$H_0 : p \geq 0.55$$

$$H_1 : p < 0.55$$

In this case the critical value is z_α and we would reject H_0 if Z<z_α.

There is also the **two-sided hypothesis test for a proportion.** In this case:

$$H_0 : p = 0.55$$

$$H_1 : p \neq 0.55$$

Now the rejection region is when Z is "far from 0" in either direction. Reject if if Z>$z_{1-\alpha/2}$ or Z<$z_{\alpha/2}$.

An alternative way of stating the result of an hypothesis test is by means of a **p-value**. The p-value is the smallest confidence level for which the test is not rejected.

The following demonstration may be helpful:
http://demonstrations.wolfram.com/HypothesisTestsAboutAPopulationMean/