Estimation

Gene Quinn

The Estimation Problem

A *probability model* generally assumes that experimental outcomes follow a probability distribution.

The Estimation Problem

A *probability model* generally assumes that experimental outcomes follow a probability distribution.

Usually, these distributions involve one or more **parameters**.

A complete specification of the probability distribution involves these parameters.

The Estimation Problem

A *probability model* generally assumes that experimental outcomes follow a probability distribution.

Usually, these distributions involve one or more **parameters**.

A complete specification of the probability distribution involves these parameters.

The process of determining suitable values for these parameters is called *estimation* and is a major topic in statistics.

The Estimation Problem

Example: Suppose an experiment has two possible outcomes, success and failure.

Suppose we conduct 10 independent trials of the experiment, and have reason to believe that the probability of success p is the same on each trial.

The Estimation Problem

Example: Suppose an experiment has two possible outcomes, success and failure.

Suppose we conduct 10 independent trials of the experiment, and have reason to believe that the probability of success p is the same on each trial.

If we associate 1 with the outcome "success" and 0 with the outcome "failure", we can state the probabilities associated with these values as:

$$P(X = x) \; = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

The Estimation Problem

Define a vector of outcomes for the 10 trials,

$$x = \{x_1, x_2, \ldots, x_{10}\}$$

The Estimation Problem

Define a vector of outcomes for the 10 trials,

$$x = \{x_1, x_2, \ldots, x_{10}\}$$

One possible result of the 10 trials is the vector

$$x \;=\; \{0, 1, 1, 0, 0, 0, 1, 0, 1, 0\}$$

The Estimation Problem

Define a vector of outcomes for the 10 trials,

$$x = \{x_1, x_2, \ldots, x_{10}\}$$

One possible result of the 10 trials is the vector

$$x = \{0, 1, 1, 0, 0, 0, 1, 0, 1, 0\}$$

We now address the problem of estimating the value of the parameter p based on this data.

Maximum Likelihood Estimation

One approach to this problem is the following:

For the *estimate* $\hat{p}$ of the *parameter* p, choose the value that *maximizes the probability of obtaining the outcome that actually occurred*

Maximum Likelihood Estimation

One approach to this problem is the following:

For the *estimate* $\hat{p}$ of the *parameter* p, choose the value that *maximizes the probability of obtaining the outcome that actually occurred*

This approach is called the *method of maximum likelihood*.

Maximum Likelihood Estimation

Define a vector of 10 random variables

$$X = \{X_1, X_2, \ldots, X_{10}\}$$

Maximum Likelihood Estimation

Define a vector of 10 random variables

$$X = \{X_1, X_2, \ldots, X_{10}\}$$

Now define the **likelihood function** of the vector of outcomes x as

$$L(p) = \prod_{i=1}^{10} P(X_i = x_i)$$

Maximum Likelihood Estimation

Define a vector of 10 random variables

$$X = \{X_1, X_2, \ldots, X_{10}\}$$

Now define the **likelihood function** of the vector of outcomes x as

$$L(p) = \prod_{i=1}^{10} P(X_i = x_i)$$

Note that the likelihood function depends on the vector of outcomes and the parameter p.

Maximum Likelihood

For the vector of outcomes:

$$x = \{0, 1, 1, 0, 0, 0, 1, 0, 1, 0\}$$

the likelihood function is:

$$L(p) = [P(X_1 = 0)]\,[P(X_2 = 1)]\,[P(X_3 = 1)] \cdots [P(X_{10} = 0)]$$

Maximum Likelihood

For the vector of outcomes:

$$x = \{0, 1, 1, 0, 0, 0, 1, 0, 1, 0\}$$

the likelihood function is:

$$L(p) = [P(X_1 = 0)]\,[P(X_2 = 1)]\,[P(X_3 = 1)] \cdots [P(X_{10} = 0)]$$

Since $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$, $L(p)$ is

$$(1-p) \cdot p \cdot p \cdot (1-p) \cdot (1-p) \cdot (1-p) \cdot p \cdot (1-p) \cdot p \cdot (1-p)$$

Maximum Likelihood

For the vector of outcomes:

$$x = \{0, 1, 1, 0, 0, 0, 1, 0, 1, 0\}$$

the likelihood function is:

$$L(p) = [P(X_1 = 0)]\,[P(X_2 = 1)]\,[P(X_3 = 1)] \cdots [P(X_{10} = 0)]$$

Since $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$, $L(p)$ is

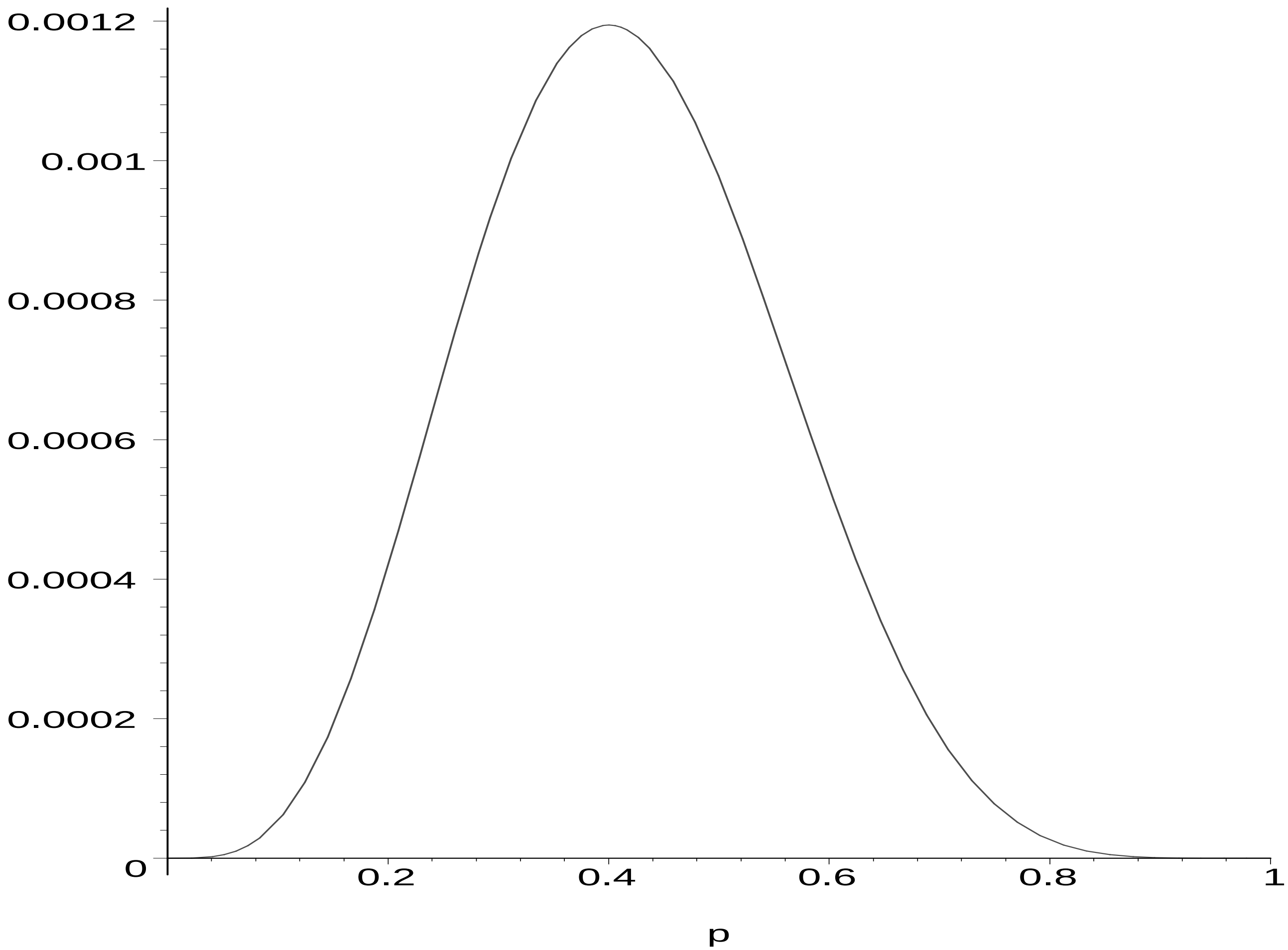
$$(1-p) \cdot p \cdot p \cdot (1-p) \cdot (1-p) \cdot (1-p) \cdot p \cdot (1-p) \cdot p \cdot (1-p)$$

Or, collecting like factors,

$$L(p) = p^4 (1-p)^6$$

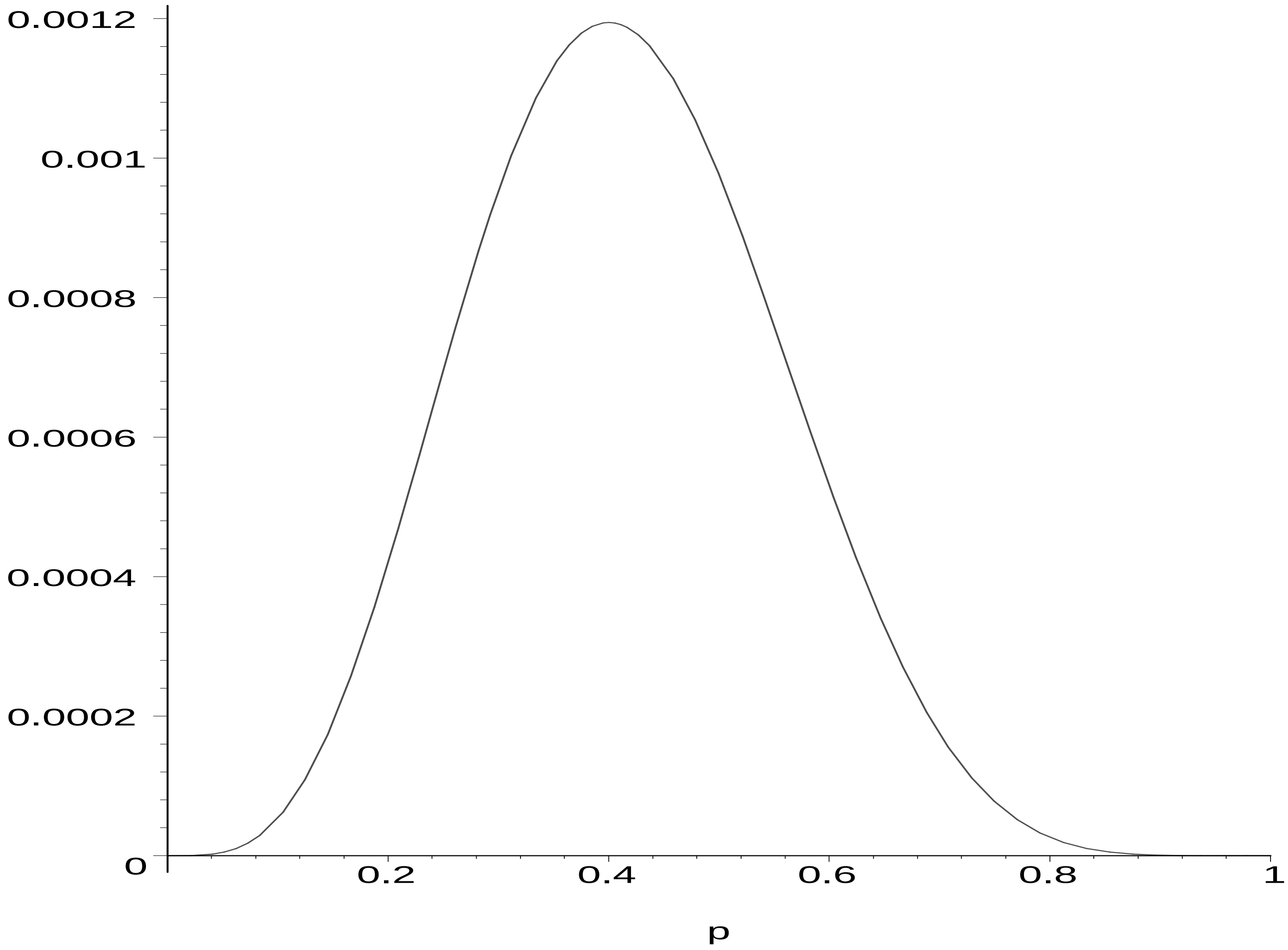
Maximum Likelihood

The graph of $L(p)$ for values of p between 0 and 1 is:



Maximum Likelihood

It appears that $L(p)$ is maximized for some value of p near 0.4:



Maximum Likelihood

To determine the value of p that maximizes $L(p)$, first evaluate

$$\frac{d}{dp}\, p^4(1-p)^6$$

Maximum Likelihood

To determine the value of p that maximizes $L(p)$, first evaluate

$$\frac{d}{dp}\, p^4(1-p)^6$$

Using the product rule, this is

$$4p^3(1-p)^6 + 6p^4(1-p)^5(-1)$$

Maximum Likelihood

To determine the value of p that maximizes $L(p)$, first evaluate

$$\frac{d}{dp}\, p^4(1-p)^6$$

Using the product rule, this is

$$4p^3(1-p)^6 + 6p^4(1-p)^5(-1)$$

Collecting factors, and equating the result to zero, we get

$$p^3(1-p)^5\left[4(1-p)-6p\right] \;=\; p^3(1-p)^5\,(4-10p) \;=\; 0$$

Maximum Likelihood

For $p \in (0, 1)$, the equation

$$p^3(1-p)^5\,(4-10p) \;=\; 0$$

is satisfied when

$$4 - 10p \;=\; 0 \quad \text{so} \quad p = \frac{4}{10}$$

Maximum Likelihood

For $p \in (0,1)$, the equation

$$p^3(1-p)^5\,(4-10p) \;=\; 0$$

is satisfied when

$$4-10p \;=\; 0 \quad \text{so} \quad p = \frac{4}{10}$$

So, the maximum likeklihood estimate of p is:

$$\hat{p} = \frac{4}{10}$$

Maximum Likelihood

We can generalize this result to a sequence of n trials that produces k successes.

For $1 < k < n-1$, the likelihood function is maximized when

$$p^{(k-1)}(1-p)^{(n-k-1)}\,(k-np) \;=\; 0$$

is satisfied when

$$k - np \;=\; 0 \quad \textbf{so} \quad \hat{p} = \frac{k}{n}$$

Maximum Likelihood

With the obvious extensions to include the cases $k = 0$ and $k = n$, we have the maximum likelihood estimate

$$\hat{p} = \frac{k}{n}$$

Maximum Likelihood

With the obvious extensions to include the cases $k = 0$ and $k = n$, we have the maximum likelihood estimate

$$\hat{p} = \frac{k}{n}$$

Now we can state the following result:

If a series of n Bernoulli trials produces k successes, the maximum likelihood estimate of the parameter p is

$$\hat{p} \; = \; \frac{k}{n}$$

Maximum Likelihood

Example 2: Exponential distribution

In this case, the density function is

$$f(x) = \lambda e^{-\lambda x}$$

Maximum Likelihood

Example 2: Exponential distribution

In this case, the density function is

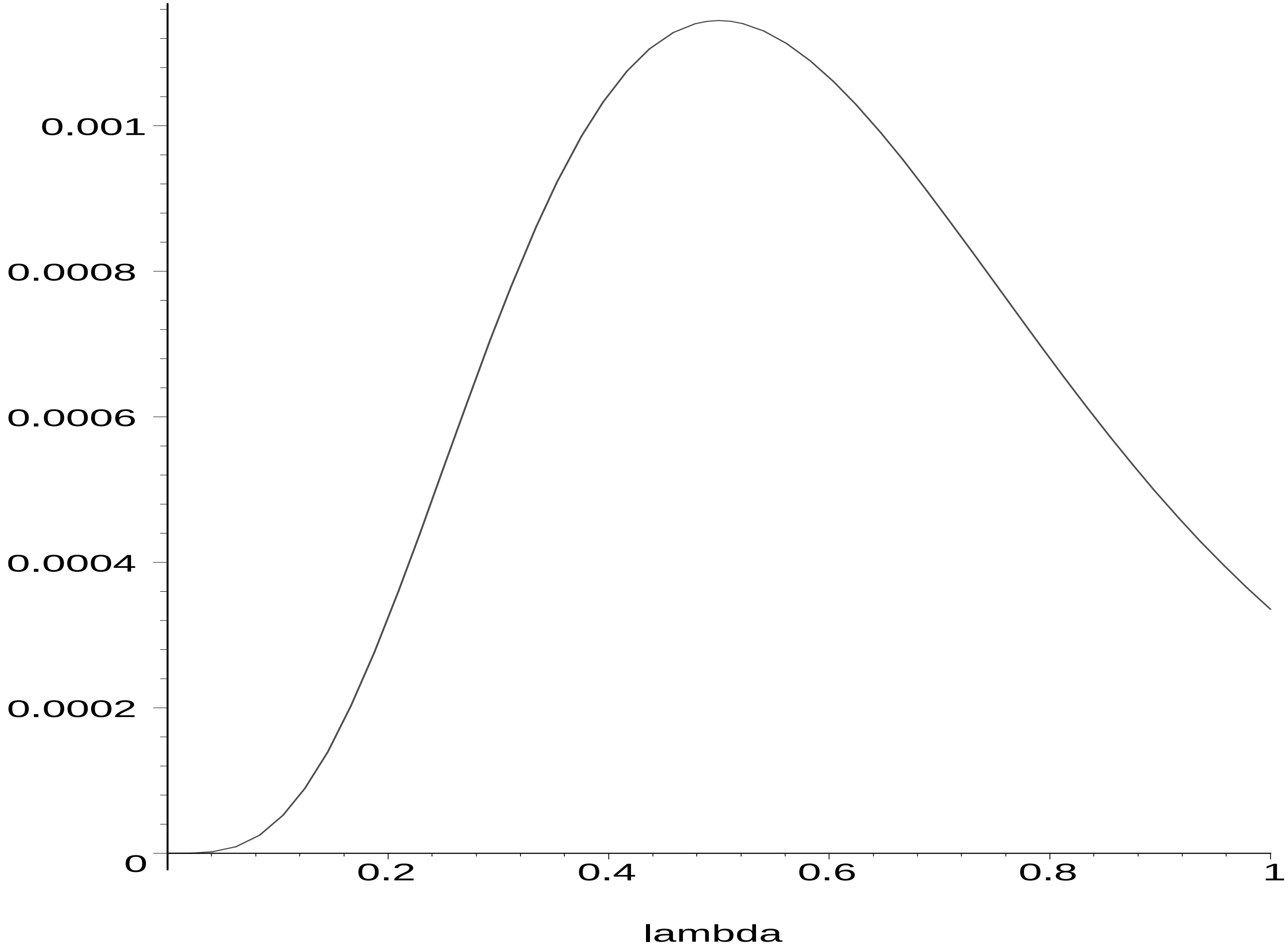
$$f(x) = \lambda e^{-\lambda x}$$

For a sample $x = \{x_1, x_2, \ldots, x_n\}$, the likelihood function is:

$$L(\lambda) \; = \; \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \; = \; \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i\right)$$

Maximum Likelihood

The graph of $L(\lambda)$ for values of λ is:



Maximum Likelihood

Differentiating the likelihood function with respect to λ and setting the result to zero, on solving the result for $lambda$ we have the maximum likelihood estimate

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}$$

for the parameter of the exponential distribution

Maximum Likelihood

Differentiating the likelihood function with respect to λ and setting the result to zero, on solving the result for $lambda$ we have the maximum likelihood estimate

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}$$

for the parameter of the exponential distribution

Maximum Likelihood

Example 3: Geometric distribution

In this case, the density function is

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

Maximum Likelihood

Example 3: Geometric distribution

In this case, the density function is

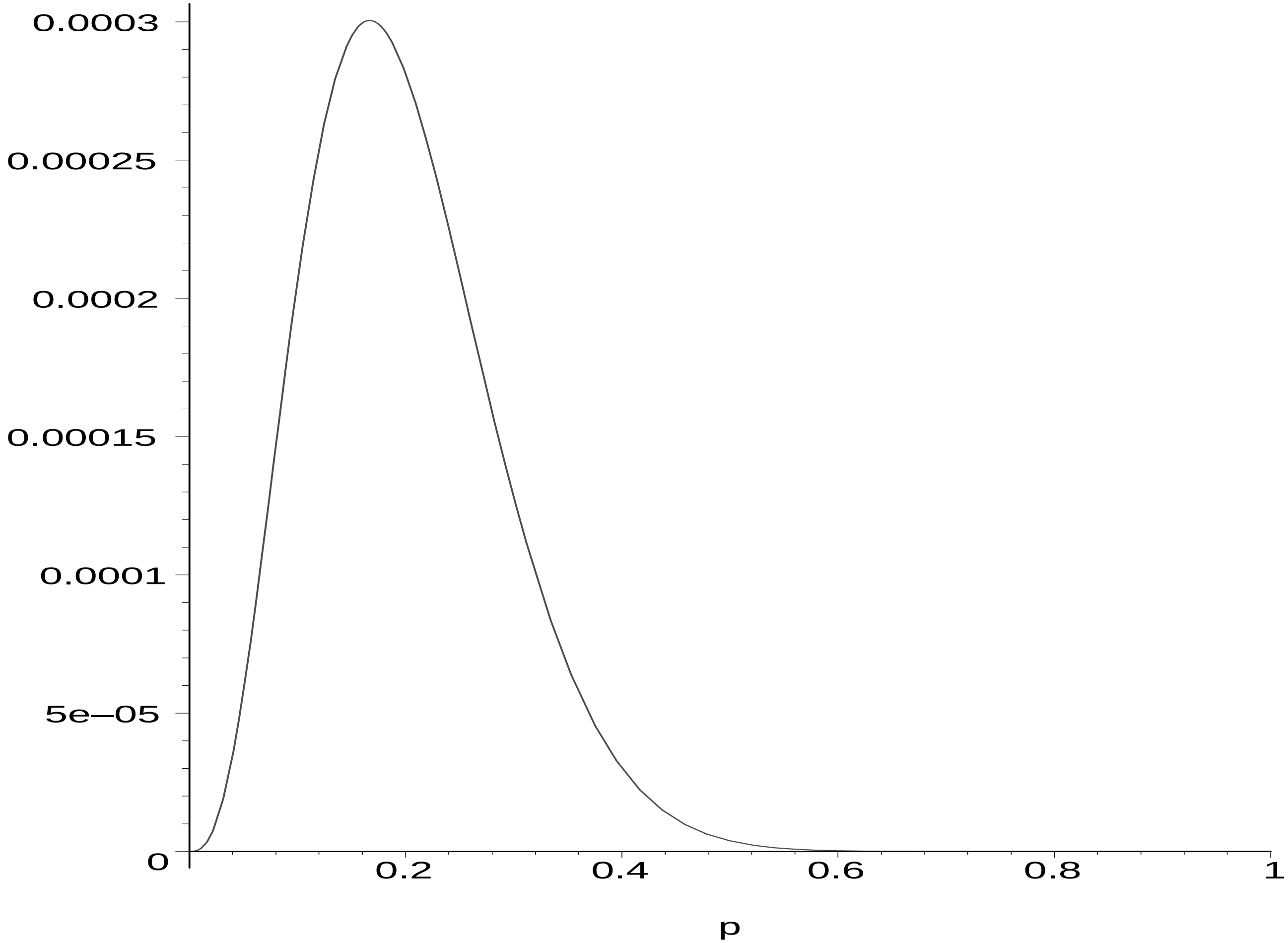
$$f(x) = (1-p)^{n-1}p$$

For a sample $x = \{x_1, x_2, \dots, x_n\}$, the likelihood function is:

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

Maximum Likelihood

The graph of $L(p)$ for values of p is:



Maximum Likelihood

Differentiating the likelihood function with respect to p and setting the result to zero, on solving the result for p we have the maximum likelihood estimate

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} x_i}$$

for the parameter of the geometric distribution.

Maximum Likelihood

Differentiating the likelihood function with respect to p and setting the result to zero, on solving the result for p we have the maximum likelihood estimate

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} x_i}$$

for the parameter of the geometric distribution.

Maximum Likelihood

Example 4: Poisson distribution

In this case, the density function is

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Maximum Likelihood

Example 4: Poisson distribution

In this case, the density function is

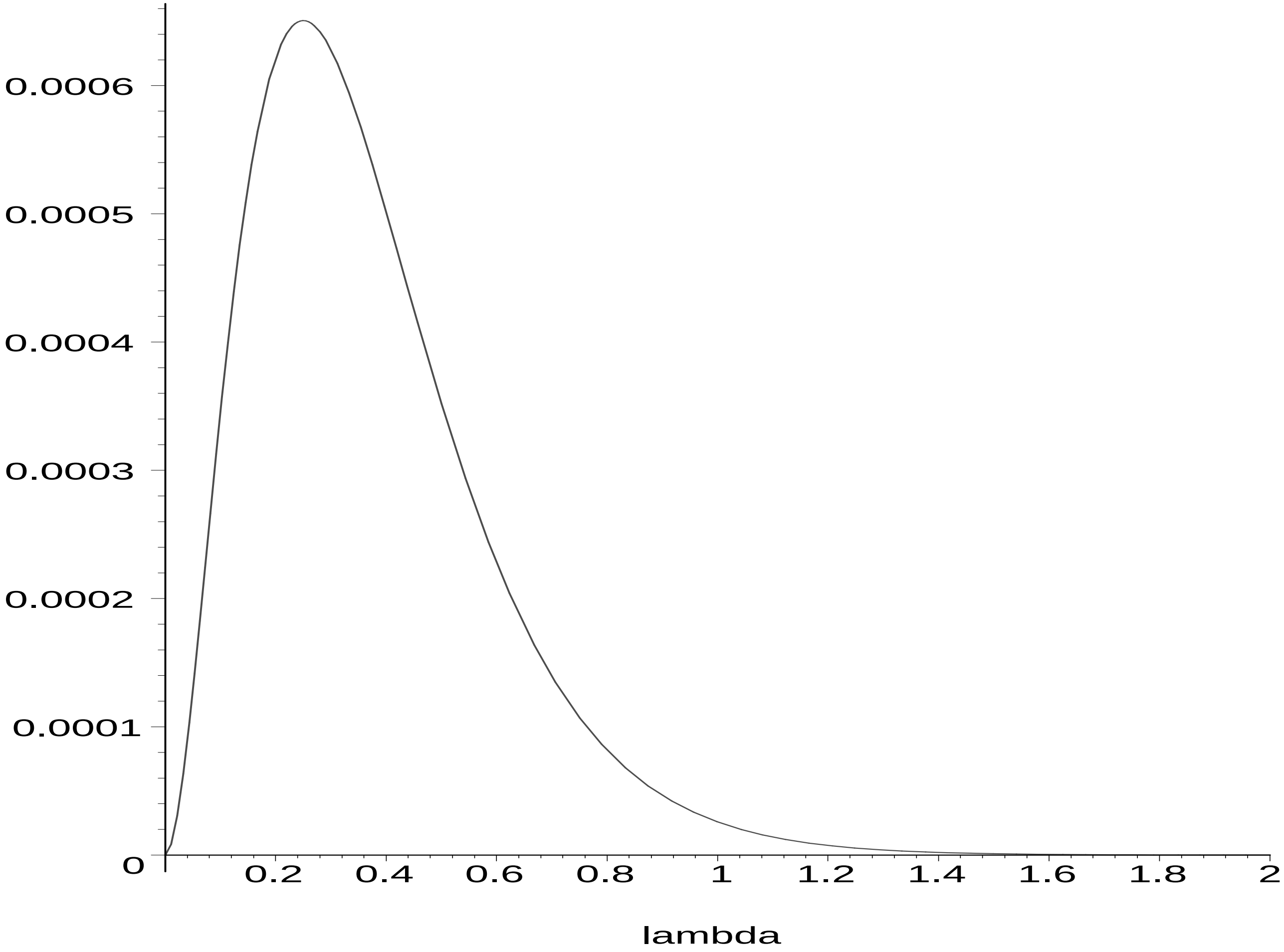
$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

For a sample $x = \{x_1, x_2, \ldots, x_n\}$, the likelihood function is:

$$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum x_i} e^{-n\lambda}}{\prod x_i!}$$

Maximum Likelihood

The graph of $L(\lambda)$ for values of λ is:



Maximum Likelihood

Differentiating the likelihood function with respect to λ and setting the result to zero, on solving the result for λ we have the maximum likelihood estimate

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$$

for the parameter of the Poisson distribution.

Maximum Likelihood

Differentiating the likelihood function with respect to λ and setting the result to zero, on solving the result for λ we have the maximum likelihood estimate

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$$

for the parameter of the Poisson distribution.

Multiple Parameters

Many important distributions have more than one parameter.

Recall that if a random variable X is normally distributed, $X \sim N(\mu, \sigma^2)$, the density function is:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

Multiple Parameters

For a sample $x = \{x_1, x_2, \ldots, x_n\}$ from a $N(\mu, \sigma^2)$ population, the likelihood function is:

$$L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

$$= \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2\right)$$

Multiple Parameters

To maximize $L(\mu, \sigma)$, recall that we set up a system of simultaneous equations in the partial derivatives with respect to μ and σ,

$$\frac{\partial L(\mu, \sigma)}{\partial \mu} = 0$$

$$\frac{\partial L(\mu, \sigma)}{\partial \sigma} = 0$$

Multiple Parameters

To maximize $L(\mu, \sigma)$, recall that we set up a system of simultaneous equations in the partial derivatives with respect to μ and σ,

$$\frac{\partial L(\mu, \sigma)}{\partial \mu} = 0$$

$$\frac{\partial L(\mu, \sigma)}{\partial \sigma} = 0$$

The complexity of this type of system is one of the drawbacks of maximum likelihood estimation.

Method of Moments

An alternative to the maximum likelihood technique for estimating parameter values is the **method of moments**.

Method of Moments

An alternative to the maximum likelihood technique for estimating parameter values is the **method of moments**.

The idea is that in general, the (theoretical) moments of a random variable X,

$$E(X^k) = \int_{-\infty}^{\infty} x^k f_X(x)\, dx$$

are functions of the unknown parameters.

Method of Moments

If $x = \{x_1, x_2, \ldots, x_n\}$ is a random sample from some population, define the k^{th} **sample moment** as:

$$\frac{1}{n}\sum_{i=1}^{n} x_i^k$$

Method of Moments

If $x = \{x_1, x_2, \ldots, x_n\}$ is a random sample from some population, define the k^{th} **sample moment** as:

$$\frac{1}{n}\sum_{i=1}^{n} x_i^k$$

The method of moments estimates are obtained by setting the sample moments equal to the theoretical moments and solving for the parameters.

$$E(X^k) \; = \; \int_{-\infty}^{\infty} x^k f_X(x)\, dx \; = \; \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

Method of Moments

Example Using moment-generating functions, we have seen that if $x = \{x_1, x_2, \ldots, x_n\}$ is a random sample from a $N(\mu, \sigma^2)$ population, the sample mean has a normal distribution with mean μ and variance σ^2/n.

In this case, the (theoretical) first moment, $E(X)$, is just μ.

Method of Moments

Example Using moment-generating functions, we have seen that if $x = \{x_1, x_2, \ldots, x_n\}$ is a random sample from a $N(\mu, \sigma^2)$ population, the sample mean has a normal distribution with mean μ and variance σ^2/n.

In this case, the (theoretical) first moment, $E(X)$, is just μ.

The method of moments estimate of μ is obtained from the equation

$$E(X^1) = \mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Method of Moments

So, for a random sample from a $N(\mu, \sigma^2)$ population, the method of moments extimate of μ is:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$