CMPT 155: Computer Applications for Life Sciences

Lecture 11: Continuous Probability - The Normal Distribtuion

Ivan E. Perez

April 19, 2022

Presentation Outline

Homework & Administrative Schedule

- Homeworks:
  - #6 Due: Friday April 22th at 6pm
  - #7 Due: Friday, April 29 rd at 6pm
  - #8 Due: Friday, May 6th at 6pm
- Final Exam Review: Tuesday, May 3rd at 6pm
- Mock Final Exam: Wednesday, May 4th
- **Final Exams**:
  - Section 01 (8am) Final Exam: May 9th 11am - 1pm
  - Section 02 (9am) Final Exam: May 10th 11 am - 1pm

Differences Between Discrete and Continuous Random Variables

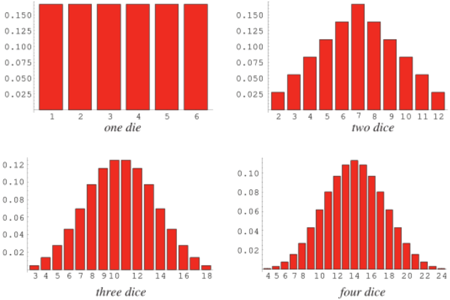
- Discrete probability is computed by *counting* the subset of outcomes that satisfy restrictions against the broader set of outcomes.
- Coninuous probability is about *measuring* the range of outcomes that satisfy restrictions against the broader range of outcomes.

Example:

- In the Binomial Model for Probability we were interested in the **direcete** number of k-success in n-trials.
- In the Poisson Model for Probability we were interested in the **discrete** number of x-arrivals in a period of length t.
- In the Guassian(Normal) Model for Probability we are interested in **ranges** of values.

Why Should I care about the Normal Distribution?

While few random variables are normally distributed on their own.
The Central Limit Theorem states that the sum of non-normally distributed random variables tends to turn into a normal distribution as more random variables are added to the sum.
Example: Adding Die to a dice roll



Examples/Descriptions of the Normal Distribution

Common Examples of Normal Distributions include:

- Distribution of Heights at different ages
- Distribution of food weights at a deli.
- Distribution of Test scores.

The Normal Distribution has the following characterstics:

- Symmetric about the Mean.
- The Mode, Mean, and Median are all equivalent values.
- 50 % of values lie below the Mean, and 50 % of values lie above the mean.
- long tails that stretch infinitely.

Features/Parameters of the Normal Distribution

Normally distributed random variables can be described using the parameters for location, μ, and dispersion σ^2.

- Mean - μ
  - Measure of central tendency/location
  - Can be estimated using AVERAGE(),
- Variance - σ^2
  - Measure of dispersion/scale
  - Can be estimated using VAR() or VAR.S()
  - its cousin Standard Deviation, σ
    - Can be estimated using STDEV()
    - is just $\sqrt{\sigma^2}$

Equations of the Normal Distribution

Probability density functions(pdf's) can be used to describe distributions. For a given μ, σ, the normal distribution's probability density function is

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \tag{1}$$

Where x is the input variable whose probability density is being computed using $f(x, \mu, \sigma)$.
The probability of a range of values is the area under the curve(AUC) between two computed probability densities.
The probability of a random variable, $\mathbf{X}$, being within the range of x_1, and x_2, (i.e., $[x_1, x_2]$), for a given μ, σ is:

$$P(x_1 \leq \mathbf{X} \leq x_2) = \int_{x_1}^{x_2} f(\mathbf{X}, \mu, \sigma)dx \tag{2}$$

Equations of the Normal Distribution

Unfortunately, NORMDIST() only computes probabilities from $-\infty$ to x in the form:

$$P(\mathbf{X} \leq x) = \int_{-\infty}^{x} f(x, \mu, \sigma)dx. \tag{3}$$

We can express (Eq. 2) as a difference of probabilities,

$$P(x_1 \leq \mathbf{X} \leq x_2) = P(\mathbf{X} \leq x_2) - P(\mathbf{X} \leq x_1) \tag{4}$$

To express probabilities where the random variable, $\mathbf{X}$, is at least x_1; we take the complement of $P(\mathbf{X} \leq x_1)$:

$$P(\mathbf{X} \geq x_1) = (P(\mathbf{X} \leq x_1))^c = 1 - P(\mathbf{X} \leq x_1) \tag{5}$$

The NORMDIST() function

NORMDIST() computes the output of the pdf, $f(x, \mu, \sigma)$, if `cumulative` is set to `FALSE`, or the probability $P(\mathbf{X} \leq x)$ if `cumulative` is set to `TRUE`.

The arguments are:

- `x` : numeric
  - The input, x, for computing either $f(x, \mu, \sigma)$ or $P(\mathbf{X} \leq x)$.
- `mean` : numeric
  - μ the population mean of the normal distribution.
- `standard_dev` : numeric
  - σ the population standard deviation.
  - Must be 0 or more.
- `cumulative` : `TRUE` or `FALSE`
  - `TRUE` : Computes $P(\mathbf{X} \leq x)$.
  - `FALSE`: Computes $f(x, \mu, \sigma)$.

Example 1a: Plotting the Normal Distribution

Use NORMDIST() to plot heights of men aged 60-69 years, given that their height is on average 5'9" with a standard deviation of 3 inches. Heights are assumed to be normally distributed about the mean.

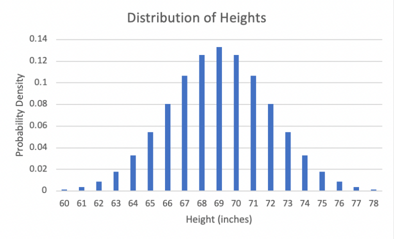
1. Label columns A, B, and C as 'Height', 'Density', and 'Cumulative' respectively.
2. In cells A2:A20 List the heights increasing in 1in increments from the mean minus 3 standard deviations, (i.e., 60 in) to the mean plus 3 standard deviations (i.e., 78 in).
3. In cell B2 compute the probability *density* that of male at this height.
   - In cell B2 type `=NORMDIST(A2, 69, 3, FALSE)`
4. Use Autofill to compute *densities* in cells B3:B20.

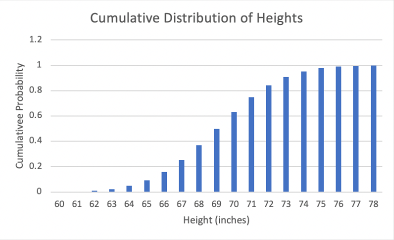
Example 1a: Solution

1. In cell C2 compute the probability that a male will be this height or shorter.
    - In cell C2 type `=NORMDIST(A2, 69, 3, TRUE)`
2. Use Autofill to compute the probabilities in cells C3:C20.
3. Plot the probability density function and cumulative probabilities using separate 2D-Column charts.
    - Make sure that the column labels reference A2:A20.
    - Remember to only plot these values as a single series.

Example 1a: Solution

Height	Density	Cumulative
60	0.0014773	0.0013499
61	0.0037987	0.0038304
62	0.0087406	0.0098153
63	0.017997	0.0227501
64	0.033159	0.0477904
65	0.05467	0.0912112
66	0.0806569	0.1586553
67	0.1064827	0.2524925
68	0.1257944	0.3694413
69	0.1329808	0.5
70	0.1257944	0.6305587
71	0.1064827	0.7475075
72	0.0806569	0.8413447
73	0.05467	0.9087888
74	0.033159	0.9522096
75	0.017997	0.9772499
76	0.0087406	0.9901847
77	0.0037987	0.9961696
78	0.0014773	0.9986501





Example 1b: Heights

According to US Census Data in 2007/2008 Men aged 60-69 years, are on average 5'9" with a standard deviation of 3 inches. Heights are assumed to be normally distributed about the mean.

What is the proportion of men who are:

1. Less than and including to 65 inches tall?
2. Less than and including to 69 inches tall?
3. Greater than 71 inches tall?
4. Greater than 68 inches tall?
5. Between 67 and 70 inches (inclusive)tall?

Example 1b: Solution

For Probabilities that include:

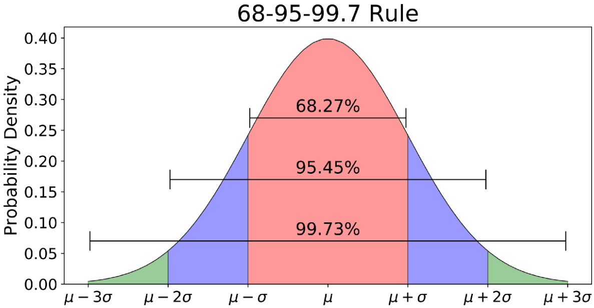
- *fewer* then compute $P(\mathbf{X} \leq x)$.
  - Example 1b Q1:
    $P(\mathbf{X} \leq 65)$ type: `=NORMDIST(65, 69, 3, TRUE)`.
- *greater* then compute $P(\mathbf{X} \geq x)$.
  - Use (Eq. 5) to express $P(\mathbf{X} \geq x)$ as a complement of $P(\mathbf{X} \leq x)$.
  - Example 1b Q3:
    $P(\mathbf{X} \geq 71) = (P(\mathbf{X} \leq x))^c = 1 - P(\mathbf{X} \leq x)$ type:
    `=1-NORMDIST(71, 69, 3, TRUE)`.
- *lower*, x_1 and *upper*, x_2 bounds then compute $P(x_1 \leq \mathbf{X} \leq x_2)$.
  - Use (Eq. 4) to express the probability as a difference.
  - Example 1b Q5:
    $P(x_1 \leq \mathbf{X} \leq x_2) = P(\mathbf{X} \leq x_2) - P(\mathbf{X} \leq x_1)$ type:
    `=NORMDIST(70, 69, 3, TRUE) - NORMDIST(67, 69, 3, TRUE)`

Example 1b: Solution

Range	Expression	Computation	Result
$(-\infty, \mathbf{65}]$	$P(X \leq \mathbf{65})$	`=NORMDIST(65, 69, 3, TRUE)`	0.0912
$(-\infty, \mathbf{69}]$	$P(X \leq \mathbf{65})$	`=NORMDIST(69, 69, 3, TRUE)`	0.5000
$[\mathbf{71}, \infty)$	$P(X \geq \mathbf{71})$	`=1-NORMDIST(71, 69, 3, TRUE)`	0.2523
$[\mathbf{68}, \infty)$	$P(X \geq \mathbf{68})$	`=1-NORMDIST(68, 69, 3, TRUE)`	0.6306
$[\mathbf{67}, \mathbf{70}]$	$P(\mathbf{67} \leq X \leq \mathbf{70})$	`=NORMDIST(70, 69,3,TRUE) - NORMDIST(67, 69,3, TRUE)`	0.3781

Facts to Remember about the Normal Distribution

1. $\sim 68\%$ of the area under the curve is enclosed within ± 1 standard deviation, 1σ, from the mean, μ.
2. $\sim 95\%$ of the area under the curve is enclosed within ± 2 standard deviations, 2σ, from the mean, μ.
3. $\sim 99\%$ of the area under the curve is enclosed within ± 3 standard deviations 3σ,from the mean, μ.



The Standard Normal Distribution

The Standard Normal Distribution is a special case of the Normal Distribution. where

- The mean, $\mu = 0$
- The standard deviation, $\sigma = 1$.

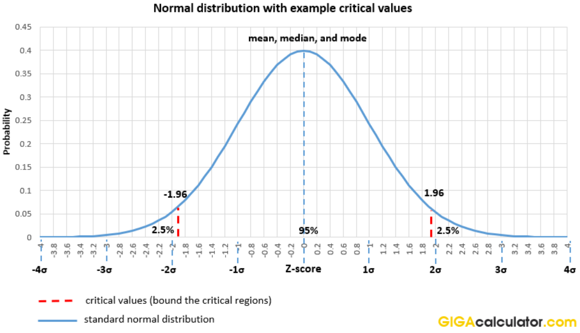


Figure: The Standard Normal Distribution

Further Reading

The topics covered in the lecture can be found in *Compter Applications for Life Sciences* p. 76-84.