15.3.9 Confidence Intervals for Probability Estimation

MO2.3 **MO2.13** **MO2.14** **MO2.15**

Now we develop the confidence interval for a probability estimates. This confidence interval is derived by re-writing the $\hat{p}_{\text{event}}$ probability estimate given in Equation (15.14) in terms of the mean of an indicator function x:

- $x = 0$: when the event did not occur for an instance
- $x = 1$: when the event did occur for an instance

With this indicator function, then we can prove that the sample mean of x is equal to $\hat{p}_{\text{event}}$:

$$\overline{x} = \frac{1}{N}\sum_{i=0}^{N-1} x_i = \frac{N_{\text{event}}}{N} = \hat{p}_{\text{event}} \tag{15.24}$$

Thus, we can use the properties of the sample mean estimate to derive the properties of the $\hat{p}_{\text{event}}$ estimate. Specifically,

- Since $\overline{x}$ is an unbiased estimate of μ_x, then $\hat{p}_{\text{event}}$ is an unbiased estimate of p_{event}.

- With some algebraic manipulations, the standard error for $\hat{p}_{\text{event}}$ can be shown to be:

$$\sigma_{\hat{p}}^2 \equiv E\left[\left(\hat{p}_{\text{event}} - p_{\text{event}}\right)^2\right] = \frac{p_{\text{event}}\left(1 - p_{\text{event}}\right)}{N} \tag{15.25}$$

- In practice, the standard error is estimated by assuming $p_{\text{event}} \approx \hat{p}_{\text{event}}$, such that,

$$\sigma_{\hat{p}}^2 \approx \frac{\hat{p}_{\text{event}}\left(1 - \hat{p}_{\text{event}}\right)}{N} \tag{15.26}$$

- A 95% confidence interval for the probability estimate is then,

$$\hat{p}_{\text{event}} - 1.96\sigma_{\hat{p}} < p_{\text{event}} < \hat{p}_{\text{event}} + 1.96\sigma_{\hat{p}} \tag{15.27}$$

The following video discusses this confidence interval for a probability estimate.

The Python script used in this video (and several others) is available here.

Video on confidence intervals for probability estimates

of the probability I'm interested

in is plus or minus 2 of the

Media Kit

Previous

Next

is the p-hat plus or minus 2 sigma p-hat 2 times