Prediction vs Estimation

Satyanath Bhat

Institute of Operations Research and Analytics
satya.bhat@nus.edu.sg

November 1, 2019

Agenda

- Statistics primer through R
  - Random variables
  - Continuous and discrete random variables
  - Normal distribution
  - Central limit theorem
  - Applications
- Prediction in high dimension
  - Losses in prediction
  - Lasso
- Estimation using machine learning ideas

Random variable

Random variable takes a value which depends on the outcome of a random phenomena.

- Weight in kilogram of a randomly picked MSBA student

Random variable

Random variable takes a value which depends on the outcome of a random phenomena.

- Weight in kilogram of a randomly picked MSBA student
- Number seen in dice throw

Random variable

Random variable takes a value which depends on the outcome of a random phenomena.

- Weight in kilogram of a randomly picked MSBA student
- Number seen in dice throw
- Outcome space of a coin toss $\Omega = \{Head, Tail\}$

Random variable

Random variable takes a value which depends on the outcome of a random phenomena.

- Weight in kilogram of a randomly picked MSBA student
- Number seen in dice throw
- Outcome space of a coin toss $\Omega = \{Head, Tail\}$
  - Let us define a random variable $X : \Omega \to \mathbb{R}$

Random variable

Random variable takes a value which depends on the outcome of a random phenomena.

- Weight in kilogram of a randomly picked MSBA student
- Number seen in dice throw
- Outcome space of a coin toss $\Omega = \{Head, Tail\}$
  - Let us define a random variable $X : \Omega \to \mathbb{R}$
  -
$$X(\omega) = \begin{cases} 0, & \text{if } \omega = T \\ 1, & \text{otherwise} \end{cases} \tag{1}$$

Random variable

Random variable takes a value which depends on the outcome of a random phenomena.

- Weight in kilogram of a randomly picked MSBA student
- Number seen in dice throw
- Outcome space of a coin toss $\Omega = \{Head, Tail\}$
  - Let us define a random variable $X : \Omega \to \mathbb{R}$
  - $$X(\omega) = \begin{cases} 0, & \text{if } \omega = T \\ 1, & \text{otherwise} \end{cases} \tag{1}$$
  - Head occurs with probability $p = 0.5$ or $p(H) = p(T) = 0.5$

Random variable

Random variable takes a value which depends on the outcome of a random phenomena.

- Weight in kilogram of a randomly picked MSBA student
- Number seen in dice throw
- Outcome space of a coin toss $\Omega = \{Head, Tail\}$
  - Let us define a random variable $X : \Omega \to \mathbb{R}$
  - $$X(\omega) = \begin{cases} 0, & \text{if } \omega = T \\ 1, & \text{otherwise} \end{cases} \tag{1}$$
  - Head occurs with probability $p = 0.5$ or $p(H) = p(T) = 0.5$
  - What is the outcome space of weight experiment in MSBA class? Roll no. of all students
  - What are the probabilities? 1/(no. of people in MSBA class)

Random variable

Random variable takes a value which depends on the outcome of a random phenomena.

- Weight in kilogram of a randomly picked MSBA student
- Number seen in dice throw
- Outcome space of a coin toss $\Omega = \{Head, Tail\}$
  - Let us define a random variable $X : \Omega \to \mathbb{R}$
  - $$X(\omega) = \begin{cases} 0, & \text{if } \omega = T \\ 1, & \text{otherwise} \end{cases} \tag{1}$$
  - Head occurs with probability $p = 0.5$ or $p(H) = p(T) = 0.5$
  - What is the outcome space of weight experiment in MSBA class?
  - What are the probabilities?
- Run estimator of p demo1.Rmd

rbinom (t, c, p)

trials

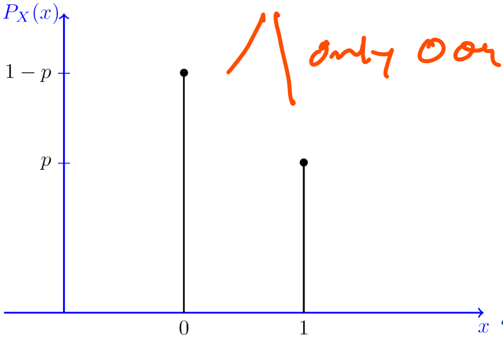
coins tossed per trial

Prob of getting 1

Discrete random variables

Discrete random variables can be represented by probability mass functions (pmf)

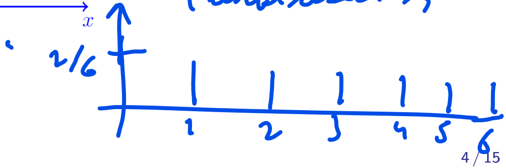
$X \sim Bernoulli(p)$



only 0 or 1 in Bernoulli

- What is the pmf of fair dice?

For a dice (unbiased),



Uniform random variables

Consider the following experiment

- Pick a flexible ruler
- Deform the ruler to a circle
- Ask a friend to very precisely choose a point in the circle
- Open the ruler and record the point chosen

The distribution of the point chosen is said to be uniform.

- What is the probability that the point is 0.5? 0

↳ How?

Remember that there is no area for a single point

→ For a finite no. of points, area is always 0

Uniform random variables

Consider the following experiment

- Pick a flexible ruler
- Deform the ruler to a circle
- Ask a friend to very precisely choose a point in the circle
- Open the ruler and record the point chosen

The distribution of the point chosen is said to be uniform.

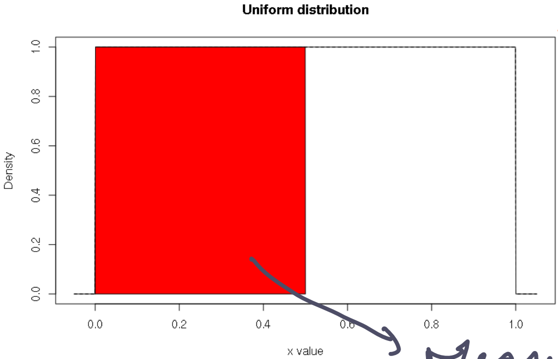
- What is the probability that the point is 0.5?
- What is the probability that point chosen lies in the first half of the ruler?

Probability density function

runif(100, 0, 1)

100 trials

uniform dist between 0,1



Run demo2.Rmd (uniform)

For a uniform distribution

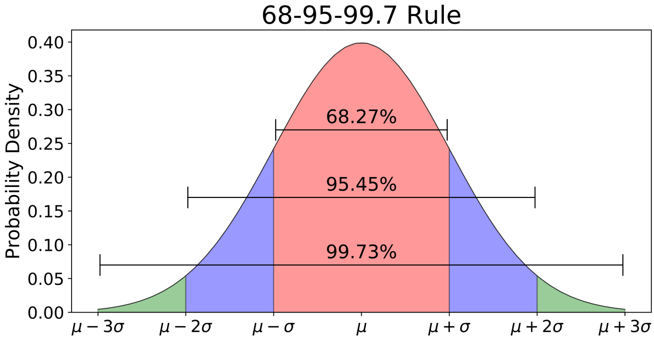
Measure the area to find prob. of a point lying between (x_1, x_2)

Normal Distribution

Defined by three parameters

- μ Mean
- σ Standard deviation
- σ^2 Variance

How does it look?



- This figure is not accurate, why?
- Run demo3.Rmd

Galton board

Play this youtube video ▸ Link

Central limit theorem informal statement

The mean estimator of a random variable with mean μ and variance σ^2 "looks like" normal with mean μ and variance $\frac{\sigma^2}{n}$ In applications, $n > 30$ is good enough.

Application 1: Exit polls and surveys

Suppose there are 10 million citizens in a country. It is election time and all the citizens have cast their votes. There are two candidates in fray Mr A and Mr B. The counting is tedious and it takes 5 days to announce the result. You are the head of statistics division of a prominent newspaper and you need to design an exit poll to estimate the winner.
Run app1.Rmd

Application 2: Sanity check for RCT

Suppose you have been hired by a marketing firm as a data scientist. Your first assignment is to look at marketing data from a client company FairGroceries. Your boss claims that he randomly chose 0.1 of their clients as targets for an advertising campaign. The experiment intends to show that advertising boosts sales. However, the sales post-treatment has been negative!; you suspect that your nincompoop boss did not truely conduct the RCT correctly. You look at the pre-treatment sales data to gather evidence.
Run app2.Rmd

Lasso for prediction

coefficients of predictors are brought down to 0

$$\min\left(\sum_{i=1}^{n}(y_i - \sum_j x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p}|\beta_j|\right)$$

$\lambda > 0$; L1 norm

λ is chosen using cross-validation

Run lassoPred.Rmd

Why lasso works?

For simple regression,

$$\beta_{OLS} = \underbrace{\operatorname*{argmin}_{\beta} \sum (y_i - \sum \beta_i x_{ij})^2}_{\text{minimise}}$$

$\beta = (2, -3, 4)$

L1 norm for β: $||\beta||_1 = \sum_{i=1}^{3} |\beta_i| = (|2| + |-3| + |4|) = 9$

Lasso for estimation?

If some variables in x and some in W are correlated.

- Suppose $y = \alpha + \beta X + \gamma W + \epsilon$
- Suppose treatment has a self-selection based on X
- Suppose there are high dimensional noise
- Using lasso here will not work
- We have to use two stage lasso
- Use Belloni, Chernozukov, and Hansen (2013)

Steps:
1. Lasso of 'y' - identify relevant variables
2. " " 'W' - identify relevant variables

→ Chuck all other variables and run on y

Thank you