Likelihood Function

CSU, Chico Math 314

2018-09-17

outline

Review: estimating parameters

Recall that statistics attempts to estimate the (population) parameters of interest via an assumed known functional form. The population parameters are in reality functions of unknown variables of a probility mass/distribution function.

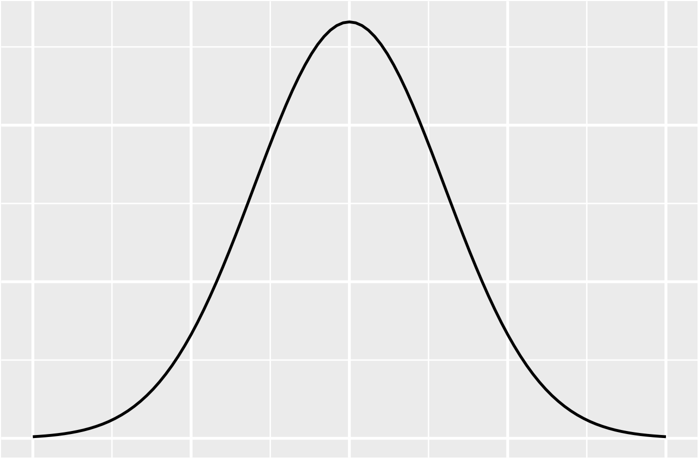
outline

Motivating Likelihood

If we know the distribution function from which a sample of data was generated, we should be able to estimate the parameter(s) of the distribution that is most likely to have generated these data.

Motivating Likelihood

Suppose we randomly sampled 5 people and found heights $69, 72, 70, 68, 71, 71$ inches. What are the parameters that are most likely to have brought about these data?



outline

Bernoulli Example

Suppose we observed $x_1, \ldots, x_n \sim_{iid}$ Bernoulli(p). What is the maximum likelihood estimator for p?

$$f(x|p) = p^x(1-p)^{1-x}$$

outline

Likelihood Function

Consider n observations $x_1, \ldots, x_n$ independently drawn from the distribution $G(x|\theta)$. We seek to find the parameter θ that is most likely to have generated the data $\mathbf{x}$.

likelihood function

$$L(\theta|\mathbf{x}) = \prod_{i=1}^{n} G(x_i|\theta).$$

outline

Likelihood Function, notes

- We think about L as a function of θ instead of $\mathbf{x}$.
- All observations are independent and identically distributed, *iid*.
- $\hat{\theta} = \arg\max_{\theta} L$, is the mathematical equivalent of "parameters most likely to have generated the observed data."
- It should be immediately clear that nobody wants to work with $L(\theta|\mathbf{x})$, itself.

outline

Maximum Likelihood

Estimators derived via the likelihood function are called **maximum likelihood estimators** or **MLE**s. MLE estimators estimate parameters of interest by maximizing the likelihood function. That is $\hat{\theta}$ is the MLE if

$$\hat{\theta} = \arg\max_{\theta} L(\theta|\mathbf{x}).$$

Maximum Likelihood, computation

So how does one find $\hat{\theta}$?

- Calculus.
- It's immensely helpful to work with the natural log of the likelihood function.
- Maximize L as a function of θ just as we would any other function f in calculus.

Detour the First

How can we rewrite (simplify?) the following equations?

- $\log(a \cdot b)$?
- $\log(x_1 \cdot x_2)$?
- $\log(x_1 \cdot x_2 \cdot \ldots \cdot x_n)$?
- $\log(\prod_{i=1}^{n} x_i)$?
- $\log(\prod_{i=1}^{n} f(x_i))$?

Likelihood Function, computation

Use properties about the natural log to manipulate the likelihood function L into a reasonable function to optimize.

$$\ell(\theta|\mathbf{x}) = \log L(\theta|\mathbf{x}) = \sum_{i=1}^{n} \log G(x_i|\theta).$$

Log Likelihood Function, notes

The log-likelihood function ℓ has a two main advantages over L.

- ℓ is easier to optimize for both humans and computers.
- $\hat{\theta} = \arg\max_\theta \ell(\theta|\mathbf{x}) = \arg\max_\theta L(\theta|\mathbf{x})$ because log() is monotone.

Detour the Second

How can we rewrite (simplify?) the following equations?

- $\frac{d}{d\theta}\{a \cdot \theta + b \cdot \theta\}$?
- $\frac{d}{d\theta}\{x_1 \cdot \theta + x_2 \cdot \theta\}$?
- $\frac{d}{d\theta}\{x_1 \cdot \theta + \ldots + x_n \cdot \theta\}$?
- $\frac{d}{d\theta}\{\sum_{i=1}^{n} x_i \cdot \theta\}$?

Detour the Third

Evaluate the following expression,

$$\frac{d}{d\theta} = x \cdot \log(1-\theta).$$

outline

Exponential Example

Suppose we observed $x_1, \ldots, x_n \sim_{iid}$ Exponential(λ). What is the MLE for λ?

$$f(x|\lambda) = \lambda \exp(-\lambda x)$$

Examples via a Computer

Note, we are going to brush over many topics that are often at least one class worth of material: optimization techniques.

```
?optim
?rbinom
?rpois
```

outline

references I

Michael G. Akritas. *Probability and Statistics with R for Engineers and Scientists*. Pearson Education, Inc., 2016.