

# Module 6 & 7: Summary

## Statistics and Parameter Estimation

Since it is practically impossible to enroll the whole target population, we take a **sample** - a subgroup representative of the population. Since we're not examining the whole population, inferences will not be certain. **Probability** is the ideal tool to model and communicate uncertainty inherent with informing the population characteristic based on a sample. Inferences are categorized into two broad categories:

1. Estimation - Estimate the value of a parameter based on a sample
2. Hypothesis Testing - Comparing parameters for two sub-populations using tests of significance

For smaller sample sizes ( $n < 30$ ) we can use a t distribution.

## Parameter Estimation

In many statistical problems we make an assumption on the probability distribution from which the data are generated. The likelihood function is a concept that indicates how likely different parameters are to fit your distribution. Maximum likelihood is an approach based on selecting the parameter values that make the observed sample most likely.

If  $X_1, \dots, X_n$  is a sample of independent observations from  $X \sim f(x; \theta)$ , the likelihood function is defined as:

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta) = \prod_{i=1}^n p^{x_i} (1 - p)^{1-x_i}$$

The product of each observation (marginals); As a joint distribution is the product of the marginals when the observations are independent (identically distributed). For binomial distributions, this can be further simplified:

$$= (p^{\sum_{i=1}^n x_i})(1 - p)^{\sum_{i=1}^n (1-x_i)} = p^x (1 - p)^{n-x}$$

For a Poisson Likelihood with the mean  $\rightarrow p(X=x) = (\mu^x * e^{-\mu})/x!$ ;  $\mu$  = mean

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta) = \prod_{i=1}^n \frac{\theta^{x_i} e^{-\theta}}{x_i!} = \frac{(\theta^{\sum_{i=1}^n x_i}) e^{-n\theta}}{\prod_{i=1}^n x_i!}$$

We could also express the Poisson likelihood function in terms of rates for each subject. So  $X_i \sim \text{Poisson}(m_i * p)$ ; where  $m$  is the number of trials and  $p$  is the probability of success, and assuming independence.

$$= \prod_{i=1}^n \frac{m_i^{x_i} p^{x_i} e^{-m_i p}}{x_i!} = \frac{(p^{\sum_{i=1}^n x_i}) (e^{-\sum_{i=1}^n m_i p}) \prod_{i=1}^n m_i^{x_i}}{\prod_{i=1}^n x_i!}$$

For a Normal Distribution likelihood can be expressed in terms of mean and variance:

$$\begin{aligned} \prod_{i=1}^n f(x_i; \mu, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(x_i - \mu)^2}{\sigma^2}\right) = \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^2}\right) \end{aligned}$$

## MLE

The **Maximum Likelihood Estimate (MLE)** is the value of the parameter that maximizes the likelihood equation. Often we work with the log-likelihood because it will lead to the same maximize (since log is a strictly increasing function). To find this with calculus we can differentiate and set the derivative to 0.

## Concepts

- An estimator  $T$  of a parameter  $\mu$  is unbiased if  $E(T) = \mu$
- The MLE  $\hat{\mu}_{\text{hat}}$  is not always unbiased. However, under general conditions the probability  $\hat{\mu} = \hat{\mu}_{\text{hat}}$  approaches 1 as the sample size  $n$  grows to infinity. (Consistency of MLE)
  - The practicality of this is that MLE on average will approximate the true population value in large samples.

- When two estimators are unbiased one can compare them by variances. The best estimator would be the one with smaller variance -> more precise.
- To compare biased estimators we use mean squared error:
  - $MSE(T) = E(T - \mu)^2 = (E(T) - \mu)^2 + V(T) = \text{bias}(T)^2 + V(T)$
- Sometimes it is preferable to have a biased estimator with a low variance (bias-variance tradeoff)

In logistic regression we often take the log of the Odds Ratio.

---


$$\text{Odds Ratio} = ad/bc = \frac{(a/m_1)/(b/m_1)}{(c/m_0)/(d/m_0)} = \frac{\hat{p}_D/(1-\hat{p}_D)}{\hat{p}_{\bar{D}}/(1-\hat{p}_{\bar{D}})}$$

$$\text{Var}(\log(\widehat{OR})) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

**Binomial:**

$$L(p) = p^x(1-p)^{n-x}$$

$$l(p) = \log(L(p)) = x \cdot \log(p) + (n-x) \cdot \log(1-p)$$

$$dl(p) / dp = x / p - (n-x) / (1-p) = 0$$

$$(x \cdot (1-p) - (n-x) \cdot (1-p)) / (p \cdot (1-p)) = 0$$

$$x = np \rightarrow p_{\text{hat}} = x/n$$

Based on CLT, when n is large:

$$X \sim N(np, np(1-p)) \text{ and } p_{\text{hat}} = X / n \sim N(p, p(1-p)/n)$$

**Poisson:**

For means:

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta) = \frac{(\theta^{\sum_{i=1}^n x_i}) e^{-n\theta}}{\prod_{i=1}^n x_i!}$$

$$l(\theta) = \log(L(\theta)) \propto \sum_{i=1}^n x_i \log(\theta) - n\theta \Rightarrow \frac{dl(\theta)}{d\theta} = \frac{\sum_{i=1}^n x_i}{\theta} - n = 0$$

$$\frac{\sum_{i=1}^n x_i - n\theta}{\theta} = 0 \Leftrightarrow \boxed{\hat{\theta} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}}$$

For Probabilities:

$$\log(L(p)) = \sum_{i=1}^n x_i \log(p) - p \sum_{i=1}^n m_i \Rightarrow \frac{d \log(L(p))}{dp} = \frac{\sum_{i=1}^n x_i}{p} - \sum_{i=1}^n m_i = 0 \Rightarrow \boxed{\hat{p} = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n m_i}}$$

Normal Distributions:

$$\frac{\partial l(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \Rightarrow \boxed{\hat{\mu} = \bar{x}}$$

$$\begin{aligned} \frac{\partial l(\mu, \sigma^2)}{\partial \sigma^2} &= \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 - \frac{n}{2\sigma^2} = 0 \Rightarrow \text{replace } \mu \text{ by } \hat{\mu} = \bar{x} \\ \Rightarrow \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \bar{x})^2 - \frac{n}{2\sigma^2} &= 0 \Rightarrow \boxed{\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \end{aligned}$$

This looks very similar to our estimate of  $S^2$  or "Sample Variance" but has  $n$  in the denominator instead of  $(n-1)$ , meaning this is biased at low sample sizes.

## Large Sample Approximation

Fisher Information Matrix can be used to forecast the precision in one observation:

$$I(\theta) = -E\left(\frac{\partial^2 \log(f(x; \theta))}{\partial \theta^2}\right)$$

This is a very important property that allows us to generate approximate distribution for MLE when the sample size is large.

---

Revision #9

Created 22 August 2022 13:22:22 by Elkip

Updated 23 August 2022 16:03:33 by Elkip