Applied Probability II

Section 7: Estimation

Professor Caroline Brophy

Semester 2, 2020-21

Applied Probability II. Section 7: Estimation

Section 7.1: Recap of standard probability laws

The distribution game!

What is the distribution of the random variable in each of the following?

- Suppose you are playing a game of Ludo and so you need to roll a 6 to get your first counter out of home and onto the board. What is the probability that you will get onto the board in between 4 and 6 goes?
- A recent study has shown that 30% of women in a certain large population suffer from anemia (iron deficiency). A random sample of 8 women is taken from the population and tested. What is the probability that three or more of the women in the sample are anemic?
- A basketball player scores a basket from the free throw line with probability 0.45. You start observing a training session on free throw attempts at a random point. What is the probability that her third basket occurs on the sixth shot?
- Suppose that items in a vending machine get stuck on the way out at random with probability 0.03. You arrive at the vending machine, put you money in and select your item. What is the probability you end up not getting your item (and are really annoyed!)?
- Consider a lotto draw with 45 balls where 6 balls are chosen at random. What is the probability of matching 4 balls?

Bernoulli

Consider an experiment with two outcomes, success and failure. Such an experiment is known as a Bernoulli trial.

Let X be a Bernoulli random variable.

$$X \sim \text{Bernoulli}(p)$$

X can take values: 0, 1, where 1 is defined as a 'success'.

Parameter: p, where $0 < p < 1$ and is the probability of success in a trial.

Probability mass function (pmf): $P(X = x) = p^x(1-p)^{1-x}$

Mean: $\mathrm{E}[X] = p$

Variance: $\mathrm{Var}[X] = p(1-p)$

Example

Suppose that items in a vending machine get stuck on the way out at random with probability 0.03. You arrive at the vending machine, put you money in and select your item. What is the probability you end up not getting your item (and are really annoyed!)?

Binomial

Bernoulli random variables provide the building blocks for defining other discrete random variables. An experiment which consists of n repeated independent Bernoulli trials, each with probability of success p, is called a **binomial** experiment.

Let X be the total number of successes in a binomial experiment with n trials. Then X is a binomial random variable with parameters n and p.

$$X \sim \text{Binomial}(n, p)$$

X can take values: 0, 1, 2, . . . , n.

Parameter: $n = 1, 2, 3...$, and p, where $0 < p < 1$ and is the probability of success in an individual trial.

Probability mass function (pmf): $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$

Mean: $\mathrm{E}[X] = np$

Variance: $\mathrm{Var}[X] = np(1 - p)$.

Example

A recent study has shown that 30% of women in a certain large population suffer from anemia (iron deficiency). A random sample of 8 women is taken from the population and tested. What is the probability that three or more of the women in the sample are anemic?

Geometric

Suppose independent Bernoulli trials with success probability p are performed until a success occurs. Let X be the number of trials required. Then

$$X \sim \text{Geometric}(p)$$

X can take values: $1, 2, 3, \ldots$.

Parameter: p, where $0 < p < 1$ and is the probability of success in an individual trial.

Probability mass function (pmf): $P(X = x) = (1 - p)^{x-1} p$

Mean: $\mathrm{E}[X] = \frac{1}{p}$.

Variance: $\mathrm{Var}[X] = \frac{(1-p)}{p^2}$.

Example

Suppose you are playing a game of Ludo and so you need to roll a 6 to get your first counter out of home and onto the board. What is the probability that you will get onto the board in between 4 and 6 goes?

Negative binomial

The negative binomial distribution is a generalisation of the geometric distribution.

Suppose independent Bernoulli trials with success probability p are performed until r successes occurs. Let X be the number of trials required.

Then $X \sim$ Negative binomial (r, p).

The probability mass function for X is:

$$P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

for $k = r, r+1, r+2, \dots$

Example

A basketball player scores a basket from the free throw line with probability 0.45. You start observing a training session on free throw attempts at a random point. What is the probability that her third basket occurs on the sixth shot?

Hypergeometric

Suppose a box contains N items, A of which are of type A and $N - A$ of which are of type B. A sample of size n is taken without replacement.

Let X denote the number of type A items drawn.

Then X is a hypergeometric random variable with parameters N, A and n.

The probability mass function is

$$p(k) = P(X = k) = \frac{\binom{A}{k}\binom{N-A}{n-k}}{\binom{N}{n}}$$

for $k = 0, 1, 2, \ldots, \min(A, n)$.

Example

Consider a lotto draw with 45 balls where 6 balls are chosen at random. What is the probability of matching 4 balls?

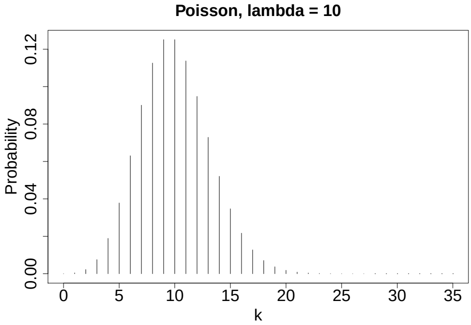
Poisson

The Poisson is a 'counting' distribution. A random variable X has a Poisson distribution with parameter λ if

$$p(k) = P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

for $k = 0, 1, 2, \ldots$ and $\lambda > 0$.

We say that $X \sim \text{Poisson}(\lambda)$. $P(X = k)$ is the probability mass function.

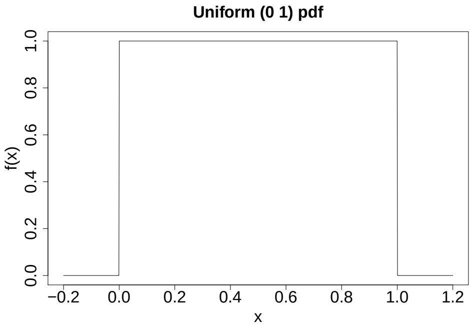


Uniform

X is said to have a uniform distribution on the interval (a, b) if

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

$$F(x) = P(X \leq x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x < b \\ 1 & x \geq b \end{cases}$$



Exponential

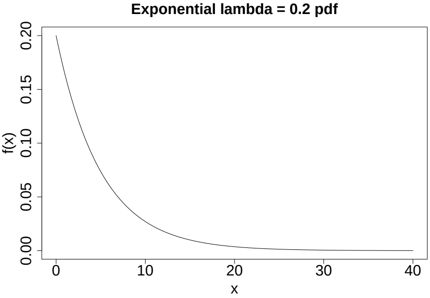
A continuous random variable whose probability density function (pdf) is given by:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

with $\lambda > 0$, is said to have an exponential distribution.

The exponential distribution is often used to model the time until a specific event occurs. The cumulative distribution function (cdf) for an exponential random variable is:

$$F(t) = P(X \leq t) = \int_0^t \lambda e^{-\lambda x} dx = -e^{-\lambda x}\Big|_0^t = 1 - e^{-\lambda t}, t \geq 0.$$

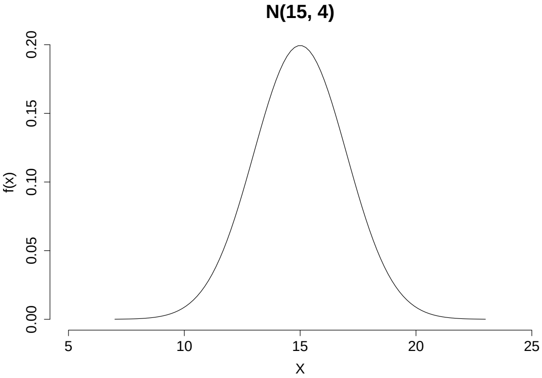


Normal

A random variable X is said to have a normal (or Gaussian) distribution with parameters μ and σ^2 if it has the probability density function (pdf):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

For example, consider $X \sim \mathrm{N}(\mu = 15, \sigma^2 = 4)$.



Section 7.2: Likelihood

Binomial example

Think about tossing a coin 10 times, but where we know nothing about the fairness of the coin. The outcome from each toss is a Bernoulli distributed random variable with parameter p.

The number of heads we see is: $X \sim \text{Binomial}(n = 10, p)$.

Suppose that we observe $x = 8$ heads.

We don't know what the value of p is, but what can we say about p from our own experiment?

- Information about p is not complete, so there is uncertainty.
- p cannot be zero and is unlikely to be small as then $\mathrm{P}(X = 8)$ would be tiny.
- Likely values for p are 0.7, 0.8 or 0.9, because:
  - if $p = 0.7$, then $\mathrm{P}(X = 8) = 0.2335$, and
  - if $p = 0.8$, then $\mathrm{P}(X = 8) = 0.3020$, and
  - if $p = 0.9$, then $\mathrm{P}(X = 8) = 0.1937$.

A way to compare candidate values of p is to compare the observed probabilities across different values for p. We can do this formally via a likelihood function.

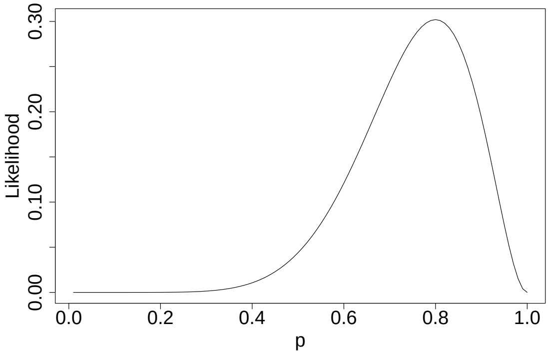
Binomial example contd.

In our experiment, we observe $x = 8$ heads in 10 trials.

The likelihood function for p is

$$L(p) = P(X = 8 \mid p) = \binom{10}{8} p^8 (1-p)^2$$

A plot of $L(p)$ versus p is:



Definition of likelihood

Definition: Assuming a statistical model is parametrised by a fixed and unknown parameter θ, the likelihood is the probability of the observed data x considered as a function of θ.

Back to the coin experiment

Suppose we repeat the coin experiment three times:

First run, observe $x_1 = 8$, second run: $x_2 = 6$, third run: $x_3 = 7$.

Now the likelihood is the probability of observing $\mathbf{x} = (x_1, x_2, x_3)$.

Assuming the experiments are independent, the probability of observing all three is

$$P(X = 8)P(X = 6)P(X = 7)$$

where, $X \sim \text{Binomial}(n = 10, p)$.

The likelihood function is

$$L(p) = P(X = x_1)P(X = x_2)P(X = x_3) = \prod_{i=1}^{3} P(X = x_i)$$

We know that $P(X = x) = \binom{10}{x} p^x (1-p)^{10-x}$. So,

$$L(p) = \prod_{i=1}^{3} P(X = x_i) = \prod_{i=1}^{3} \binom{10}{x_i} p^{x_i} (1-p)^{10-x_i}$$

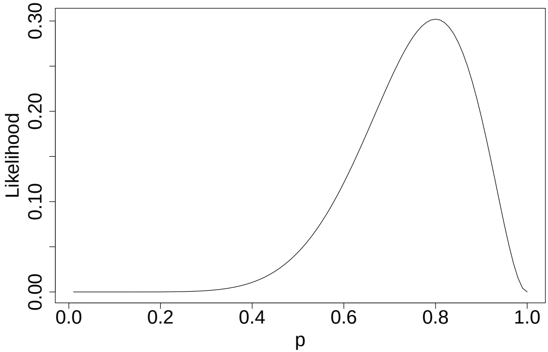
$$= \left[p^{\sum_{i=1}^{3} x_i}\right] \left[(1-p)^{30-\sum_{i=1}^{3} x_i}\right] \left[\prod_{i=1}^{3} \binom{10}{x_i}\right]$$

Section 7.3: Maximum likelihood estimation

How does maximum likelihood (ML) estimation work?

If we have an unknown parameter θ in a statistical model, the maximimum likelihood estimate (MLE) $\hat{\theta}$ is the value of θ which maximises $L(\theta)$.

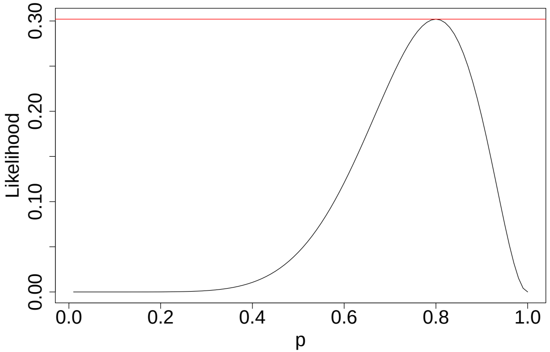
Think back to the coin example earlier, when we observed $x = 8$ heads from 10 independent trials.



How does maximum likelihood (ML) estimation work?

If we have an unknown parameter θ in a statistical model, the maximimum likelihood estimate (MLE) $\hat{\theta}$ is the value of θ which maximises $L(\theta)$.

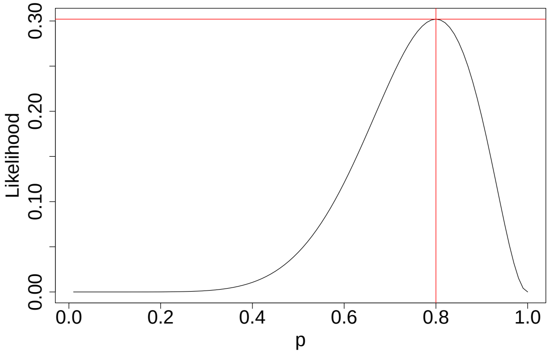
Think back to the coin example earlier, when we observed $x = 8$ heads from 10 independent trials.



How does maximum likelihood (ML) estimation work?

If we have an unknown parameter θ in a statistical model, the maximimum likelihood estimate (MLE) $\hat{\theta}$ is the value of θ which maximises $L(\theta)$.

Think back to the coin example earlier, when we observed $x = 8$ heads from 10 independent trials.



How does maximum likelihood (ML) estimation work?

In general, based on a sample of size n independent observations $x_1, x_2, \ldots, x_n$, the likelihood function can be written as

$$L(\theta) = \prod_{i=1}^{n} P(X = x_i \mid \theta) \qquad \text{for } X \text{ for discrete}$$

$$L(\theta) = \prod_{i=1}^{n} f_X(x_i \mid \theta) \qquad \text{for } X \text{ for continuous}$$

Effectively we want to maximise $L(\theta)$ with respect to θ.

It is often much easier to work with the log (natural log) of the likelihood function $l(\theta)$.

$$l(\theta) = logL(\theta)$$

Maximising $l(\theta)$ with respect to θ and taking $\hat{\theta}$ as the maximiser, gives the MLE.

Example 1 Poisson - likelihood function

The number of cars arriving at a car park per hour from 9 to 10am is assumed to be Poisson with rate λ. The number of cars from 9-10am on six randomly selected days were: 50, 47, 82, 91, 46, 64.

We have a random sample of $X_1, X_2, ..., X_6$ from a Poisson(λ) distribution.

Remember, $P(X_i = k) = \frac{\lambda^k e^{-\lambda}}{k!}$

The likelihood function for λ is:

$$\begin{aligned} L(\lambda) &= P(X_1 = 50)P(X_2 = 47)P(X_3 = 82)P(X_4 = 91)P(X_5 = 46)P(X_6 = 64) \\ &= \left(\frac{\lambda^{50} e^{-\lambda}}{50!}\right)\left(\frac{\lambda^{46} e^{-\lambda}}{46!}\right)\cdots\left(\frac{\lambda^{64} e^{-\lambda}}{64!}\right) \end{aligned}$$

Or more generically:

$$L(\lambda) = \prod_{i=1}^{6} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum_{i=1}^{6} x_i} e^{-6\lambda}}{\prod_{i=1}^{6} x_i!}$$

Example 1 Poisson - MLE

What is the maximimum likelihood estimate for λ?

Starting with the likelihood function:

$$L(\lambda) = \frac{\lambda^{\sum_{i=1}^{6} x_i} e^{-6\lambda}}{\prod_{i=1}^{6} x_i!}$$

We take the log of the likelihood function:

$$\begin{aligned} log(L(\lambda)) = l(\lambda) &= log\left(\frac{\lambda^{\sum_{i=1}^{6} x_i} e^{-6\lambda}}{\prod_{i=1}^{6} x_i!}\right) \\ &= log(\lambda^{\sum_{i=1}^{6} x_i}) + log(e^{-6\lambda}) - log(\prod_{i=1}^{6} x_i!) \\ &= \left(\sum_{i=1}^{6} x_i\right) log(\lambda) - 6\lambda - \sum_{i=1}^{6} log(x_i!) \end{aligned}$$

Example 1 Poisson - MLE contd.

$$l(\lambda) = \left(\sum_{i=1}^{6} x_i\right) log(\lambda) - 6\lambda - \sum_{i=1}^{6} log(x_i!)$$

We maximise $l(\lambda)$ with respect to λ, by first differentiating:

$$\frac{dl}{d\lambda} = \frac{\sum_{i=1}^{6} x_i}{\lambda} - 6$$

And then setting equal to 0 and evaluating at $\hat{\lambda}$:

$$\frac{\sum_{i=1}^{6} x_i}{\hat{\lambda}} - 6 = 0$$

$$\hat{\lambda} = \frac{\sum_{i=1}^{6} x_i}{6} = \bar{x} = 63.33$$

Our MLE is $\hat{\lambda} = 63.33$.

Example 2 exponential - likelihood function

Consider the time to failure for a piece of equipment. The observed times to failure are: 30.4, 7.8, 1.4, 13.1 and 67.3 hours.

Assume the lifetime follows an exponential distribution with parameter λ.

Let X be the lifetime, then

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

The likelihood function is

$$L(\lambda) = f_X(x_1) f_X(x_2) f_X(x_3) f_X(x_4) f_X(x_5)$$

$$= (\lambda e^{-\lambda 30.4})(\lambda e^{-\lambda 7.8})(\lambda e^{-\lambda 1.4})(\lambda e^{-\lambda 13.1})(\lambda e^{-\lambda 67.3})$$

Or more generically:

$$L(\lambda) = \prod_{i=1}^{5} \lambda e^{-\lambda x_i} = \lambda^5 e^{-\lambda \sum_{i=1}^{5} x_i}$$

Example 2 exponential - MLE

Take the log of the likelihood function:

$$log(L(\lambda)) = l(\lambda) = log(\lambda^5 e^{-\lambda \sum_{i=1}^{5} x_i})$$
$$= log(\lambda^5) + log(e^{-\lambda \sum_{i=1}^{5} x_i}) = 5 log\lambda - \lambda \sum_{i=1}^{5} x_i$$

Maximise $l(\lambda)$ with respect to λ, by first differentiating:

$$\frac{dl}{d\lambda} = \frac{5}{\lambda} - \sum_{i=1}^{5} x_i$$

And then setting equal to 0 and evaluating at $\hat{\lambda}$:

$$\frac{5}{\hat{\lambda}} - \sum_{i=1}^{5} x_i = 0$$
$$\frac{1}{\hat{\lambda}} = \frac{\sum_{i=1}^{5} x_i}{5} = \bar{x}$$
$$\hat{\lambda} = \frac{1}{\bar{x}} = 0.042 \qquad \text{this is our MLE}$$

Final word on maximum likelihood estimation

- Widely used and powerful tool.

- Here we have focused on some simple examples, but ML estimation is also very useful in estimating complex models.