Part III: Introduction to Statistical Inference

Yi Luo

May 21, 2015

1 Parameters and Estimators

Suppose a random variable $X$ follows a model with some parameter $p$. In order to obtain an estimate of $p$, we observe $X$ $n$ times. Denote these observations as $X_1, X_2, \ldots, X_n$ which are also random variables and usually assumed to be independent and identically distributed (i.i.d.) as $X$. Let the actual observed data now be $x_1, x_2, \ldots, x_n$.

An estimator $\hat{p}$ for the parameter $p$ is a random variable that is calculated from $X_1, X_2, \ldots, X_n$, which is intended to approximate $p$. It can be seen as a function of observations $\hat{p} = \hat{p}(X_1, X_2, \ldots, X_n)$ and thereby the actual observed value now is $\hat{p}(x_1, x_2, \ldots, x_n)$.

$\hat{p}$ is said to be unbiased if $E(\hat{p}) = p$. When comparing two unbiased estimators, the one with a smaller variance is better, or said to be more efficient.

2 Sampling Distributions of $\hat{p} = \bar{X}_n$

Suppose the parameter $p = \mathbb{E}(X)$. Let $\sigma = \text{SD}(X)$ and the unbiased estimator for $p$

$$\hat{p} \triangleq \bar{X}_n \triangleq \frac{X_1 + X_2 + \cdots + X_n}{n}. \quad (2)$$

By the Central Limit Theorem, when $n$ is large enough,

$$\hat{p} \approx N\left(p, \text{SD}(\hat{p}) = \frac{\sigma}{\sqrt{n}}\right), \quad (3)$$

The gathering of $n$ random samples without replacement from a total population of cases is called a survey. A census is a survey of the entire population. If the population is finite, then $X_1, X_2, \ldots, X_n$ can not be i.i.d., since we sampled without replacement. However if the sample size is less than 10% of the population, the i.i.d. assumption is still reasonable.

If Independence Assumption, Randomization Condition, 10% Condition and $n$ Large Enough Condition are met, $X_1, X_2, \ldots, X_n$ can be assumed to be i.i.d., and the sampling distribution of $\hat{p}$ is roughly normal with mean $p$ and standard deviation $\sigma/\sqrt{n}$, which is (3).
3 Confidence Intervals for $p = \mathbb{E}(X)$

After we get an estimator $\hat{p}$ for the unknown parameter $p$, we would like to find a $\delta > 0$ such that

$$P(\hat{p} - \delta \leq p \leq \hat{p} + \delta) = 1 - \alpha. \quad (4)$$

(4) can be interpreted as we are $100 \cdot (1 - \alpha)$ percent confident that $p$ lies in $[\hat{p} - \delta, \hat{p} + \delta]$, which is called the confidence interval for $p$ with confidence level $1 - \alpha$. $\alpha$ is called the significance level. $\delta$ is called the margin of error:

$$\text{ME}(\hat{p}) \triangleq \delta. \quad (5)$$

3.1 One-sample $z$-interval for the Mean

1. Verify $p = \mathbb{E}(X)$ and $\hat{p} = \bar{X}_n = (X_1 + X_2 + \cdots + X_n)/n$.

2. If the model of $X$ has only one parameter, write $\sigma$ as a function of $p$:

$$\sigma = \sigma(p). \quad (6)$$

3. A standard error of $\hat{p}$, denoted as $\text{SE}(\hat{p})$, is an estimate for $\text{SD}(\hat{p})$. To this end, we replace $p$ with $\hat{p}$ in the expression of $\text{SD}(\hat{p})$:

$$\text{SD}(\hat{p}) = \frac{\sigma}{\sqrt{n}} = \frac{\sigma(p)}{\sqrt{n}} \Rightarrow \text{SE}(\hat{p}) = \frac{\sigma(\hat{p})}{\sqrt{n}}. \quad (7)$$

Then by (3) and (7),

$$\frac{\hat{p} - p}{\text{SE}(\hat{p})} \approx Z \sim N(0, 1). \quad (8)$$

4. The critical value with significance level $\alpha$ is the $z^* > 0$ that satisfies

$$P(-z^* \leq Z \leq z^*) = 1 - \alpha. \quad (9)$$

It can then be shown that

$$\text{ME}(\hat{p}) \approx z^* \cdot \text{SE}(\hat{p}). \quad (10)$$

5. The $100 \cdot (1 - \alpha)$ percent confidence interval for the parameter $p$ is

$$[\hat{p} - \text{ME}(\hat{p}), \hat{p} + \text{ME}(\hat{p})]. \quad (11)$$
3.2 One-Sample $t$-interval for the Mean

If the model of $X$ has multiple parameters, (6) fails, and even if (6) holds, when $n$ is not large enough, it turns out that (8) introduces too big an error. In such scenarios, the $t$-distribution with a parameter known as degree of freedom (df) may be useful.

1. Verify $p = E(X)$ and $\hat{p} = \bar{X}_n = (X_1 + X_2 + \cdots + X_n)/n$.

2. Instead of (6) we calculate

$$s_X = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}{n-1}}. \quad (12)$$

3. We replace $\sigma$ with $s_X$ to get the standard error

$$SD(\hat{p}) = \frac{\sigma}{\sqrt{n}} \Rightarrow SE(\hat{p}) = \frac{s_X}{\sqrt{n}}. \quad (13)$$

If the Nearly Normal Condition is met, then

$$\frac{\hat{p} - p}{SE(\hat{p})} \approx T, \quad (14)$$

where $T$ is a $t$-distribution with $n - 1$ degrees of freedom.

4. Redefine the critical value for the $t$-distribution:

$$P(-t_{n-1}^* \leq T \leq t_{n-1}^*) = 1 - \alpha, \quad (15)$$

and similar to (10), we have

$$ME(\hat{p}) \approx t_{n-1}^* \cdot SE(\hat{p}). \quad (16)$$

5. The $100 \cdot (1 - \alpha)$ confidence interval expression (11) does not change:

$$[\hat{p} - ME(\hat{p}), \hat{p} + ME(\hat{p})].$$

3.3 Confidence Intervals for Functions of the Mean

If the function $f$ is strictly increasing, then

$$P(p_1 \leq p \leq p_2) = P(f(p_1) \leq f(p) \leq f(p_2)). \quad (17)$$

Therefore the confidence interval for $f(p)$ is $[f(p_1), f(p_2)]$.

Similarly, if the function $g$ is strictly decreasing, then

$$P(p_1 \leq p \leq p_2) = P(f(p_2) \leq f(p) \leq f(p_1)). \quad (18)$$

Therefore the confidence interval for $f(p)$ is $[f(p_2), f(p_1)]$. 
4 Hypothesis Testing about $p = \mathbb{E}(X)$

The goal of Hypothesis Testing is to determine whether the data provides enough evidence to “reject the null hypothesis” or otherwise “fail to reject it”.

The Null Hypothesis, denoted $H_0$, is a claim of the form:

$$H_0 : p = p_0.$$  \hspace{1cm} (19)

Its competing counterpart, the Alternative Hypothesis, denoted $H_A$, contains values of $p$ that we consider plausible if we reject $H_0$. It can be two-sided:

$$H_A : p \neq p_0,$$  \hspace{1cm} (20)

or one-sided:

$$H_A : p < p_0,$$  \hspace{1cm} (21)

$$H_A : p > p_0.$$  \hspace{1cm} (22)

4.1 One Sample $z$-test

1. State a null hypothesis $H_0$ and an alternative hypothesis $H_A$.

2. Verify $p = \mathbb{E}(X)$ and $\hat{p} = \bar{X}_n = (X_1 + X_2 + \cdots + X_n)/n$.

3. Assume the model of $X$ has only one parameter. If the null hypothesis is true, then by (3), (6) and (19) we have

$$\text{SD}(\hat{p}) = \frac{\sigma(p_0)}{\sqrt{n}},$$  \hspace{1cm} (23)

and

$$\frac{\hat{p} - p_0}{\text{SD}(\hat{p})} \approx Z \sim N(0,1).$$  \hspace{1cm} (24)

4. The P-value is the conditional probability of seeing the observed data or something less likely in the region of $H_A$ (as measured by the test statistic), given $H_0$ is true. Now calculate the test statistic from the observed data:

$$z = \frac{\hat{p} - p_0}{\text{SD}(\hat{p})},$$  \hspace{1cm} (25)

and it can be shown that

$$\text{P-value} = \begin{cases} \mathbb{P}(|Z| \geq |z|) & \text{if (20)} \\ \mathbb{P}(Z \leq z) & \text{if (21)} \\ \mathbb{P}(Z \geq z) & \text{if (22)} \end{cases}$$  \hspace{1cm} (26)

5. If the P-value is less than the significance level $\alpha$, the observed data would be too usual for the null hypothesis to be true, therefore we “reject $H_0$”. Otherwise, we “fail to reject $H_0$”.
4.2 One Sample t-test

Similar to Section 3.2, if the model of $X$ has multiple parameters or when $n$ is not large enough, we might want to use the $t$-distribution.

1. State a null hypothesis $H_0$ and an alternative hypothesis $H_A$.
2. Verify $p = E(X)$ and $\hat{p} = \bar{X}_n = (X_1 + X_2 + \cdots + X_n)/n$.
3. Instead of (23) we calculate
   \[ s_X = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X}_n)^2}{n-1}}. \] (27)

Then
   \[ \text{SD}(\hat{p}) \approx \frac{s_X}{\sqrt{n}}. \] (28)

If the null hypothesis is true and the Nearly Normal Condition is met,
   \[ \frac{\hat{p} - p_0}{s_X/\sqrt{n}} \approx T, \] (29)

where $T$ is a $t$-distribution with $n-1$ degrees of freedom.

4. Now calculate the test statistic from the observed data:
   \[ t = \frac{\hat{p} - p_0}{s_X/\sqrt{n}}, \] (30)

and it can be shown that
   \[ \text{P-value} = \begin{cases} 
   \mathbb{P}(|T| \geq |t|) & \text{if (20)} \\
   \mathbb{P}(T \leq t) & \text{if (21)} \\
   \mathbb{P}(T \geq t) & \text{if (22)} 
\end{cases} \] (31)

5. The conclusion is similar to Section 4.1.

4.3 Two Types of Error

**Type I Error** Reject $H_0$ when $H_0$ is true. By the definion of P-value, we know it happens with probability equal to the significance level $\alpha$.

**Type II Error** Fail to reject $H_0$ when $H_0$ is false ($H_A$ true). Denote by $\beta$ the probability of a type II error, which depends on the value of $p$. The power of the test is defined to be $1 - \beta$; the capability to reject $H_0$ when it is false.

For a fixed sample size $n$, if we choose a lower $\alpha$, then we also lower the power. The only way to reduce the probabilities of type I and type II error is to increase the sample size.