Lecture #12

- One of the most fundamental observations of probability is that the LLN can be strengthened to give quantitative estimates on the behavior of random sums.
- That is, if $\{X_i\}_{i=1}^n$ are iid samples from a common r.v. X, then not only is it the case that $S_n := \frac{1}{n}\sum_{i=1}^n X_i$ converges to $\mu := \mathbb{E}(X)$

$$S_n \xrightarrow{\mathbb{P}} \mathbb{E}(X) \quad (\text{WLLN})$$

but the deviations of S_n from μ are (after a suitable normalization) approximately normal: $\frac{\sqrt{n}\,[S_n - \mu]}{\sigma} \rightsquigarrow \mathcal{N}(0,1)$. (CLT)

$$=$$

$$\underbrace{\left(\frac{\sigma}{\sqrt{n}}\right)^{-1}}_{\text{"normalizing factor"}} \cdot [S_n - \mu] \rightsquigarrow \mathcal{N}(0,1)$$

- We call this "asymptotic normality."
- A fundamental observation in statistics is that is very, very useful. A broad family of hypothesis tests exploit the CLT. These are known as the Wald test when it is in general form

Wald Test: Consider the hypotheses $H_0: \theta = \theta_0$
$H_1: \theta \neq \theta_0$

Suppose $\hat{\theta}$ is asymptotically normal, i.e. $\frac{(\hat{\theta} - \theta_0)}{\hat{se}} \rightsquigarrow N(0,1)$ for some standard error $\hat{se}$. The Wald test rejects H_0 at the level α if $|W| > Z_{\alpha/2}$, where (1) $W = \frac{(\hat{\theta} - \theta_0)}{\hat{se}}$

(2.) $Z_{\frac{\alpha}{2}}$ is s.t. $\mathbb{P}\left(|Z| > \frac{Z_\alpha}{2}\right) = \alpha$, $Z \sim N(0,1)$

- Why does this make sense? Well, if $W = \frac{\hat{\theta} - \theta_0}{\hat{se}}$, then we are assuming $W \stackrel{\sim}{\sim} N(0,1)$ (with precise convergence as $n \to \infty$). Therefore, we are just treating W like $N(0,1)$ and proceeding accordingly.
- Recall that if the test is at level α, then $\mathbb{P}(\text{reject } H_0 \mid H_0 \text{ true}) = \alpha$. Under the Wald hypotheses, $W \stackrel{\sim}{\sim} N(0,1)$, so H_0 holding and our test choosing to reject ~~h~~ happens when the observed $W = \frac{\hat{\theta} - \theta_0}{\hat{se}}$ deviates strongly, made precise by $\underset{W}{\underset{\parallel}{|Z|}} > Z_{\frac{\alpha}{2}}$.

Remark: (1) Once we have data, W is determined: $\frac{\hat{\theta} - \theta_0}{\hat{se}}$

data-driven

fixed by H_0

data driven

(ii) On the other hand, before we see any data, W is a r.v.:

$$\frac{\hat{\theta} - \theta_0}{\hat{se}}$$

random $\rightarrow \hat{\theta}$; $\theta_0 \leftarrow$ fixed by H_0; random $\rightarrow \hat{se}$

Balancing (i) & (ii) is a key conceptual foundation of the application of probability (CLT) to statistics (hypothesis testing).

- We can apply a Wald test whenever $\hat{\theta}$ is asymptotically normal.

ex: Suppose $X \sim Bin(n, p_X)$, $Y \sim Bin(m, p_Y)$ are independent binomial r.v. Consider

The hypotheses that
$$H_0: p_X - p_Y = 0$$
$$H_1: p_X - p_Y \neq 0$$

Recall that the MLE for a binomial r.v. $X \sim Bin(n,p)$ is $\hat{p} = \frac{X}{n}$.

Then by equivariance, $\widehat{(p_X - p_Y)} = \frac{X}{n} - \frac{Y}{m}$.

What about the standard error? Well, $\hat{se} = \sqrt{Var(\hat{p_X} - \hat{p_Y})}$

$$\overset{\text{independence}}{=} \sqrt{Var(\hat{p_X}) + Var(\hat{p_Y})}$$

$$= \sqrt{Var\left(\frac{X}{n}\right) + Var\left(\frac{Y}{m}\right)}$$

$$= \sqrt{\frac{1}{n^2} Var(X) + \frac{1}{m^2} Var(Y)}$$

$$= \sqrt{\frac{1}{n^2} \cdot n \cdot (p_X)(1-p_X) + \frac{1}{m^2} \cdot (p_Y) \cdot (1-p_Y)}$$

$$\approx \sqrt{\frac{1}{n} \hat{p}_X \cdot (1-\hat{p}_X) + \frac{1}{m} \cdot \hat{p}_Y \cdot (1-\hat{p}_Y)}$$

So, the Wald statistic is $W = \frac{\hat{\theta} - \theta_0}{\hat{se}}$

$$= \frac{\hat{p}_X - \hat{p}_Y - [0]}{\sqrt{\frac{1}{n}\hat{p}_X(1-\hat{p}_X) + \frac{1}{m}\hat{p}_Y(1-\hat{p}_Y)}}$$

$$= \frac{\hat{p}_X - \hat{p}_Y}{\text{'' \quad ''}}$$

We would reject H_0 at the α level if $|W| > z_{\frac{\alpha}{2}}$, where as usual $z_{\frac{\alpha}{2}}$ is s.t. $\mathbb{P}(|Z| > z_{\alpha/2}) = \alpha$.

Let's pick some numbers. Suppose $n = 200$, $m = 100$ and we observe $X = 93$, $Y = 51$. Can we reject $H_0: p_X - p_Y = 0$ at the $\alpha = .01$ level?

Well, if we ran a Wald test, we estimate $\hat{p}_X = \frac{93}{200}$

$$\hat{p}_Y = \frac{51}{100}$$

$$\hat{se} = \sqrt{200 \cdot \left(\frac{93}{200}\right) \cdot \left(1 - \frac{93}{200}\right) + 100 \cdot \left(\frac{51}{100}\right)}$$

$\left(1-\frac{51}{100}\right)$

Hence, the Wald statistic is

$$\frac{\hat{p}_X - \hat{p}_Y}{\sqrt{\frac{1}{n}(\hat{p}_X)(1-\hat{p}_X)+\frac{1}{m}(\hat{p}_Y)(1-\hat{p}_Y)}}$$

computer

$= -.7355$

So, for $\alpha = .01$, $Z_{\frac{\alpha}{2}} \approx 2.576$. So, we reject H_0 iff $|-.7355| > 2.576$. So, we don't reject H_0.

<u>Remark</u>. We can also estimate the standard error using se_0, which is the standard error under H_0. Both this formulation (se_0) and the original $(\widehat{se})$ are valid.