The problem of econometric modeling

Susan Thomas
IGIDR, Bombay

August 5, 2008

Defining the economists problem

- The real world contains economic variables, Y.
- The job of an economist is to:
  1. Understand what drives the behaviour of the variables.
  2. Forecast what is going to happen next
- Economists state hypotheses in the form of a model:

$$\begin{aligned} Y_t &= f(\tilde{I}_{t-1}) + \epsilon_t \\ I_{t-1} &= Y_{t-1}, Y_{t-2}, \ldots, \vec{X_{t-1}}, \vec{X_{t-2}}, \ldots, t \end{aligned}$$

- The hypothesis is a mathematical description of the behaviour of Y, with the description of:
  - f, or the functional form of the model
  - $\tilde{I}_{t-1}$, or what a subset of I_{t-1} is required to understand Y_t.
  - Distribution of what is unknown – called the error term, ϵ.

Defining the econometric problem

- Start with a hypothesis.
- Take real world data on $\vec{Y}, \vec{I}$.
- Using the sample of data, econometricians need to
  1. Estimate the model.
  2. Validate the model.

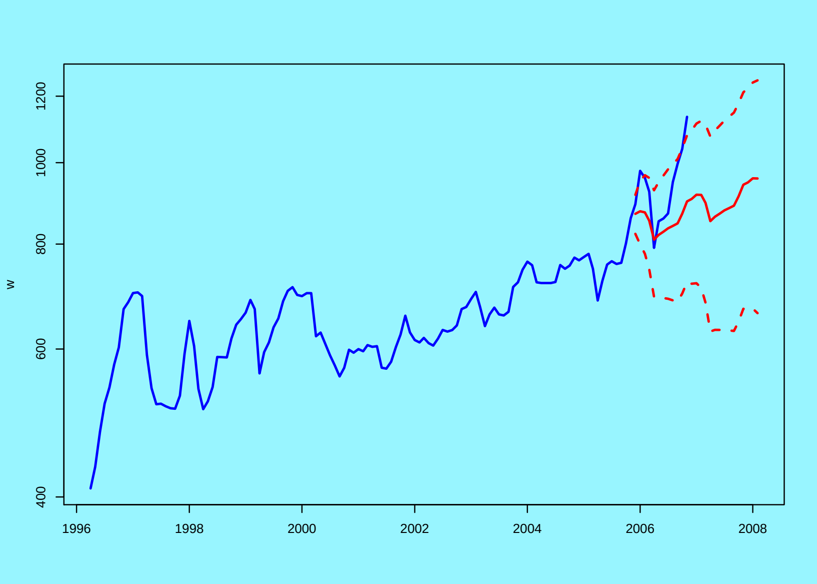
Estimating models

- The generic functional form, f is well defined if it specifies:
  1. A well-defined vector of input variables, I_{t-1}, called the **explanatory variables**.
  2. Weights on each of the input variables called the **coefficients**.
  3. The distribution of the error term.
     This includes the **type** of the distribution and the **parameters** of the distribution.
- The estimation step involves finding the values for the coefficients as well as the distribution parameters.

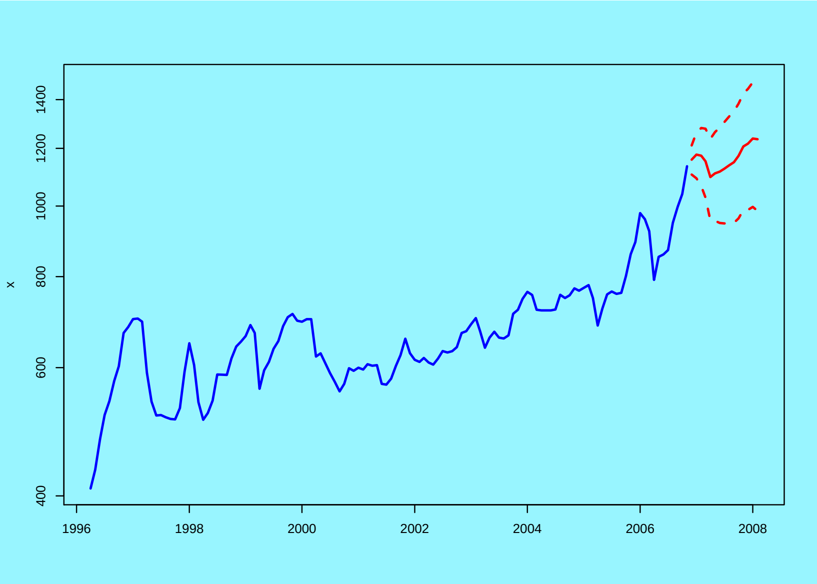
Validating models

- A well developed hypothesis includes expected values for the model "parameters".
- Validation, thus, involves establishing that:
  - f is the best defined function.
  - $\tilde{I}$ is the best subset required to describe Y.
  - That the errors are indeed distributed as the hypothesis.
- This is done by establishing that the model and the real world agree on:
  - the sensitivity of Y to changes in the values of $\tilde{I}$
  - forecasted values of Y from the model and observed values from the real world.
- Complication on the problem: we have $\vec{Y}, \vec{I}$ as sample observations. It is not the whole set of observations.

Example: Forecasting price of Indian wheat, case I



Example: Forecasting price of Indian wheat, case II

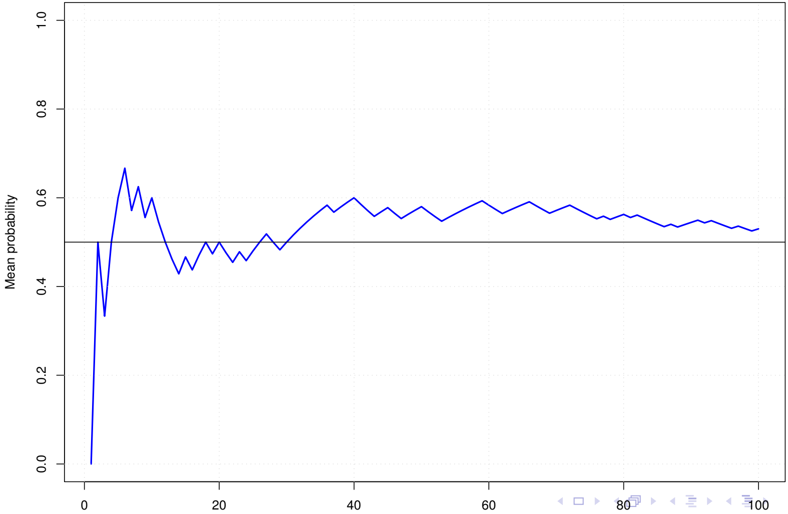


Recap on probability

Definitions in probability

- Every observation is an **outcome** or an **event**.
  Eg., an experiment: does a rat die after having been injected with the bird flu virus.
- **Event space**: the set of all possible events that could happen.
  Eg., the rat can either die or not.
- **Probability**: how frequently does the event appear compared to all the other events in the event space occurs.
  Eg., in the toss of a fair coin, probability of heads is 0.5.
- Problem: we always work with samples. Therefore, we always have to simulate or model probabilities.
  Eg., the probability of a 50% chance of heads in a coin toss is known because it has been empirically calculated by tossing a coin a very large number of times.

Results from a simulation of 100 tosses of a coin



Definitions and properties

- **Mutually exclusive** events cannot take place together.
- The probability of two mutually exclusive events happening together is 0.
- The probability of two of these events taking place is the sum of their probabilities.
- **Conditional probability** is the probability of one event happening given that another event has already happened.

The joint distribution of X and Y

- Suppose there are two random outcomes, X and Y.
- $X = 0, 1, 2$. $Y = 0, 1$.
- The **joint distribution** $\Pr(X = x, Y = y)$ shows all probabilities of the **events** of the combination of the two that can come about.
- These correspond to statements $\Pr((X = x) \text{and} (Y = y))$.

Joint distribution: $\Pr(X = x, Y = y)$

	Y		
X	0	1	
0	0	0.2	
1	0.1	0.1	
2	0.2	0.4	
			1

- All the cells contain joint probabilities.
- They add up to 1
- This joint pdf is the most you can know about the joint variation of X and Y.

Recovering Pr(X)

	0	1	
0	0	0.2	0.2
1	0.1	0.1	0.2
2	0.2	0.4	0.6
			1

- How to reduce from $\Pr(X = x, Y = y)$ to $\Pr(X)$?
- Addup all the ways in which you can get $X = 2$
- Addup along the rows of the joint to get $\Pr(X)$
- Takes us back to the pdf of X = (0.2, 0.2, 0.6).

Same idea for recovering Pr(Y)

	0	1	
0	0	0.2	
1	0.1	0.1	
2	0.2	0.4	
	0.3	0.7	1

- Suppose we're interested in $\Pr(Y = 1)$.
- Y can be 1 in 3 different ways. Adding up, we get 0.7.
- Similarly, we get $\Pr(Y = 0)$.
- Now we know the full distribution of Y.

“Joint” versus “Marginal” distribution

	0	1	
0	0	0.2	0.2
1	0.1	0.1	0.2
2	0.2	0.4	0.6
	0.3	0.7	1

- The joint distribution contains all knowable facts.
  From the joint, we got the 2 univariate distributions.
- Written in the margins of the table, so the name “marginal” distributions.
- The joint is the fundamental underlying information; the marginals flow from that.

Independence

- If the probability of one event does not depend upon whether another event has taken place, the events are **independent.**
- Using the joint and marginal distributions: X and Y are independent if and only if

$$\Pr(X = x, Y = y) = \Pr(X = x) \cdot \Pr(Y = y)$$

- If A and B are not independent of each other, the Prob(A and B) is

$$Prob(A) * Prob(B|A)$$

Random variables and probability distributions

- A random variable (RV) is what the value of an event is. There are discrete and continuous RV.
- For a discrete RV, the probability distribution (PD) is a table of all the events and their related probabilities.
- For example, in the roll of a die:

Value	Probability	Value	Probability
1	1/6	4	1/6
2	1/6	5	1/6
3	1/6	6	1/6

- A probability distribution will contain **all** the outcomes and their related probabilities, and the probabilities will sum to 1.

How to read a probability distribution

- From the distribution, we can find:
  - X = 3
  $$\Pr(X = 3) = 1/6$$
  - X = even number
  $$\Pr(X = 2 \text{ or } X = 4 \text{ or } X = 6) = 3/6 = 1/2$$
- More interesting, we can also find:
  $$\Pr(X > 3) = 3/6 = 1/2$$
  This is called a **cumulative probability.**

A cumulative probability distribution (CD)

- A table of the probabilities cumulated over the events.

Value	Probability	Value	Probability
X≤1	1/6	X≤4	4/6
X≤2	2/6	X≤5	5/6
X≤3	3/6	X≤6	1

- The CD is a monotonically increasing set of numbers
- The CD always ends with at the highest value of 1.

What is a probability density function?

- The probability density function (PDF) is the PD of a continuous random variable.
- Since continuous random variables are uncountable, it is difficult to write down the probabilities of all possible events.
  Therefore, the PDF is always a function which gives the probability of one event, x.
- If we denote the PDF as function f, then

$$\Pr(X = x) = f(x)$$

- A probability distribution will contain **all** the outcomes and their related probabilities, and the probabilities will sum to 1.

The problem with estimating PDFs

- In a set of continuous random variables, the probability of picking out a value of exactly x is zero.
- We define the Pr(X = x) as the following difference:

$$\Pr(X \leq (x + \Delta)) - \Pr(X \leq x)$$

  as Δ becomes an infinitesimally small number.
- Here, $\Pr(X \leq x)$ is the **cumulative density function** of X.

What is the cumulative density function (CDF)?

- Analogous to the discrete RV case, the CDF is the cumulation of the probability of all the outcomes upto a given value.
- Or, the CDF is the probability that the RV can take any value less than or equal to X.
- If we assume that the RV X can take values from $-\infty$ to ∞, then theoretically,

$$F(X) = \int_{-\infty}^{X} f(x) d(x)$$

Reformulating the PDF in calculus

- Pr(x = X) is given as:

$$\begin{aligned} \Pr(x = X) &= F(X + \Delta) - F(X) \\ &= d(F(x))/d(x) \end{aligned}$$

  for infinitesimally small Δ.
- For continuous RVs, we approach the Pr(x) as the derivative of the CDF.

Examples of a PDF: Uniform RVs

Uniform distribution: The outcome is any number that can take a value between a minimum (A) and a maximum (B) with equal probability.

- For a uniform RV,

$$\Pr(X = x) = 1/(B - A)$$

- The uniform density has two parameters, A, B.

Testing concepts: Uniform density CD

What is the form of the Uniform CDF, given that the maximum = B and minimum = A?

$$
\begin{aligned}
F(X) &= \int_{\text{minimum}}^{X} f(x)d(x) \\
&= \int_{A}^{X} 1/(B-A)d(x) \\
&= (X-A)/(B-A)
\end{aligned}
$$

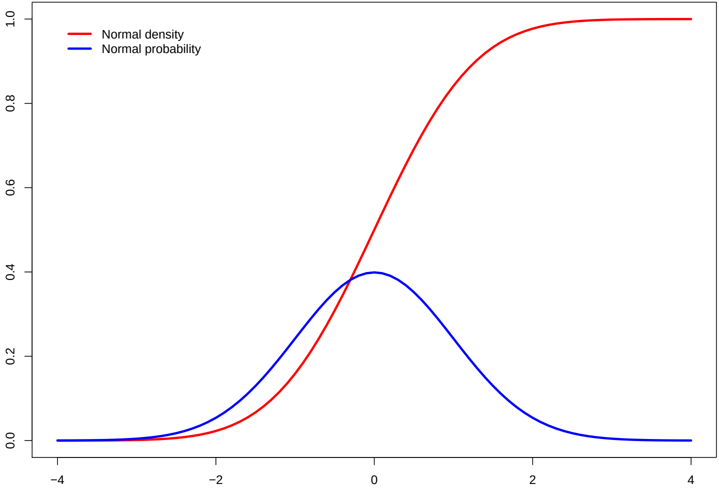
Examples of a PDF: Normal/Gaussian RV

- $f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}((x-\mu)/\sigma)^2}$
- The normal has **two** parameters, μ, σ.
- A lot of economic variables are assumed to be normal distributed.

Features of the normal PDF

- RVs can take values from $-\infty$ to ∞.
- It is symmetric: $\Pr(-x) = \Pr(x)$
- When X is a normal RV with parameters μ, σ, then $Y = 5.6 + 0.2X$ will also be a normal RV, with known parameters $(5.6 + 0.2\mu), (0.2\sigma)$.
- Note: Special case of a normal distribution is $\mu = 0, \sigma = 1$. This is called a **standard normal distribution**.

Probability distributions



Coming back to the econometrician's problem: sample vs. population distributions

Examples of a PD: Bernoulli RVs

Bernoulli distribution: The outcome is either a “failure” (0) or a “success” (1).

- Y is a bernoulli RV when

$$\begin{aligned} \Pr(Y = 0) &= p \\ \Pr(Y = 1) &= (1 - p) \end{aligned}$$

  For example, the USD-INR rises or not at the end of the day.
- The bernoulli distribution has **one** parameter, p.
- The value of p falls in $[0, 1]$.
- Mathematically:

$$B(Y) = p^Y (1 - p)^{1-Y}, \text{ where } Y = 0, 1$$

The CD of a Bernoulli RV

- We need to know what p of the RV is. Say $p = 0.35$.
- The CD of this Bernoulli RV is:

Value	Probability
$Y \leq 0$	0.35
$Y \leq 1$	1.00

The econometrician's problem in the bernoulli world

- The economic hypothesis is that the observed RV, Y, comes out of a bernoulli PD.
  Ie, model: $Y \sim B(p)$
  where B(p) stands for a bernoulli PD with probability of "success" p.
- The model defines how many "parameters" are required to be fully specified.
  Ie, in the bernoulli, number of parameters = 1, p.
- Econometrician's problem: estimate the parameters of the "model" such that it "explains the observed $\vec{Y}$".
  Ie, in the bernoulli, what is the value of p that matches the sample probability of "success"?

Using what we know so far

Describing data concisely

- One of the uses of the concepts of statistics to better describe data.
- The questions we try to answer are:
  1. What is the likely probability distribution/density?
  2. What are the parameters for the distribution/density?
- There are two kinds of tools: visual and numerical.

Graphical tools

- Eyeballing the data: is it continous or discrete?
- Most popularly used graphical tool: histograms or frequency distribution plots.
- Density plots: smoothed versions of histograms for continuous RVs.

Histograms

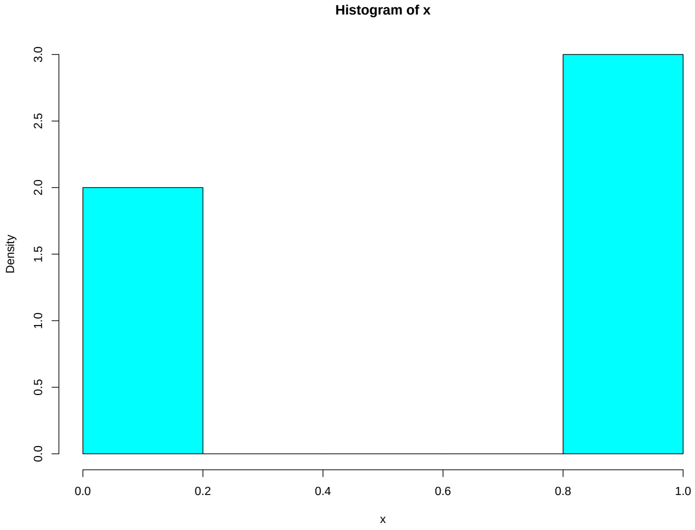
- The histogram is the plot of the unique values in a sample and the frequency with which they are observed.
- RV Value on the x-axis, frequency (with which the value occur in the data) on the y-axis.
- Histograms for the discrete case is easy: have to rework the goal a little for the continuous case.

Example 1: Discrete RV

1. The data is a set of 20 values.
2. $x = 0\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 0\ 0\ 1\ 0\ 1\ 1\ 0\ 0\ 1\ 0\ 1$
3. It looks discrete. It looks binary.
4. Frequency table:

x	Freq
0	13
1	7

Histogram of x



Guessing the PD

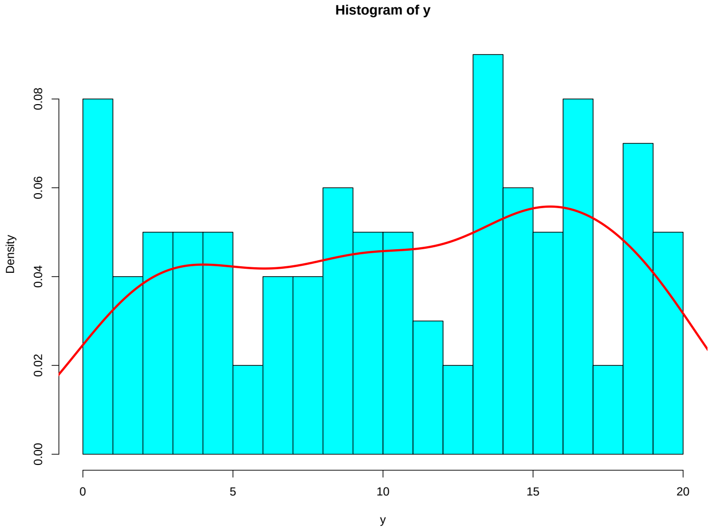
- Binary distribution – Bernoulli?
- $p = 0.45$

Example 2: Discrete RV

1. Sample = 100 values
2. (The first 13 values)
   $y = 11 \;\; 9 \;\; 6 \;\; 13 \;\; 15 \;\; 18 \;\; 17 \;\; 11 \;\; 8 \;\; 10 \;\; 2 \;\; 0 \;\; 12$

y	Freq	y	Freq	y	Freq	y	Freq
0	4	6	5	11	7	16	5
1	7	7	6	12	5	17	7
2	4	8	9	13	4	18	4
3	1	9	5	14	3	19	0
4	5	10	5	15	4	20	3
5	7						

Histogram of y



Guessing the PD

- Discrete RV from 0 to 20
- Could be a uniform discrete PD

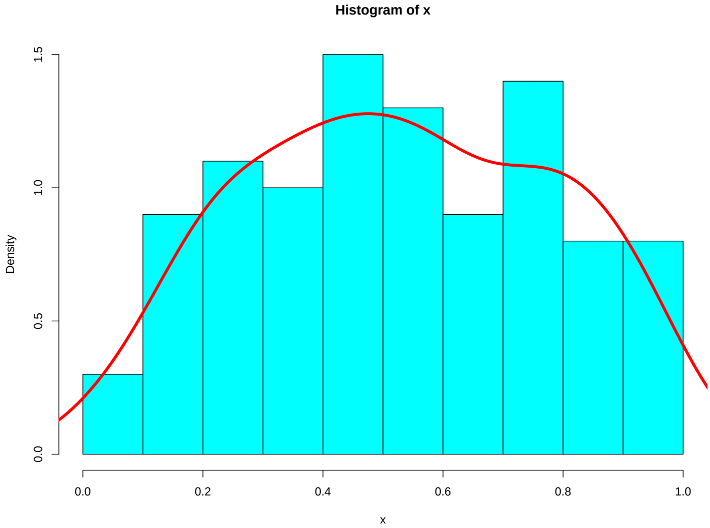
Example 4: Continuous RV

1. Sample = 100 values
2. $k =$

0.993941516	0.612212929	0.201375686
0.240819249	0.142533204	0.430064859
0.697499793	0.030674237	0.944907661
...	...	...

3. Frequency table: each element has a frequency of one.

Histogram of k



Guessing the PD

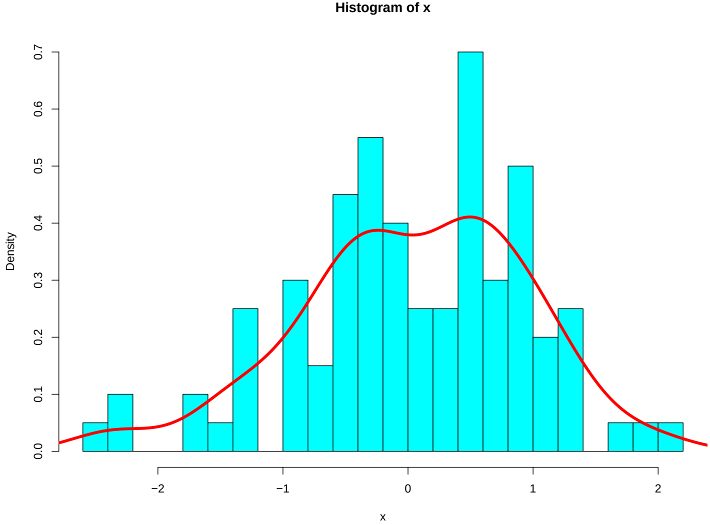
- RVs are continuous
- Could be a uniform distribution? (No negative values, data appears range bound.)

Example 5: Continuous RV

1. Sample = 100 values
2. $n =$

0.724773431	0.500281917	0.903696952
0.612282267	-0.185570961	0.247409823
-0.820567376	-1.413818678	1.368954272
...	...	...

Histogram of n



Guessing the PD

- Continuous RV
- Could be normally distributed?

Sample vs. population distributions

Population vs. ample distributions of a “fair coin” toss

- Event: discrete, bernoulli distributed
- Model parameters: p.
- Population parameter value: 50%
- Parameter value for a sample of 100 tosses, $p_{n=100}$? For a sample of 20 tosses, $p_{n=20}$?
- Econometrician’s problem: we only get samples, never populations.
  How do we work backward from sample values to population distribution parameter values?

Economic problem: gender representation in demographic data

- Economic RV, G: fraction of girls in the human race.
- G: discrete, bernoulli.
- Parameters: p_G
- Hypothesis: $p_G = 0.5$

Econometric problem: is $p_G = 0.5$ a good hypothesis?

- Sample data set: number of children born in the UK, 2004.
- $N = 715996$, $N_{\text{boys}} = 367586$, $N_{\text{girls}} = 348410$

	0	1
N_G	367586	348410
$\hat{p}_G$	0.513	0.487

- Here, the events are mutually exclusive. Thus, p_G is the fraction of girls in the population.
- $\hat{p}_G$ is the "estimated" or "sample" or "observed" parameter.
- Difference with the coin toss problem: population distribution can be simulated. Here, the population distribution cannot be simulated – our only observation is the sample.

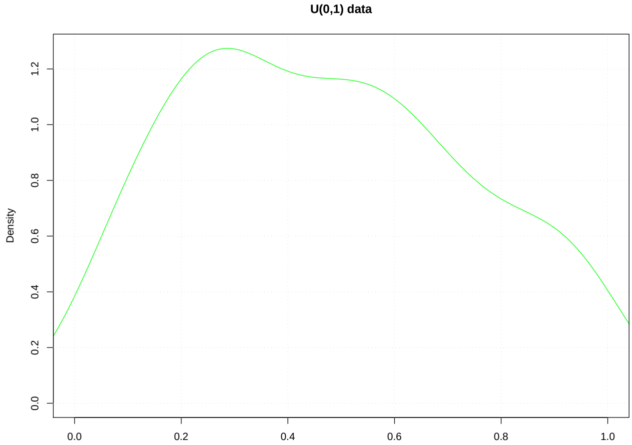
Simulating samples from PDFs

Sampling from a uniform distribution

- The PD/PDF is a description of all likely values that exist in the **population.**
- In reality, all the data that we can observe is a **sample**, which is a subset of the population.

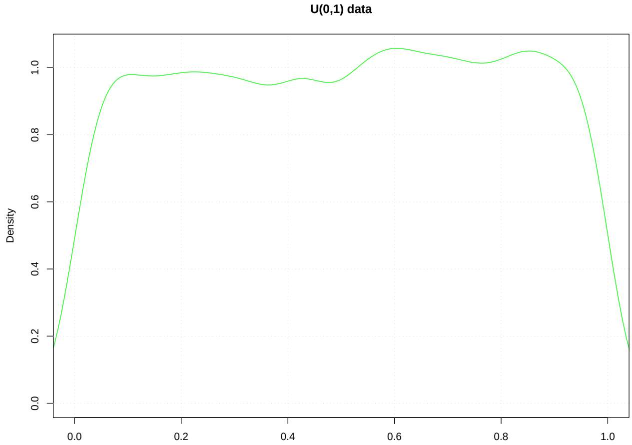
Samples of UniformPDF

Density of simulation from a continuous uniform PDF, with min = 0, max = 1, size = 100.

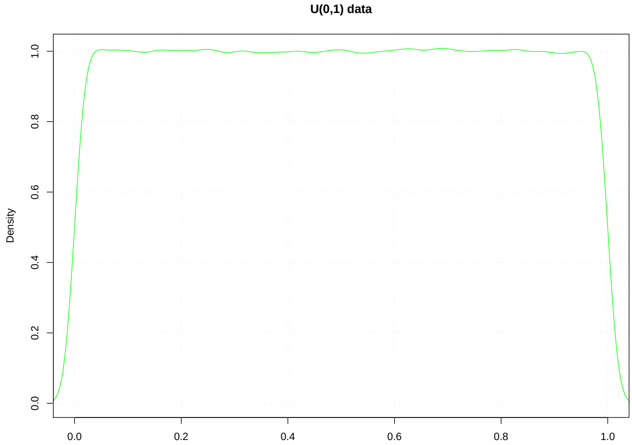


Samples of UniformPDF, 2

Density of simulation from a continuous uniform PDF, with min = 0, max = 1, size = 10000.



Density of simulation from a continuous uniform PDF, with min = 0, max = 1, size = 1000000.

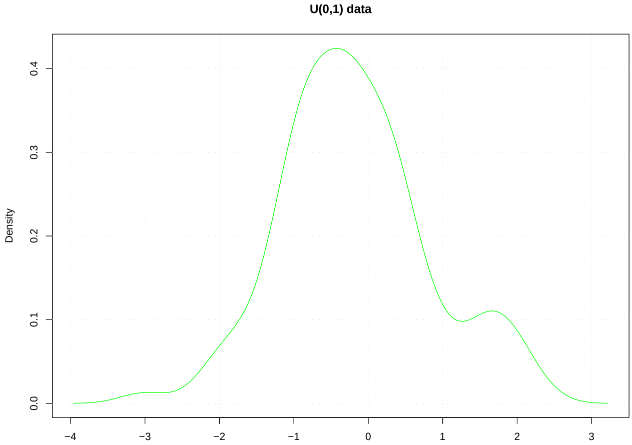


Samples are imperfect subsets of populations!

- We went from a sample size of 100 to 10000 to a million, and still the sample density shows a visible deviation from the “population” density.
- Most modelling exercises deals with creating summary statistics of the sample data.

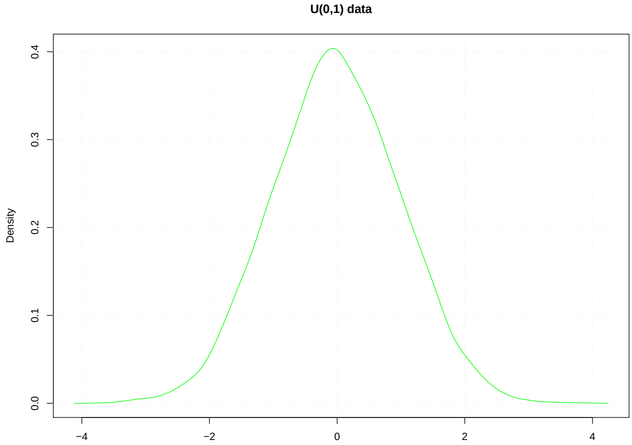
Samples of Normal PDF, 1

Density of simulation from a normal PDF, with $\mu = 0, \sigma = 1$, size = 100.



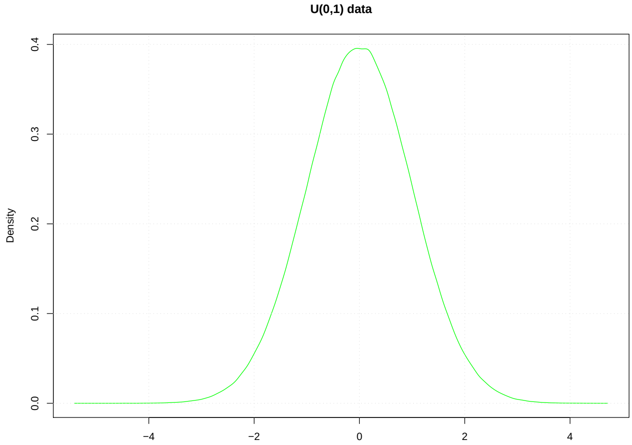
Samples of Normal PDF, 2

Density of simulation from a normal PDF, with $\mu = 0, \sigma = 1$, size = 10000.



Samples of Normal PDF, 3

Density of simulation from a normal PDF, with $\mu = 0, \sigma = 1$, size = 1000000.



References

- Chapter 1, SHELDON ROSS., “Introduction to Probability Models.”, Harcourt India Pvt. Ltd., 2001, 7th edition.