Chapter 3: Continuous Random Variables

Mariona Segú

Assistant Professor
Thema, CY Cergy Paris Université

Fall 2023

Contents

Contents

1. Introduction

- Many types of data can take any value in some interval (sometimes all) of the real numbers.
- Here, the probability density function for discrete random variables is not enough because
  1. the number of possible outcomes is uncountable, so we can't just add up all probabilities
  2. the probability of any particular value on the continuum typically has to be zero.
- We have to deal with this type of random variables separately from the discrete case.

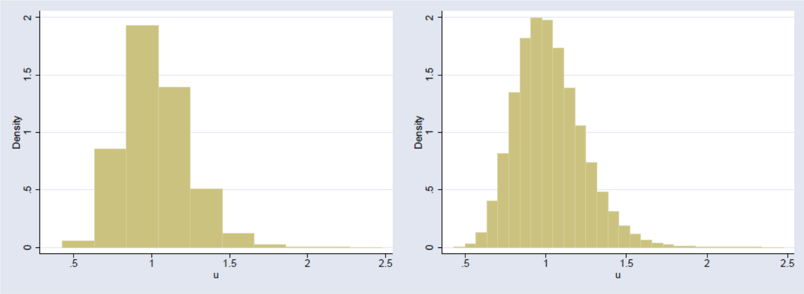
1. Introduction - Definition

Definition A random variable Y has a **continuous distribution** if Y can take on any values in some interval, bounded or unbounded, of the real line.

- We can "discretize" the distribution by putting the possible values the random variable can take into "bins"
- i.e. instead of looking at the probabilities $P(Y = y)$, we'll look at probabilities for intervals, i.e. $P(y_1 \leq Y \leq y_2)$.
- Then, we can plot the bins into a histogram

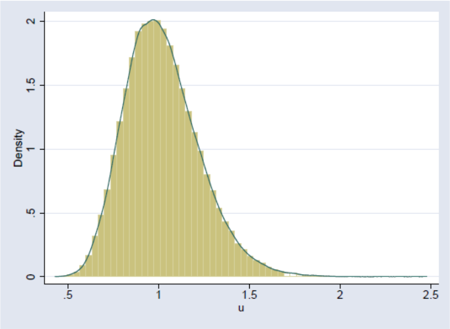
1. Introduction - Definition

Histograms of the same Distribution for 10 and 30 Bins, respectively



1. Introduction - Definition

Histogram with 60 Bins and Continuous Density



1. Introduction - Definition

We can express the probability over a wider interval as the sum of smaller intervals:

$$P(y_j \leq Y \leq y_k) = \sum_{i=j+1}^{k} P(y_{i-1} \leq Y \leq y_i)$$

- As we make the intervals smaller, the histogram approaches a smooth curve.
- In the limit, we find the area under a curve —the **integral** of a function.

$$F(y) = P(a \leq y \leq b) = \int_a^b f(y)dx$$

Contents

2. The CDF of a Continuous Variable

Definition. A random variable Y with cumulative distribution function (CDF) $F(y)$ is said to be **continuous** if $F(y)$ is a continuous function for all $-\infty < y < \infty$. **Interpretation.**

- Just like in the discrete case, the CDF gives the probability that Y takes a value less than or equal to y:

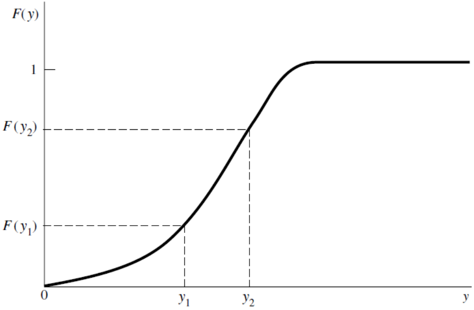
$$F(y) = P(Y \leq y).$$

- It always increases from 0 to 1 as y moves from $-\infty$ to $+\infty$.
- The CDF is useful because it fully describes the distribution of Y.

2. The Shape of the CDF

Discrete vs. Continuous CDFs

- In the **discrete case**, $F(y)$ is a **step function**: it jumps at each possible value of Y.
- In the **continuous case**, $F(y)$ is **smooth and continuous** because Y can take infinitely many values.



2. PDF and CDF of a Continuous Variable

A Continuous RV will have $P(Y = y) = 0$
What does it mean to have $P(Y = y) = 0$?

- If this were not true and $P(Y = y_0) = p_0 > 0$, then $F(y)$ would have a discontinuity (jump), violating the continuity assumption.

Rainfall

Consider the example of measuring daily rainfall. What is the probability that we will see a daily rainfall measurement of exactly 2.193 cm? It is quite likely that we would never observe that exact value even if we took rainfall measurements for a lifetime, although we might see many days with measurements between 2 and 3 cm.

2. PF for the Continuous case?

- In a **continuous distribution**, Y can take infinitely many values in any interval.
- The probability of observing one exact value (e.g. $Y = 2.193$) is therefore **infinitesimally small** —effectively **zero**.
- **Unlike the discrete case**, where we define a **probability function (PF)** giving $P(Y = y)$ for each value, the PF has no meaning here since $P(Y = y) = 0$ for all y.

Instead, we use the *probability density function (PDF)*.

- The PDF $f(y)$ is **not a probability**, but a **density of probability**.
- It indicates how probability is *concentrated* around different values of y.

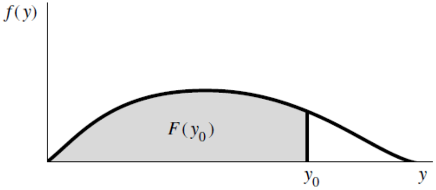
2. PDF and CDF of a Continuous Variable

The probability density function PDF is the derivative of $F(y)$:

$$f(y) = \frac{dF(y)}{dy} = F'(y)$$

It then follows that

$$F(y) = \int_{-\infty}^{y} f(t)dt$$



2. PDF and CDF of a Continuous Variable

Properties:

The PDF must satisfy that:

1. Positive probability

$$f(y) \geq 0 \qquad \forall y \in \mathbb{R}$$

2. Add up to 1

$$\int_{-\infty}^{\infty} f(y)dy = 1$$

Note that for any $Y \in \mathbb{R}$, $P(Y = y) = 0$

2. PDF and CDF of a Continuous Variable

Properties of a CDF:
The CDF must satisfy that:

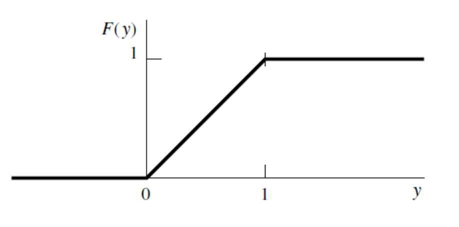
1. $F(-\infty) \equiv \lim_{y \to -\infty} F(y) = 0.$
2. $F(\infty) \equiv \lim_{y \to \infty} F(y) = 1.$
3. $F(y)$ is a nondecreasing function of y. [If y_1 and y_2 are any values such that $y_1 < y_2$, then $F(y_1) \leq F(y_2)$.]

2. PDF and CDF of a Continuous Variable

Numerical example 1

Find the PDF of

$$F(y) = \begin{cases} 0 & \text{if } y < 0 \\ y & \text{if } 0 \leq y \leq 1 \\ 1 & \text{if } y > 1 \end{cases}$$

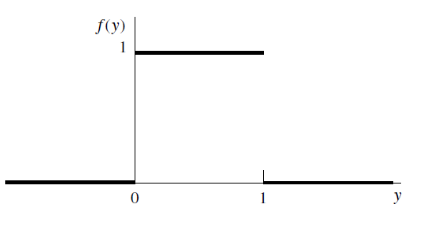


2. PDF and CDF of a Continuous Variable

Numerical example 1

We need to derivate $F(y)$

$$f(y) = \begin{cases} 0 & \text{if } y < 0 \\ 1 & \text{if } 0 \leq y \leq 1 \\ 0 & \text{if } y > 1 \end{cases}$$

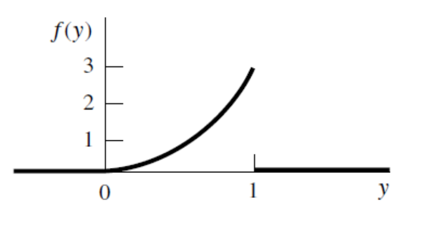


2. PDF and CDF of a Continuous Variable

Numerical example 2

Find F(y)

$$f(y) = \begin{cases} 3y^2 & \text{if } 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$



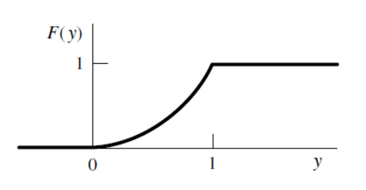
2. PDF and CDF of a Continuous Variable

Numerical example 2

We need to integrate $f(y)$ ▸ How to integrate

$$F(y) = \int_0^y 3t^2 dt = t^3]_0^y = y^3 - 0^3$$

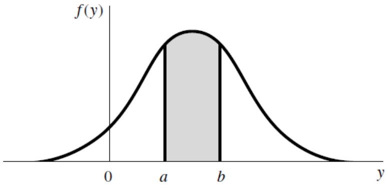
$$F(y) = \begin{cases} 0 & \text{if } y < 0 \\ y^3 & \text{if } 0 \leq y \leq 1 \\ 1 & \text{if } y > 1 \end{cases}$$



2. PDF and CDF of a Continuous Variable

Here is how we can work with Continuous RV:
If we want to know the proba that Y falls in a given interval $[a, b]$, we can compute

$$P(Y \in [a, b]) = P(a \leq Y \leq b) = \int_a^b f(y)dy$$



Here the equality sign does not matter as much as in the discrete case.

2. Why the Equality Sign Does Not Matter

For continuous random variables:

- The probability of taking any **exact value** is zero:

$$P(Y = a) = 0 \quad \text{and} \quad P(Y = b) = 0$$

- Therefore, including or excluding the endpoints in an interval does not change the probability:

$$P(a \leq Y \leq b) = P(a < Y \leq b) = P(a \leq Y < b) = P(a < Y < b)$$

- This is because probability is represented by the **area under the PDF curve**, and a single point has no area.

2. PDF and CDF of a Continuous Variable

Find c

Given

$$f(y) = \begin{cases} cy^2, & \text{if } 0 \leq y \leq 2 \\ 0, & \text{elsewhere} \end{cases}$$

Find the value of c for which $f(y)$ is a valid density function.

2. PDF and CDF of a Continuous Variable

Find c

We require a value for c such that

$$F(\infty) = \int_{-\infty}^{\infty} f(y)\, dy = 1$$

Given the function $f(y)$, this can be written as:

$$\int_0^2 cy^2\, dy = \frac{cy^3}{3}\Big|_0^2 = \frac{8c}{3}.$$

Thus, $\frac{8}{3}c = 1$, and we find that $c = \frac{3}{8}$.

2. PDF and CDF of a Continuous Variable

Find c

Find $P(1 \leq Y \leq 2)$ for the previous example. Also find $P(1 < Y < 2)$.

2. PDF and CDF of a Continuous Variable

Find c

Find $P(1 \leq Y \leq 2)$ for the previous example. Also find $P(1 < Y < 2)$.

We have:

$$P(1 \leq Y \leq 2) = \int_1^2 f(y)\,dy = \frac{3}{8}\int_1^2 y^2\,dy = \frac{3}{8}\left[\frac{y^3}{3}\right]_1^2 = \frac{7}{8}.$$

Because Y has a continuous distribution, it follows that:

$$P(Y = 1) = P(Y = 2) = 0$$

and, therefore, that:

$$P(1 < Y < 2) = P(1 \leq Y \leq 2) = \frac{7}{8}.$$

Wooclap

Question #22, #23 and #24

Contents

3. The Expected Value of a Continuous RV

Sometimes, it is difficult to find the PDF of a continuous random variable. We can then use its *moments*.

Definition The expected value of a continuous random variable Y is

$$E(Y) = \int_a^b y f(y)\, dy$$

- The integral is taken over all possible values of Y on its support (a, b).
- $f(y)dy$ corresponds to $p(y)$ for the discrete case
- integration corresponds to summation
- Hence, $E(Y)$ is also a *mean*

3. The Expected Value of a Continuous RV

As in the discrete case...

- We can compute the expected value of a function $g(Y)$

$$E(g(Y)) = \int_a^b g(y)f(y)dy \quad (1)$$

- $E(c) = c$
- $E[cg(Y)] = cE[g(Y)]$
- $E[g_1(Y) + g_2(Y) + ... + g_k(Y)] = E[g_1(Y)] + E[g_2(Y)] + ... + E[g_k(Y)]$

3. The Expected Value of a Continuous RV

Example

If, Y has density function

$$f(y) = \begin{cases} \frac{1}{2}(2-y), & 0 \leq y \leq 2, \\ 0, & \text{elsewhere,} \end{cases}$$

find the mean and variance of Y.

3. The Expected Value of a Continuous RV

Example

Mean of Y:

$$E(Y) = \int_a^b y f(y) dy$$

For the given range:

$$\int_0^2 y\left(\frac{1}{2}(2-y)\right)dy = \frac{1}{2}\int_0^2 (2y - y^2)dy = \frac{1}{2}\left[y^2 - \frac{1}{3}y^3\right]_0^2$$

$$= \frac{1}{2}\left[4 - \frac{8}{3}\right] = \frac{1}{2}\left[\frac{4}{3}\right] = \frac{2}{3}$$

Variance of a Continuous Random Variable

Definition: The variance of a continuous random variable Y with PDF $f(y)$ is defined as

$$V(Y) = E[(Y-\mu)^2] = \int_a^b (y-\mu)^2 f(y)\, dy$$

- This is the **fundamental definition** of variance.
- However, using it directly is often **more complicated**, because the expression inside the integral becomes heavy to expand.
- For this reason, we usually prefer the simplified equivalent formula:

$$V(Y) = E[Y^2] - (E[Y])^2$$

which is easier to compute in practice.

3. The Expected Value of a Continuous RV

Example

The variance is given by (slide 30, CH2):

$$\sigma^2 = E[Y^2] - (E[Y])^2$$

First, compute $E(Y^2)$:

$$E[Y^2] = \int_0^2 y^2 \left(\frac{1}{2}(2-y) \right) dy = \frac{1}{2} \int_0^2 (2y^2 - y^3) dy$$

$$E[Y^2] = \frac{1}{2} \left[\frac{2}{3} y^3 - \frac{1}{4} y^4 \right]_0^2 = \frac{1}{2} \left[\frac{16}{3} - 4 \right] = \frac{1}{2} \left[\frac{4}{3} \right] = \frac{2}{3}$$

3. The Expected Value of a Continuous RV

Example

The variance is given by (slide 30, CH2):

$$\sigma^2 = E[Y^2] - [E(Y)]^2$$

Now, using the formulas:

$$Var(Y) = E[Y^2] - (E[Y])^2$$

$$Var(Y) = \frac{2}{3} - \left(\frac{2}{3}\right)^2$$

$$Var(Y) = \frac{2}{3} - \frac{4}{9}$$

$$Var(Y) = \frac{2}{9}$$

Recap: PDF, CDF, and Expected Value

- The **Probability Density Function (PDF)** $f(y)$ gives the probability distribution of a continuous random variable Y.

$$P(a \leq Y \leq b) = \int_a^b f(y)\, dy$$

- The **Cumulative Distribution Function (CDF)** $F(y)$ is obtained by integrating the PDF over the variable y:

$$F(y) = P(Y \leq y) = \int_a^y f(t)\, dt$$

- The **Expected Value** (mean) is obtained using the exact range of possible values of Y:

$$E(Y) = \int_a^b y\, f(y)\, dy$$

where $[a, b]$ is the support of Y (the interval where $f(y) > 0$).

Wooclap

Question #25 and #26

Contents

4. The Uniform Distribution

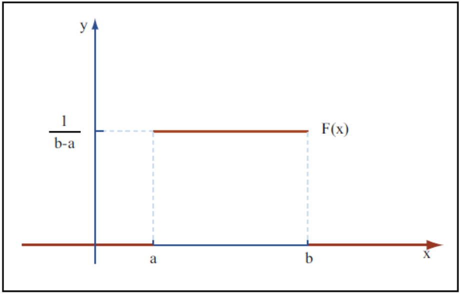
Definition A random variable Y is **uniformly distributed** on the interval $[a, b]$, with $a < b$, if it is equally likely to take any value within this interval. It has the probability density function:

$$f(y) = \begin{cases} \frac{1}{b-a}, & \text{if } a \leq y \leq b, \\ 0, & \text{otherwise.} \end{cases}$$

We write $Y \sim U(a, b)$.

4. The Uniform Distribution

p.d.f for a Uniform Random Variable, $Y \sim U(a, b)$



4. The Uniform Distribution: The CDF

Cumulative Distribution Function (CDF) For a random variable $Y \sim U(a,b)$, the CDF is obtained by integrating the PDF:

$$F(y) = P(Y \leq y) = \int_a^y f(t)\,dt$$

$$F(y) = \begin{cases} 0, & y < a, \\ \dfrac{y-a}{b-a}, & a \leq y \leq b, \\ 1, & y > b. \end{cases}$$

- $F(y)$ increases linearly from 0 to 1 between a and b.

4. The Uniform Distribution

CDF of a uniform distribution

Find the CDF of $Y \sim U[0,1]$

4. The Uniform Distribution

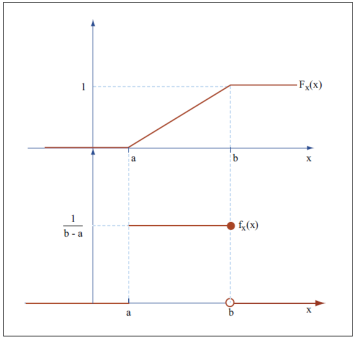
CDF of a uniform distribution

If $Y \sim U[0,1]$, then the CDF is

$$F(y) = \begin{cases} 0 & \text{if } y < 0 \\ y & \text{if } 0 \leq y \leq 1 \\ 1 & \text{if } y \geq 1 \end{cases}$$

4. The Uniform Distribution

CDF of a uniform distribution



4. The Uniform Distribution

Uniform distribution

For example, if $Y \sim U[0, 10]$, can you find $P(3 \leq Y \leq 4)$?

4. The Uniform Distribution

Uniform distribution

For example, if $Y \sim U[0, 10]$, then, its PDF is

$$f(y) = \frac{1}{b-a} = \frac{1}{10-0} = \frac{1}{10}$$

Then we can find

$$P(3 \leq Y \leq 4) = \int_3^4 \frac{1}{10} dy = \left[\frac{y}{10}\right]_3^4 = \frac{4}{10} - \frac{3}{10} = \frac{1}{10}$$

4. The Uniform Distribution

Checkout counter

It is known that, during a given 30-minute period, one customer arrived at a checkout counter. Find the probability that the customer arrived during the last 5 minutes of the 30-minute period. The actual time of arrival follows a uniform distribution over the interval of (0, 30).
First, what is the pdf?

4. The Uniform Distribution

Checkout counter

It is known that, during a given 30-minute period, one customer arrived at a checkout counter. Find the probability that the customer arrived during the last 5 minutes of the 30-minute period. The actual time of arrival follows a uniform distribution over the interval of (0, 30). If Y denotes the arrival time, then

$$P(25 \leq Y \leq 30) = \int_{25}^{30} \frac{1}{30} dy = \frac{30-25}{30} = \frac{5}{30} = \frac{1}{6}$$

4. The Uniform Distribution

Expected value of a Uniform distribution

$$\mu = E(Y) = \frac{b+a}{2}$$

Note that the mean is simply the mid-value between the two parameters.

Variance of a Uniform distribution

$$\sigma^2 = V(Y) = \frac{(a-b)^2}{12}$$

Wooclap

Question #27 and #28

Contents

The Normal distribution

Many measurements are closely approximated by a normal distribution (or bell-shaped).

Definition A random variable Y is normally distributed if the density function of Y is

$$f(y) = \frac{e^{(y-\mu)^2/2\sigma^2}}{\sigma\sqrt{2\pi}} \tag{2}$$

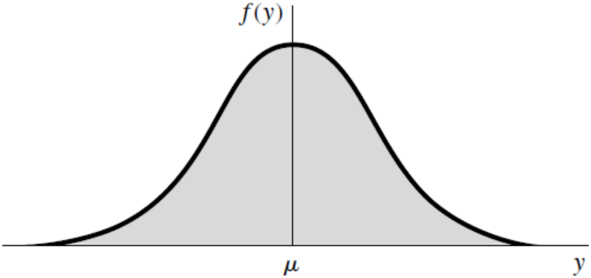
It contains 2 parameters μ and σ such that

$$E(Y) = \mu \qquad \text{and} \qquad V(Y) = \sigma^2$$

We write $Y \sim N(\mu, \sigma)$

The Normal distribution

The parameter μ is located at the center of the distribution and σ measures its spread. It is symmetric with respect to μ.

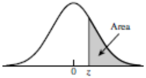


The Normal distribution

But DON'T WORRY, we will not integrate the complicated expression of f(y) to obtain F(Y). We will use an approximation presented in next slide's Table.
We use the standardized normal distribution Z, having $Z \sim N(0,1)$.
Next Table show all F(Y) values for each z point in the random variable Z.

The Normal distribution

Table 4 Normal Curve Areas
Standard normal probability in right-hand tail
(for negative values of z, areas are found by symmetry)



	Second decimal place of z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233

The Normal distribution

A Normal example

Let Z denote a normal random variable with mean 0 and standard deviation 1.

1. Find $P(Z > 2)$.
2. Find $P(-2 \leq Z \leq 2)$.
3. Find $P(0 \leq Z \leq 1.73)$.

The Normal distribution

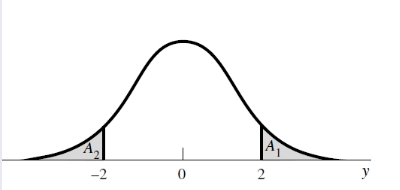
A Normal example

1. Find $P(Z > 2)$.
   Since $\mu = 0$ and $\sigma = 1$, the value 2 is actually $z = 2$. Proceed down the first (z) column in Table 4, and read the area opposite $z = 2.0$. This area, denoted by the symbol A(z), is $A(2.0) = 0.0228$. Thus, $P(Z > 2) = 0.0228$.

The Normal distribution

A Normal example

2. Find $P(-2 \leq Z \leq 2)$.



In part (1) we determined that $A_1 = A(2.0) = 0.0228$. Because the density function is symmetric about the mean, it follows that $A_2 = A_1 = 0.0228$ and hence that

$$P(-2 \leq Z \leq 2) = 1 - A1 - A2 = 1 - 2(0.0228) = 0.9544$$

The Normal distribution

A Normal example

3. Find $P(0 \leq Z \leq 1.73)$. Because $P(Z > 0) = A(0) = 0.5$, we obtain that $P(0 \leq Z \leq 1.73) = 0.5 - A(1.73)$, where A(1.73) is obtained by proceeding down the z column in Table 4, to the entry 1.7 and then across the top of the table to the column labeled .03 to read $A(1.73) = 0.0418$. Thus,

$$P(0 \leq Z \leq 1.73) = 0.5 - 0.0418 = 0.4582.$$

The Normal distribution

We can always transform a normal random variable Y to a standard normal random variable Z by using the relationship

$$Z = \frac{Y - \mu}{\sigma}$$

So we go from $Y \sim N(\mu, \sigma)$ to $Z \sim N(0, 1)$

The Normal distribution

Test scores

The achievement scores for a college entrance examination are normally distributed with mean 75 and standard deviation 10. What fraction of the scores lies between 80 and 90?

The Normal distribution

Test scores

Recall that z is the distance from the mean of a normal distribution expressed in units of standard deviation. Thus,

$$z = \frac{y - \mu}{\sigma}$$

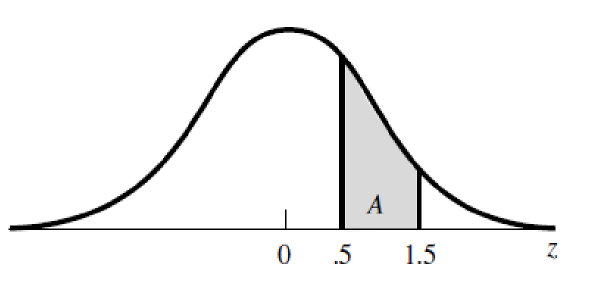
Then the desired fraction of the population is given by the area between $z_1 = \frac{80-75}{10} = 0.5$ and $z_2 = \frac{90-75}{10} = 1.5$.

$$A = A(0.5) - A(1.5) = 0.3085 - 0.0668 = 0.2417.$$

The Normal distribution

Test scores

$$A = A(0.5) - A(1.5) = 0.3085 - 0.0668 = 0.2417.$$



Wooclap

Question #29, #30 and #31

How to integrate

The integral of any polynomial is the sum of the integrals of its terms. A general term of a polynomial can be written as

$$ax^n$$

and the indefinite integral of that term is

$$\int ax^n \, dx = \frac{a}{n+1}x^{n+1} + C$$

where a and C are constants. The expression applies for both positive and negative values of n except for the special case of $n = -1$. In general, C is set equal to zero. Back

How to integrate

If definite limits are set for the integration, it is called a definite integral.

The definite integral of any polynomial is the sum of the integrals of its terms. A general term of a polynomial can be written as

$$ax^n$$

and the definite integral of that term is

$$\int_b^c ax^n\,dx = \left[\frac{a}{n+1}x^{n+1}\right]_b^c = \frac{a}{n+1}c^{n+1} - \frac{a}{n+1}b^{n+1}$$

where b and c are constants, called the limits of the integral. The procedure is basically the same as in the indefinite integral except for the evaluation at the two limits.

Back

Wooclap

Question #29 and #30