Recall: for continuous distributions, we can't compute the probability of a random variable taking on an *exact* value; only intervals

$$\Rightarrow P(Y = 1.1; \theta) = \int_{1.05}^{1.15} f(y_i; \theta) \approx \underbrace{(0.1)}_{1.15-1.05} f(1.1; \theta)$$

Ex. Let Y_i represent the time until a light bulb breaks down, where we assume $Y_i \sim \text{Exp}(\theta)$. Consider a sample of data $\{y_1, \ldots, y_n\}$ where $f(y_i; \theta) = \frac{1}{\theta} e^{-y_i/\theta}$. Show that the MLE for θ is $\hat{\theta} = \bar{y}$.

$$P(Y = y; \theta) = P(Y_1 = y_1)\, P(Y_2 = y_2) \cdots P(Y_n = y_n)$$

$$= \prod_{i=1}^{n} P(Y_i = y_i; \theta)$$

$$= \prod_{i=1}^{n} \frac{1}{\theta} e^{-y_i/\theta}$$

$$= \theta^{-n} \cdot e^{-\frac{1}{\theta}\sum_{i=1}^{n} y_i}$$

$$l(\theta) = \ln\left(\theta^{-n} \cdot e^{-\frac{1}{\theta}\sum_{i=1}^{n} y_i}\right)$$

$$= -n\ln(\theta) - \theta^{-1}\sum_{i=1}^{n} y_i$$

$$\frac{d\,l(\theta)}{l(\theta)} = \frac{-n}{\theta} + \theta^{-2}\sum_{i=1}^{n} y_i = 0$$

set derivative to 0 since we are solving for maximum

$$\Rightarrow \hat{\theta} = \sum_{i=1}^{n} y_i / n = \bar{y}$$

(sample mean)

Example: Suppose $f(x_i; b, \theta) = \theta b^{\theta} x^{-\theta-1}$ for $b > 0$, and let $X_1, \ldots, X_n$ be an i.i.d sample

Find the MLE of θ.

$$f(x_i; b, \theta) = \theta b^{\theta} x^{-\theta-1}$$

$$L(\theta) = P(X = x;\, \theta) = \prod_{i=1}^{n} \theta b^{\theta} x_i^{(-\theta-1)}$$

$$= \theta^n b^{\theta n} \prod_{i=1}^{n} x_i^{(-\theta-1)}$$

$$\ell(\theta) = n \cdot \ln(\theta) + n\theta \cdot \ln(b) + \ln\left[\prod_{i=1}^{n} x_i^{(-\theta-1)}\right]$$

$$= n \cdot \ln(\theta) + n\theta \cdot \ln(b) + \sum_{i=1}^{n} \ln\left[x_i^{(-\theta-1)}\right]$$ log rules

$$= n \cdot \ln(\theta) + n\theta \cdot \ln(b) - (\theta+1) \sum_{i=1}^{n} \ln(x_i)$$

$$= n \cdot \ln(\theta) + n\theta \cdot \ln(b) - \theta \sum_{i=1}^{n} \ln(x_i) - \sum_{i=1}^{n} \ln(x_i)$$

$$\frac{d\ell(\theta)}{\ell(\theta)} = \frac{n}{\theta} + n\ln(b) - \sum_{i=1}^{n} \ln(x_i) = 0$$ simply isolate θ

Ex.

8. Suppose $y_1, y_2, \ldots, y_n$ is an observed random sample from the distribution with probability density function

$$f(y;\theta) = (\theta+1)y^{\theta} \quad \text{for } 0 < y < 1 \text{ and } \theta > -1$$

(a) Find the likelihood function $L(\theta)$, the log likelihood function $l(\theta)$, and the maximum likelihood estimate $\hat{\theta}$.

$$L(\theta) = \prod_{i=1}^{n} (\theta+1) y_i^{\theta}$$

$$= (\theta+1)^n \prod_{i=1}^{n} y_i^{\theta}$$

$$\ell(\theta) = n \cdot \ln(\theta+1) + \ln\left(\prod_{i=1}^{n} y_i^{\theta}\right)$$

$$= n \cdot \ln(\theta+1) + \theta \sum_{i=1}^{n} \ln(y_i)$$

$$\Rightarrow \frac{d\ell(\theta)}{d\theta} = \frac{n}{\theta+1} + \sum_{i=1}^{n} \ln(y_i) = 0$$

$$\Rightarrow \hat{\theta} = \frac{n}{-\sum_{i=1}^{n} \ln(y_i)} - 1$$

b) Find the log relative likelihood function $r(\theta) = \log R(\theta)$.

$$R(\theta) = \frac{L(\theta)}{L(\hat{\theta})}; \quad r(\theta) = \ln(R(\theta)) = \ln(L(\theta)) - \ln(L(\hat{\theta}))$$ log rules

$$= \ell(\theta) - \ell(\hat{\theta})$$

$$= n \cdot \ln(\theta + 1) + \theta \sum_{i=1}^{n} \ln(y_i) - n \cdot \ln(\hat{\theta} + 1) - \hat{\theta} \sum_{i=1}^{n} \ln(y_i)$$

Multinomial Distributions

Suppose we had 20 students pick from k=3 options.

We can have a frequency table

- Option 1: 6
- Option 2: 11
- Option 3: 3

The **likelihood function** for a sample $y_1, \ldots, y_n$ from the multinomial distribution is expressed as:

$$L(\theta) = \frac{n!}{y_1! \ldots y_k!} \theta_1^{y_1} \ldots \theta_k^{y_k}$$

θ_i : proportion of each category

y_i : count of each category

Finding the log likelihood function, we eventually get

$$\ell(\theta) = y_1 \ln(\theta_1) + \cdots + y_n \ln(\theta_n)$$

We could then try to calculate the MLE for each θ_i using partial derivatives, but that would give us something like

$$\frac{\partial \ell(\theta_i)}{\partial \theta_i} = \frac{y_i}{\theta_i} = 0 \rightarrow \text{all } y_i = 0 \;???$$

Instead:

- We can either...
  - Maximize $\ell(\theta)$ subject to the constraint $\sum_{i=1}^{k} \theta_i = 1$ using Lagrange Multpliers
  - Consider each Y_i as having a binomial distribution, where $Y_i \sim \text{Bin}(n, \theta_i)$
- $P(Y_i = y_i; \theta_i) = \binom{n}{y_i} \theta_i^{y_i}(1-\theta_i)^{n-y_i}$

$$\hat{\theta}_i = \frac{y_i}{n}$$

Invariance Property of Maximum Likelihood Estimates: If $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_k)$ is the MLE of $\theta = (\theta_1, ..., \theta_k)$, then $g(\hat{\theta})$ is the MLE of $g(\theta)$

Invariance Property of MLE: Example

- Suppose the number of Canadians who are regular coffee-drinkers can be modelled with a binomial distribution, $Y \sim \text{Bin}(n, \theta)$, and from a sample of 1000 Canadians, we observed 684 individuals identifying as regular coffee drinkers.
  - $\hat{\theta} = y/n = 0.684$
  - The MLE of $g(\theta) = e^{\theta}$ would be $e^{\hat{\theta}}$