Notes on MLE

January 27, 2020

In our example

$X \sim \text{Binomial}(16, p)$

- p is an unknown, fixed number that we want to estimate
- X is a random variable: we may observe a different number every time we run the experiment due to random chance (which babies are selected or the study, etc.)

We observe $x = 14$ and will estimate p by choosing the value for which the probability of the observed data is highest:

$\hat{p} = \arg\max f_X(x|p) = \binom{n}{x} p^x (1-p)^{n-x}$

Think of this as a function of p, holding fixed the observed value of x.

$\mathcal{L}(p|x) = f_X(x|p) = \binom{n}{x} p^x (1-p)^{n-x}$

We want to maximize this: find $\hat{p}^{MLE}$ such that $\mathcal{L}(p|x) \leq \mathcal{L}(\hat{p}^{MLE}|x)$ for every value of p

Note: it's almost always easier to maximize the log-likelihood, $\ell(p|x) = \log\{\mathcal{L}(p|x)\}$.

This will give us the same estimate:

- $\log(w)$ is an increasing function: $w_1 < w_2 \Leftrightarrow \log(w_1) < \log(w_2)$
- Plug in $\mathcal{L}(p|x)$ for w_1 and $\mathcal{L}(\hat{p}^{MLE}|x)$ for w_2:
  - $\mathcal{L}(p|x) < \mathcal{L}(\hat{p}^{MLE}|x) \Leftrightarrow \log(\mathcal{L}(p|x)) < \log(\mathcal{L}(\hat{p}^{MLE}|x))$
  - So $\hat{p}^{MLE}$ maximizes the likelihood if and only if it maximizes the log-likelihood

In this example, we find the log-likelihood, differentiate, solve for p to find a critical point, and verify the critical point is a maximum with a second derivative test.

More generally

We have random variables $X_1, \ldots, X_n$

We model them as following some distribution with unknown parameters θ.

We will estimate θ by choosing the value for which the probability of the observed data is highest.

The likelihood function is:

$$\mathcal{L}(\theta|x_1, \ldots, x_n) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n|\theta)$$

If $X_1, \ldots, X_n$ are independent and identically distributed, then we can go one step further:

$$\begin{aligned}\mathcal{L}(\theta|x_1, \ldots, x_n) &= f_{X_1,\ldots,X_n}(x_1, \ldots, x_n|\theta) \\ &= f_{X_1}(x_1|\theta) \cdots f_{X_n}(x_n|\theta) \\ &= \prod_{i=1}^{n} f_{X_i}(x_i|\theta)\end{aligned}$$

In this case, the log-likelihood is:

$$
\begin{aligned}
\ell(\theta|x_1,\ldots,x_n) &= \log\left\{\mathcal{L}(\theta|x_1,\ldots,x_n)\right\} \\
&= \log\left\{\prod_{i=1}^{n} f_{X_i}(x_i|\theta)\right\} \\
&= \sum_{i=1}^{n} \log\left\{f_{X_i}(x_i|\theta)\right\}
\end{aligned}
$$