9.7 The Method of Maximum Likelihood

- Recall that the likelihood function is the joint pdf considered as a function of the parameter(s), given the observed data.
- Loosely speaking, the value of θ that maximizes $L(\theta \mid \underset{\sim}{y})$ is the parameter value that is "most likely" to have produced the data we did observe.
- The maximum likelihood estimator (MLE) of a parameter θ is:

Note: In MANY cases, it is easier to maximize $\ln L(\theta)$ than to maximize $L(\theta)$ itself.

- Since $\ln(\cdot)$ is an increasing function, the θ-value that maximizes $\ln L(\theta)$ will also maximize $L(\theta)$.
- So maximizing the log-likelihood $\ln L(\theta)$ will yield the MLE.

- Often the MLE is found by:

1. Writing out the (log) likelihood as a function of the parameter (say, θ).
2. Taking the derivative with respect to θ.
3. Setting the derivative equal to 0 and solving for $\hat{\theta}$.
4. Checking that the <u>2nd</u> <u>derivative</u> is ________ at $\hat{\theta}$ to ensure the solution is a <u>maximum</u>.

<u>Example 1</u>: $Y_1, \ldots, Y_n \overset{iid}{\sim}$ Bernoulli(p). Find the MLE of p.

- Example 2: Let $Y_1, \ldots, Y_n \overset{iid}{\sim} \text{Unif}(0, \theta)$. Find the MLE of θ.

MLE's with multiple parameters

- If we are using maximum likelihood to estimate several parameters, we must take <u>partial</u> derivatives of the (log) likelihood with respect to <u>each</u> parameter.
- We set each partial derivative to zero and solve the equations simultaneously.
- <u>Example 3</u>: $Y_1, \ldots, Y_n \overset{iid}{\sim} N(\mu, \sigma^2)$ where μ and σ^2 are unknown. Find MLEs of μ and σ^2.

  (Let's write $v = \sigma^2$ to make notation easier.)

- **Exercise**: Let $Y_1, \dots, Y_n$ be iid with pdf

$$f_Y(y) = \begin{cases} \frac{1}{\theta}\, y^{\frac{1-\theta}{\theta}} & \text{for } 0 < y < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Find the MLE of θ.

$L(\theta) =$

$\ln L(\theta) =$

Properties of MLEs

- Note that if U is a sufficient statistic for θ, then:

$\Rightarrow$ Maximizing $L(\theta)$ with respect to θ is equivalent to maximizing $g(u, \theta)$ with respect to θ.

$\Rightarrow$ The MLE $\hat{\theta}$ will be a function of the sufficient statistic U.

- This tells us: If we find an MLE, and adjust it so it is unbiased, this adjusted estimator will (often) be the MVUE.

Invariance Property

- We are often interested in estimating a <u>function</u> of a parameter.
- The invariance property of MLEs states that if $\hat{\theta}$ is a MLE of θ, and $g(\cdot)$ is any function, then:

- <u>Example 4</u>: $Y_1, \ldots, Y_n \overset{iid}{\sim}$ Bernoulli(p). Use ML to estimate $\text{var}\left(\sum_{i=1}^{n} Y_i\right)$.

- Note $\sum_{i=1}^{n} Y_i \sim$

- <u>Exercise:</u> Let $Y_1, \dots, Y_n \overset{iid}{\sim} \text{Pois}(\lambda)$. Show that the MLE $\hat{\lambda} = \bar{Y}$ and use the invariance property to estimate $P(Y=0) = e^{-\lambda}$ with ML.

Maximizing the Likelihood Numerically

- Sometimes it is too difficult to take derivatives of $L(\theta)$ or $\ln L(\theta)$.
- We can use software to find MLEs <u>numerically</u>.
  - <u>Example 5:</u> $Y_1, \dots, Y_{30} \overset{iid}{\sim} \text{Gamma}(\alpha, \beta)$. Find MLEs of α and β, given the 30 data values.

    It can be shown that

  - We cannot maximize this analytically, but using R, we can find MLEs numerically, given our sample data.