Sufficient statistics

You are given a data set $x = x_1^n$ of n i.i.d. geometrically distributed observations. Show that $\sum_{i=1}^n$ is a sufficient statistic for the geometric distribution.

Sufficient statistics

We will use the factorisation theorem for this exercise.

$$P(X_1^n|\Theta = \theta) = \prod_{i=1}^{n} P(X_i = x_i|\Theta = \theta) = \theta^n(1-\theta)^{\sum_{i=1}^n x_i}$$

By letting $g(\theta, \sum_{i=1}^n x_i) = \theta^n(1-\theta)^{\sum_{i=1}^n x_i}$ and $h(x, \sum_{i=1}^n x_i) = \frac{1}{c}$ the result follows from the factorisation theorem. Here, c is the number of sequences of n geometric draws whose sum is equal to $\sum_{i=1}^n x_i$.

Coin Tosses

A coin is taken from a box containing three coins, which give heads with probability $p = 1/3, 1/2$, and $2/3$. The mystery coin is tossed 80 times, resulting in 49 heads and 31 tails.

a) What is the likelihood of this data for each type on coin? Which coin gives the maximum likelihood?

b) Now suppose that we have a single coin with unknown probability p of landing heads. Find the likelihood and log likelihood functions given the same data. What is the maximum likelihood estimate for p?

Coin Tosses

a) The data x is 49 heads in 80 tosses. We have three hypotheses: the coin has probability $p = 1/3, p = 1/2, p = 2/3$. So the likelihood function $L_x : p \mapsto P(X = x | P = p)$ takes 3 values:

$$L_x(\frac{1}{3}) = \binom{80}{49} \left(\frac{1}{3}\right)^{49} \left(\frac{2}{3}\right)^{31} = 6.24 \cdot 10^{-7}$$

$$L_x(\frac{1}{2}) = \binom{80}{49} \left(\frac{1}{2}\right)^{49} \left(\frac{1}{2}\right)^{31} = 0.024$$

$$L_x(\frac{2}{3}) = \binom{80}{49} \left(\frac{2}{3}\right)^{49} \left(\frac{1}{3}\right)^{31} = 0.082$$

Thus the maximum likelihood is achieved under $p = 2/3$.

Coin Tosses

b) We already know that the MLE for the Bernoulli (and binomial) distribution is $\frac{k}{n}$ where k is the number of successes and n is the total number of Bernoulli trials. Thus we get

$$p^* = \frac{49}{80} = 0.6125$$

as the MLE in the present example.

Dice

There are five fair dice each with a different number of sides: 4,6,8,12,20. Jon picks one of them uniformly at random rolls it and reports a 13.

a) Compute the posterior probability for each die to have generated this outcome.

b) Compute the posterior probabilities if the result had been a 5 instead. *Hint: Drawing a table may help here. And please do use a calculator!*

Dice

a) For all but the 20-sided die the likelihood and hence the posterior probability is 0. By Bayes' rule this means that the posterior probability of the 20-sided die is 1. The numerator and the denominator in Bayes' rule take on the same value in this example.

Dice

We use a table to display the solution. We identify the dice by their number of sides.

Die	P(Die)	P(5—Die)	P(5—Die) P(Die)	P(Die—5)
4	$1/5$	0	0	0
6	$1/5$	$1/6$	$1/30$	0.392
8	$1/5$	$1/8$	$1/40$	0.294
12	$1/5$	$1/12$	$1/60$	0.196
20	$1/5$	$1/20$	$1/100$	0.118
total	1		0.085	1

Geometric MLE

For an i.i.d. data set $x = x_1^n$ find the MLE for the geometric distribution:

$$P(X = x) = (1 - \theta)^x \theta$$

Geometric MLE

The likelihood function is

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

and thus the log-likelihood is

$$\mathcal{L}_x(\theta) = n\log(\theta) + \sum_{i=1}^{n} x_i \log(1-\theta) \ .$$

Geometric MLE

In the next step we compute the score function.

$$\begin{aligned}
\frac{d}{d\theta}\mathcal{L}_x(\theta) &= \frac{d}{d\theta} n \log(\theta) + \frac{d}{d\theta}\sum_{i=1}^{n} x_i \log(1-\theta) \\
&= \frac{n}{\theta} - \frac{\sum_{i=1}^{n} x_i}{1-\theta}
\end{aligned}$$

Geometric MLE

Setting this to 0 gives

$$\begin{aligned} 0 &= \frac{n}{\theta} - \frac{\sum_{i=1}^{n} x_i}{1-\theta} && \Leftrightarrow \\ \frac{n}{\theta} &= \frac{\sum_{i=1}^{n} x_i}{1-\theta} && \Leftrightarrow \\ n - n\theta &= \theta \sum_{i=1}^{n} x_i && \Leftrightarrow \\ \theta &= \frac{n}{n + \sum_{i=1}^{n} x_i} \end{aligned}$$

Geometric MLE, variant 2

For an i.i.d. data set $x = x_1^n$ find the MLE for the geometric distribution:

$$P(X = x) = (1 - \theta)^{x-1}\theta$$

Geometric MLE, variant 2

The likelihood function is

$$L_x(\theta) = \prod_{i=1}^{n} \theta(1-\theta)^{x_i - 1} = \theta^n (1-\theta)^{\left(\sum_{i=1}^n x_i\right) - n}$$

and thus the log-likelihood is

$$\mathcal{L}_x(\theta) = n \log(\theta) + \left(\left(\sum_{i=1}^{n} x_i \right) - n \right) \log(1-\theta) \ .$$

Geometric MLE, variant 2

In the next step we compute the score function.

$$\frac{d}{d\theta}\mathcal{L}_x(\theta) = \frac{d}{d\theta} n \log(\theta) + \frac{d}{d\theta}\left((\sum_{i=1}^{n} x_i) - n \right) \log(1-\theta)$$

$$= \frac{n}{\theta} - \frac{(\sum_{i=1}^{n} x_i) - n}{1-\theta}$$

Geometric MLE, variant 2

Setting this to 0 gives

$$
\begin{aligned}
0 &= \frac{n}{\theta} - \frac{((\sum_{i=1}^{n} x_i) - n)}{1-\theta} && \Leftrightarrow \\
\frac{n}{\theta} &= \frac{((\sum_{i=1}^{n} x_i) - n)}{1-\theta} && \Leftrightarrow \\
n - n\theta &= (\theta \sum_{i=1}^{n} x_i) - n\theta && \Leftrightarrow \\
\theta &= \frac{n}{\sum_{i=1}^{n} x_i}
\end{aligned}
$$