Stat 342 Example 17

Suppose $X_1, X_2, \ldots, X_n$ are iid Bernoulli p. Then

$$\hat{p} = \overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

$$= \frac{\#\text{ of } X_i\text{'s that are } 1}{n}$$

$$= \text{sample fraction of } X_i\text{'s that are } 1$$

is covered by the LLN. That is, we are guaranteed that for any $\epsilon > 0$ with probability going to 1 as $n \to \infty$, $\hat{p} = \overline{X}$ is within ϵ of $EX_1 = p$.

This is an important result for example in simulations. If I make independent simulations of a probability model and note the sample fraction of those in which event A occurs, I may expect

$$\hat{p}_n = \text{sample relative frequency of event } A$$

to approximate

$$p = P(A) = \text{the model probability of } A$$

(with large probability) when n is big.

The CLT also applies to the distribution of $\hat{p}$ promising that after standardization approximate probabilities for $\hat{p}$ can be gotten from the standard normal dsn. ($\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$ is approximately standard normal. Note that this is $\frac{\overline{x} - \mu}{\sigma/\sqrt{n}}$ for the case of the X_i being Bernoulli.

So, for example, if $n = 400$ and $p = .5$

$$P\left[.49 < \hat{p} < .51\right] = P\left[\underbrace{\frac{.49 - .5}{\sqrt{\frac{(.5)(.5)}{400}}}}_{-.4} < \frac{\hat{p} - .5}{\sqrt{\frac{(.5)(.5)}{400}}} < \underbrace{\frac{.51 - .5}{\sqrt{\frac{(.5)(.5)}{400}}}}_{.4}\right]$$

$$\approx \Phi(.4) - \Phi(-.4)$$