PROBABILITY & STATISTICS

Continuous Random Variables
Normal Distribution

DR/ TAMER RAGEH

The Binomial Distribution

The requirements for using the binomial distribution are as follows:

- The outcome is determined completely by chance.
- There are only two possible outcomes.
- All trials have the same probability for a particular outcome in a single trial. That is, the probability in a subsequent trial is independent of the outcome of a previous trial. Let this constant probability for a single trial be p.
- The number of trials, n, must be fixed, regardless of the outcome of each trial.

IF the random variable X has Binomial distribution, then we write P(X)= b(X;n,p):

$$P(X = x) = \binom{n}{x} p^x q^{(n-x)}$$

$$x = 0, 1, 2, \dots, n$$

n – Number of trials;
x – Number of successes;
p – Probability of a success in one trial
q – probability of failures q = 1 - P.

Consider the following scenarios:

- The number of heads/tails in a sequence of coin flips.
- Vote counts for two different candidates in an election.
- The number of male/female employees in a company.
- The number of accounts that are in compliance or not in compliance with an accounting procedure.
- The number of successful sales calls.
- The number of defective products in a production run.
- The number of days in a month your company's computer network experiences a problem.

1) Mean $\mu = E(X) = np$
2) Variance $\sigma^2 = npq$
3) Standard Deviation $\sigma = \sqrt{npq}$

2.2.5) The Poisson distribution

This is a discrete distribution that is used in two situations. It is used, when certain conditions are met, as a probability distribution in its own right, and it is also used as a convenient approximation to the binomial distribution in some circumstances. The distribution is named for **S.D. Poisson**, a French mathematician of the nineteenth century.

Consider the following scenarios:

- The hourly number of customers arriving at a bank
- The daily number of accidents on a particular stretch of highway
- The hourly number of accesses to a particular web server
- The daily number of emergency calls in Dallas
- The number of typos in a book
- The monthly number of employees who had an absence in a large company
- Monthly demands for a particular product

Examples of occurrences to which the Poisson distribution often applies include counts from a *Geiger counter*, *collisions of cars* at a specific intersection under specific conditions, *flaws in a casting*, and *telephone calls* to a particular telephone or office under particular conditions. For the Poisson distribution to apply to these outcomes, they must occur randomly.

$$P(X) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x = 0, 1, 2, \dots$$

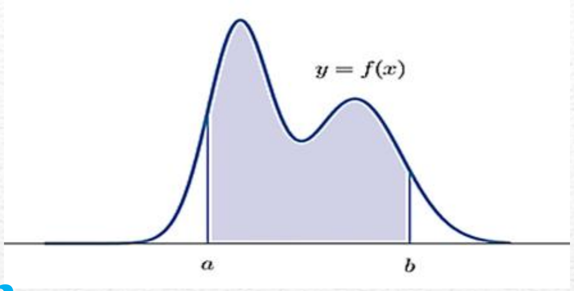
where λ is the average number of occurrences per base unit, and t is the number of base units inspected

1) Mean $\mu = E(X) = \lambda$

2) Variance $\sigma^2 = \lambda$

3) Standard Deviation $\sigma = \sqrt{\lambda}$

Continuous Random Variables



A **continuous random variable** X can take any value in an interval of the real number line like an interval [a,b].

Continuous Probability Distribution Function

A continuous random variable has a probability of 0 of assuming *exactly* any of its values. Consequently, its probability distribution cannot be given in tabular form.

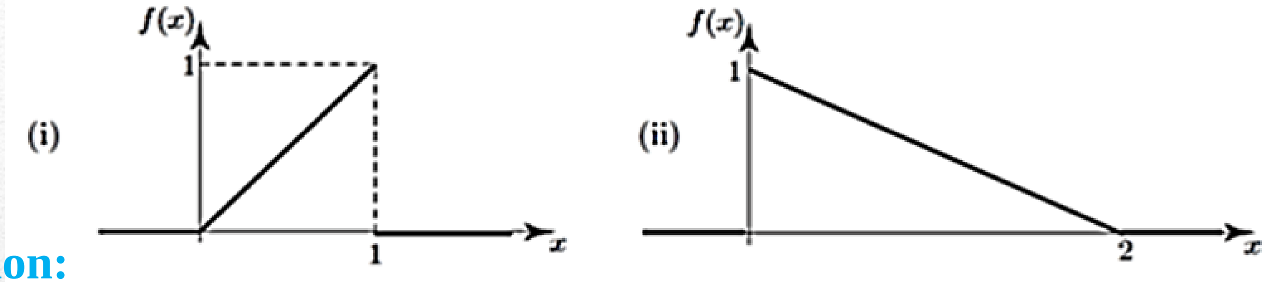
The function f(x) is a probability density function (pdf) for the continuous random variable X, defined over the set of real numbers, if

$$1)\ f(x) \geq 0, for\ all\ x\ \in \mathbb{R}.$$

$$2)\ \int_{-\infty}^{\infty} f(x)\,dx = 1.$$

$$3)\ P(a < X < b) = \int_{a}^{b} f(x)\,dx = F(b) - F(a).$$

Example 2.21: Which of the following are not probability density functions?



Solution:

i) We can write $f(x)=\begin{cases} x, & 0\le x\le 1 \\ 0, & \text{elsewhere}\end{cases}$ $\quad f(x)\ge 0, for\ all\ x\ \in \mathbb{R}$

$$but\quad \int_{-\infty}^{\infty} f(x)\,dx=\int_0^1 x\,dx=\frac{1}{2}\neq 1.$$

Thus this function is not a valid pdf because the integral's value is not 1.

ii) We have $f(x)\ =\begin{cases} 1-\frac{1}{2}x & 0\le x\le 2 \\ 0 & \text{elsewhere}\end{cases}$

$$f(x)\ge 0, for\ all\ x\ \in \mathbb{R}$$

$$\int_{-\infty}^{\infty} f(x)\,dx=\int_0^2\left(1-\frac{1}{2}x\right)dx=1.$$

(Alternatively, the area of the triangle is 1/2× 1 × 2 = 1) This implies that $f(x)$ is a valid pdf

iii) $f(x) = \begin{cases} x^2 - 4x + \frac{10}{3}, & 0 \le x \le 3 \\ 0\,, & \text{elsewhere} \end{cases}$

iii) $\int_{-\infty}^{\infty} f(x)\,dx = \int_0^3 (x^2 - 4x + \frac{10}{3})\,dx \;\; = [x^2/3 - 2x^2 + 10/3\;x]_0^3 = 1$

but, $f(x) < 0, for\;\; some\; x \in\;]0,3[$ hance $f(x)$ is not pdf.

The Cumulative Distribution Functions

The **cumulative distribution function** *F*(*x*) of a continuous random variable *X* with density function *f*(*x*) is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt, \qquad for -\infty < x < \infty.$$

Example 2.23: The cumulative distribution function of random variable X is

$$F(x) = \begin{cases} 0 & x < -1 \\ \dfrac{x+1}{2} & -1 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

Find the pdf f(x) of X.

Solution: From the CDF, we can find the PDF by direct differentiation.

$$f(x) = \frac{dF(x)}{dx} = \begin{cases} \dfrac{1}{2} & -1 \leq x < 1, \\ 0 & elsewhere \end{cases}.$$

Mean and Variance of a Continuous Random Variable

Suppose X is a continuous random variable with probability density function $f(x)$.

The **mean** or **expected value** of X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \int_{-\infty}^{\infty} x\, f(x)dx$$

The **variance** of X, denoted as $Var(X)$ or σ^2 is

$$\sigma^2 = Var(X)$$

$$= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$$

The **standard deviation** of X is $\sigma = \sqrt{\sigma^2}$.

Example 2.25: For the R.V X with pdf

$$f(x) = \begin{cases} \frac{1}{2}x & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Find E(X) and Var(X)

$$E(X) = \int_0^2 x\left(\frac{1}{2}x\right)\,dx = \left[\frac{1}{6}x^3\right]_0^2 = \frac{4}{3}$$

$$E(X^2) = \int_0^2 x^2\left(\frac{1}{2}x\right)\,dx = \left[\frac{1}{8}x^4\right]_0^2 = 2$$

$$Var(X) = E(X^2) - \left(E(X)\right)^2 = 2 - \frac{16}{9} = \frac{2}{9}.$$

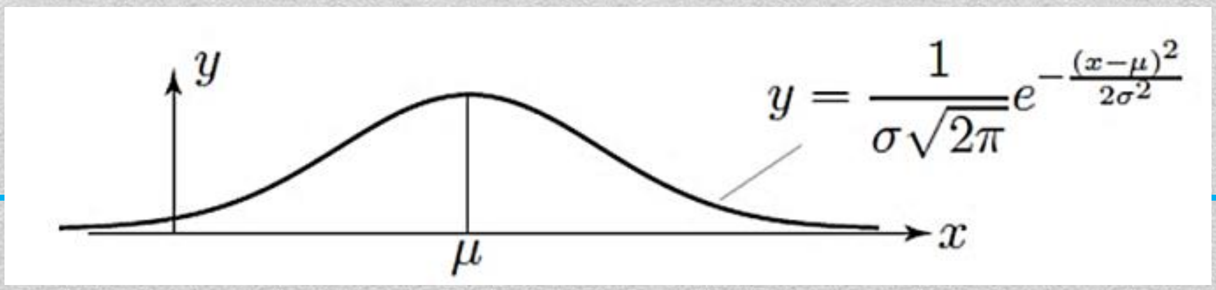
Normal Distribution

The Normal (or Gaussian) distribution is a function gives the probability that an event will fall between any two real number limits as the curve approaches zero on either side of the mean. Area underneath the normal curve is always equal to 1. The curve itself is approximately bell shaped, and is therefore informally called the Bell Curve.

The probability density function of a normal distribution with mean μ and variance σ^2 is given by the formula

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

This curve is always bell-shaped with the center of the bell located at the value of μ. The height of the bell is controlled by the value of σ



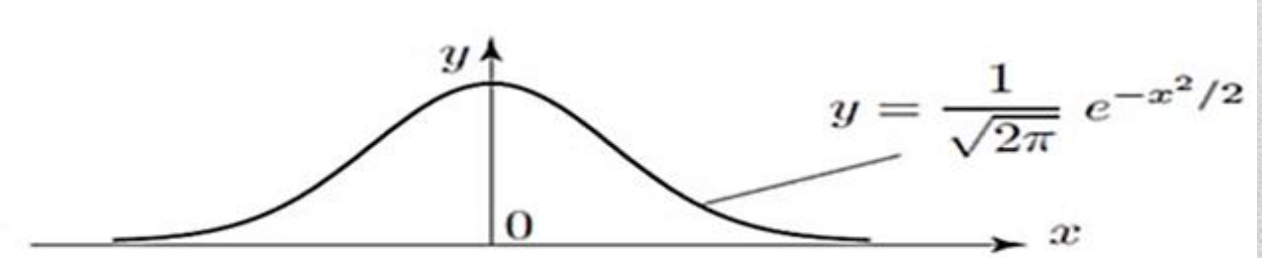
The Standard Normal Distribution

At this stage we shall, for simplicity, consider what is known as a standard normal distribution which is obtained by choosing particularly simple values for μ and σ^2.

The standard normal distribution has a mean of zero and a variance of one.

In the following Figure we show the graph of the standard normal distribution which has probability density function

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



The standard normal distribution curve

f the behaviour of a continuous random variable X is described by the distribution $N(\mu, \sigma^2)$ then the behaviour of the random variable $Z = \frac{\boldsymbol{X - \mu}}{\boldsymbol{\sigma}}$ is described by the standard normal distribution $N(0, 1)$.

We call Z the standardised normal variable and we write

$$Z \sim N(0, 1)$$

Example 2.28: If the random variable X is described by the distribution $N(45, 0.000625)$ then, what is the transformation required to obtain the standardised normal variable

Solution:

$\mu = 45\ and\ \sigma^2 = 0.000625$ so that $\sigma = 0.025$. Hence $Z = \frac{X-45}{0.025}$ is the required transformation.

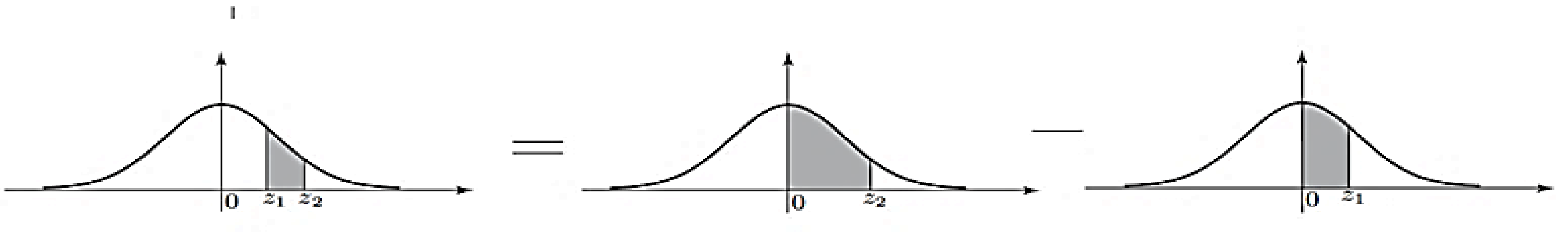
z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3304	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.483	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.8	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999

Case 1

Direct from table

Case 2

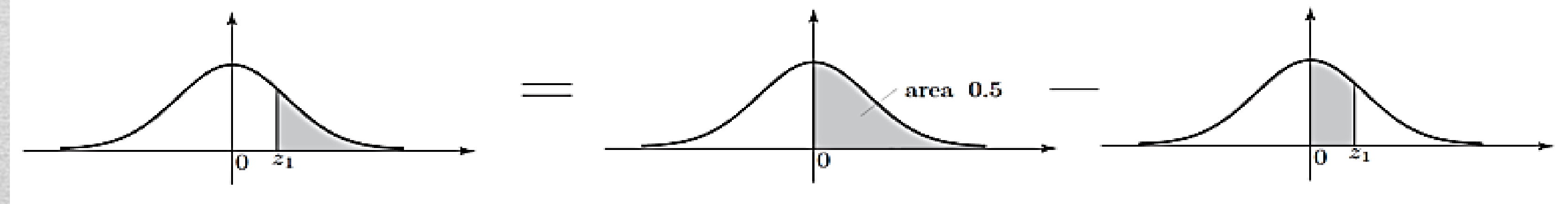
The figure illustrates what we do if both Z values are positive. By using the properties of the standard normal distribution we can organise matters so that any required area is always of 'standard form'.



Here the shaded region can be represented by the difference between two shaded areas.

Case 3

The following diagram illustrates the procedure to be followed when finding probabilities of the form $P(Z > z_1)$.

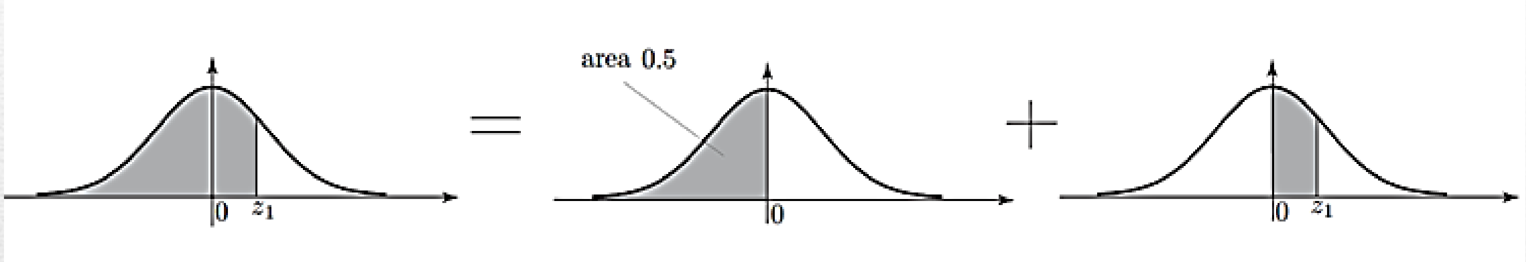


This time the shaded area is the difference between the right-hand half of the total area and an area which can be read off from Table.

Case 4

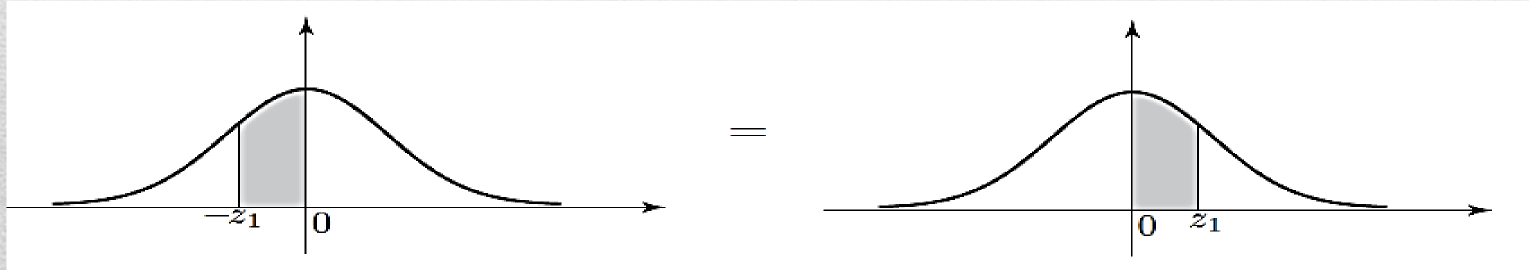
Here we consider the procedure to be followed when calculating probabilities of the form $P(Z < z_1)$.

Here the shaded area is the sum of the left-hand half of the total area and a ‘standard’ area.



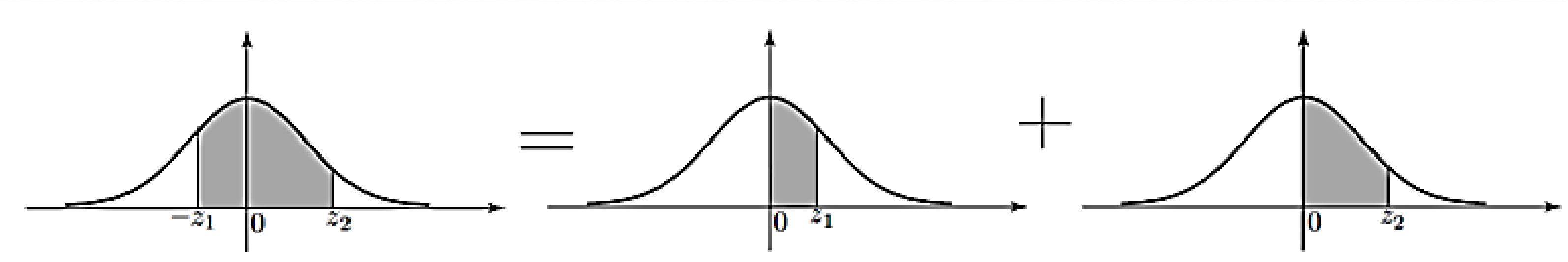
Case 5

Here we consider what needs to be done when calculating probabilities of the form $P(-z_1 < Z < 0)$ where z_1 is positive. This time we make use of the symmetry in the standard normal distribution curve



Case 6

Finally we consider probabilities of the form $P(-z_2 < Z < z_1)$. Here we use the sum property and the symmetry property.



Example 2.38: The resistance of a strain gauge is normally distributed with a mean of 100 ohms and a standard deviation of 0.2 ohms. To meet the specification, the resistance must be within the range 100 ± 0.5 ohms.What percentage of gauges are unacceptable?

z	0	0.01
0	0	0.0040
0.1	0.0398	0.0438
0.2	0.0793	0.0832
0.3	0.1179	0.1217
0.4	0.1554	0.1591
0.5	0.1915	0.1950
0.6	0.2257	0.2291
0.7	0.2580	0.2611
0.8	0.2881	0.2910
0.9	0.3159	0.3186
1.0	0.3413	0.3438
1.1	0.3643	0.3665
1.2	0.3849	0.3869
1.3	0.4032	0.4049
1.4	0.4192	0.4207
1.5	0.4332	0.4345
1.6	0.4452	0.4463
1.7	0.4554	0.4564
1.8	0.4641	0.4649
1.9	0.4713	0.4719
2.0	0.4772	0.4778
2.1	0.4821	0.4826
2.2	0.4861	0.4864
2.3	0.4893	0.4896
2.4	0.4918	0.4920
2.5	0.4938	0.4940

Solution:

$x_1 = 99.5$, $x_2 = 100.5$ $\quad Z = \dfrac{X-\mu}{\sigma} = \dfrac{X-100}{0.2}$

so that $z_1 = -2.5$ and $z_2 = 2.5$

$P(-2.5 < Z < 2.5) = 2\,P(0 < Z < 2.5)$

$$= 2 \times 0.4938 = 0.9876.$$

Hence the proportion of acceptable gauges is 98.76%.

Therefore the proportion of unacceptable gauges is 1.24%.

Example 2.39: A sample of 250 students takes the final exam test at 2^{nd} year in Benha faculty of engineering, the test score normally distributed with a mean of 70 and a standard deviation of 8.

a) How many students take score from 62 to 78?

b) How many students exceed 80?

Solution: $a)\ P(62 < X < 78) = P(\frac{62-70}{8} < \frac{X-\mu}{\sigma} < \frac{78-70}{8})$

$$= P(-1 < Z < 1)$$

z	0	0.01
0	0	0.0040
0.1	0.0398	0.0438
0.2	0.0793	0.0832
0.3	0.1179	0.1217
0.4	0.1554	0.1591
0.5	0.1915	0.1950
0.6	0.2257	0.2291
0.7	0.2580	0.2611
0.8	0.2881	0.2910
0.9	0.3159	0.3186
1.0	0.3413	0.3438
1.1	0.3643	0.3665

$$= 2P(0 < Z < 1)$$

$$= 2 * 0.3413 = 0.6826$$

The number of students=250*0.6826=170 student

$$b)\ P(X > 80) = P\left(\frac{X-\mu}{\sigma} > \frac{80-70}{8}\right)$$
$$= P(Z > 1.25)$$
$$= 0.5 - P(0 < Z < 1.25)$$
$$= 0.5 - 0.3944 = 0.1056$$

The number of students=250*0.1056=26 student

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3304	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015

THANK YOU