Probability Theory And Statistics

The Course of Lectures.

Lecture 9

Lecture 9

Uniform Distribution. Normal (Gaussian Distribution) Distributions.

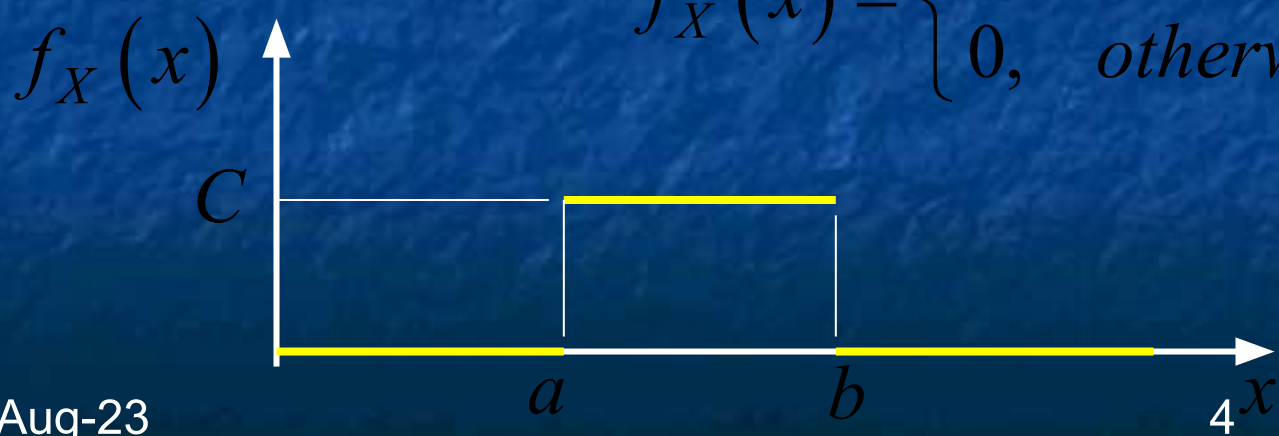
Contents

- Uniform distribution – the general view
- The Normal Distribution

Uniform Distribution: general view

- The random variable X is said to have uniform distribution on $[a,b]$, if its probability density function $f_X(x)$ is
  - constant for $x \in [a,b]$
  - and is equal to 0 for $x \notin [a,b]$

$$f_X(x) = \begin{cases} C, & for \quad a \le x \le b \\ 0, & otherwise \end{cases}$$



To find the constant C:

- We use the property of continuous random variable
  - (this property sometimes is called normalizing condition)

$$\int_{-\infty}^{+\infty} f_X(x)\,dx = 1$$

$$\underbrace{\int_{-\infty}^{a} 0\,dx}_{=0} + \int_{a}^{b} C\,dx + \underbrace{\int_{b}^{+\infty} 0\,dx}_{=0} = 1$$

$$C\int_{a}^{b} dx = 1; \quad C\,x\Big|_{a}^{b} = 1; \quad C\left(b-a\right) = 1; \quad C = \frac{1}{b-a}$$

So we have in the general terms:

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & for \quad a \le x \le b \\ 0, & for\ all\ other\ values\ of\ x \end{cases}$$

- For example, if we are interesting in probability density function for uniform distribution on $[2,5]$

$$f_X(x) = \begin{cases} \dfrac{1}{5-2} = \dfrac{1}{3}, & for \quad 2 \le x \le 5 \\ 0, & for\ all\ other\ values\ of\ x \end{cases}$$

To find the expectation of the uniformly distributed random variable:

- We remember the definition of expectation for continuous random variable

$$E[X] = \int_{-\infty}^{+\infty} x\, f_X(x)\, dx$$

- Substituting

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & for \quad a \le x \le b \\ 0, & otherwise \end{cases}$$

we obtain

To find the expectation of the uniformly distributed random variable:

$$E(X) = \underbrace{\int_{-\infty}^{a} 0\,dx}_{=0} + \int_{a}^{b} \frac{1}{b-a} x\,dx + \underbrace{\int_{b}^{+\infty} 0\,dx}_{=0} =$$

$$= \frac{1}{b-a}\int_{a}^{b} x\,dx = \frac{1}{b-a}\frac{x^2}{2}\Big|_a^b = \frac{1}{b-a}\frac{1}{2}\left(b^2 - a^2\right) =$$

$$= \frac{1}{2}\frac{(b-a)(b+a)}{b-a} = \frac{b+a}{2}$$

For the uniform distribution $E(X) = \mu_X = \dfrac{a+b}{2}$

To find the variance of the uniformly distributed random variable:

- We remember the definition of variance for continuous random variable

$$Var\left(X\right)=\sigma_X^2=E\left[\left(X-\mu_X\right)^2\right]=$$

$$=\int_{-\infty}^{+\infty}\left(x-\mu_X\right)^2 f_X\left(x\right)dx$$

To find the variance of the uniformly distributed random variable:

- Substituting the value of $\mu_X = \dfrac{a+b}{2}$ and probability density function for uniform distribution

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & for \quad a \le x \le b \\ 0, & otherwise \end{cases}$$

We have

To find the variance of the uniformly distributed random variable:

$$Var(X) = E\left[\left(X - \mu_X\right)^2\right] = \int_{-\infty}^{+\infty} \left(x - \mu_X\right)^2 f_X(x)\,dx =$$

$$= \int_a^b \left(x^2 - 2x\mu_X + \mu_X^2\right) \frac{1}{b-a}\,dx =$$

$$= \frac{1}{b-a} \int_a^b \left(x^2 - 2x\mu_X + \mu_X^2\right) dx =$$

$$= \frac{1}{b-a} \left(\frac{x^3}{3}\Big|_a^b - 2\mu_X \frac{x^2}{2}\Big|_a^b + \mu_X^2 x\Big|_a^b \right) =$$

So the variance of the uniformly distributed random variable:

$$Var(X) = \frac{1}{b-a}\left(\frac{x^3}{3}\Big|_a^b - 2\mu_X \frac{x^2}{2}\Big|_a^b + \mu_X^2 x\Big|_a^b\right) =$$

$$= \frac{1}{b-a}\left(\frac{b^3 - a^3}{3} - \frac{a+b}{2}\left(b^2 - a^2\right) + \left(\frac{a+b}{2}\right)^2 (b-a)\right) =$$

$$= \ldots = \frac{(b-a)^2}{12}$$

It could be calculated as $Var(X) = \sigma_X^2 = \frac{(b-a)^2}{12}$

To find the cumulative distribution function for uniform distribution

- We use the definition of cumulative distribution function

$$F_X(x_0) = \int_{-\infty}^{x_0} f_X(x)\,dx$$

- So we have for 3 different intervals
- 1. For $x < a$

$$F_X(x) = \int_{-\infty}^{x} 0\,dx = 0$$

To find the cumulative distribution function for uniform distribution

- 2. For $x \in [a,b]$

$$F_X(x) = \underbrace{\int_{-\infty}^{a} 0\,dx}_{=0} + \int_{a}^{x} \frac{1}{b-a}\,dx = \frac{1}{b-a}\,x\Big|_{a}^{x} = \frac{x-a}{b-a}$$

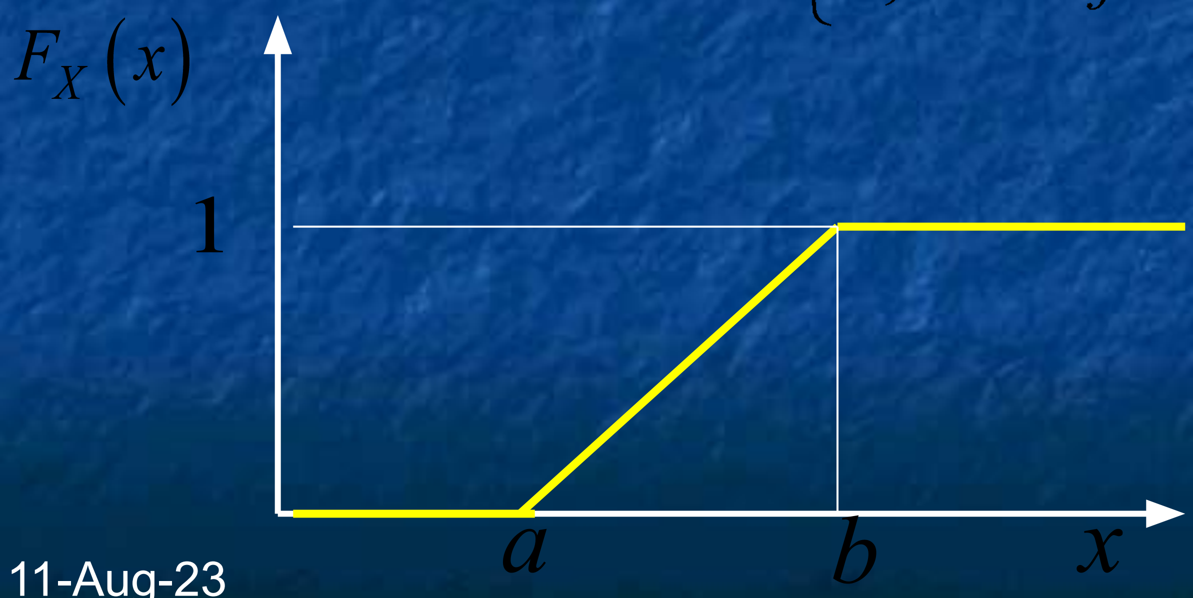
- 3. For $x > b$ $\quad F_X(x) =$

$$= \underbrace{\int_{-\infty}^{a} 0\,dx}_{=0} + \int_{a}^{b} \frac{1}{b-a}\,dx + \underbrace{\int_{b}^{x} 0\,dx}_{=0} = \frac{1}{b-a}\,x\Big|_{a}^{b} = \frac{b-a}{b-a} = 1$$

Finally we obtain the cumulative distribution function for uniform distribution

- We have

$$F_X(x) = \begin{cases} 0, & if \;\; x < a \\ \dfrac{x-a}{b-a}, & if \;\; x \in [a, b] \\ 1, & if \;\; x > b \end{cases}$$



The Normal Distribution

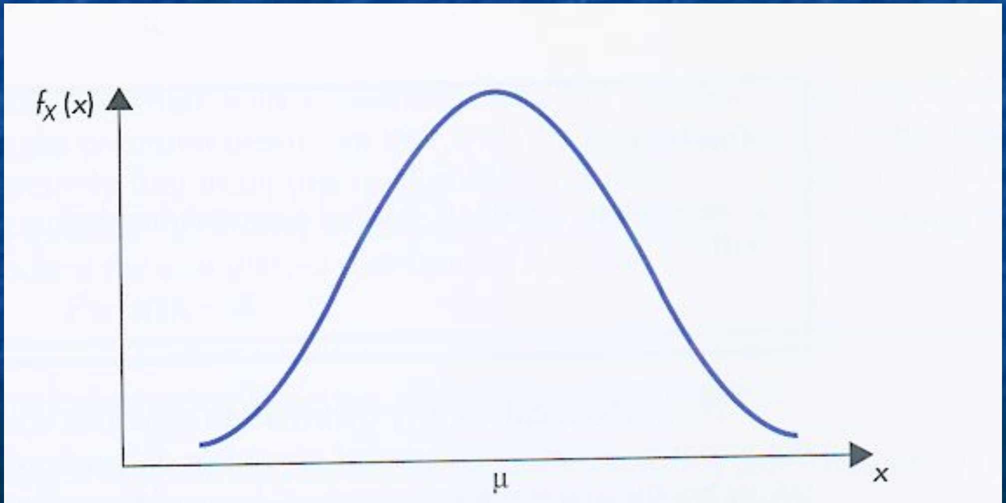
- We introduce now a continuous distribution that plays a central role in a very large body of statistical analysis.
- For example, suppose that a big group of students takes a test. A large proportion of their scores are likely to be concentrated about the mean, and the numbers of scores in ranges of a fixed width are likely to "tail off" away from the mean.

The Normal Distribution

- If the average score on the test is 60, we would expect to find, for instance, more students with scores in the range 55-65 than in the range 85-95
- These considerations suggest a probability density function that peaks at the mean and tails off at its extremities. One distribution with these properties is the **normal distribution**, whose probability density function is shown below.
- As can be see this density function is *bell-shaped.*

Probability Density Function of the Normal Distribution

- The shape of the probability density function is a symmetric bell-shaped curve centered on the mean μ, that peaks at the mean and tails off at its extremities



Probability Density Function of the Normal Distribution

- If the random variable X has probability density function

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} \quad for \ -\infty < x < +\infty$$

- where μ and σ^2 are any number such that $\mu \in (-\infty, +\infty)$ and $\sigma^2 \in (0, +\infty)$ and where e and π are physical constants, e = 2.71828 ... and π = 3.14159 ...,
- then X is said to follow a **normal distribution.**

Comments

- It can be seen from the definition that there is not a single normal distribution but a whole family of distributions, resulting from different specifications of μ and σ^2 .
- These two parameters have very convenient interpretations

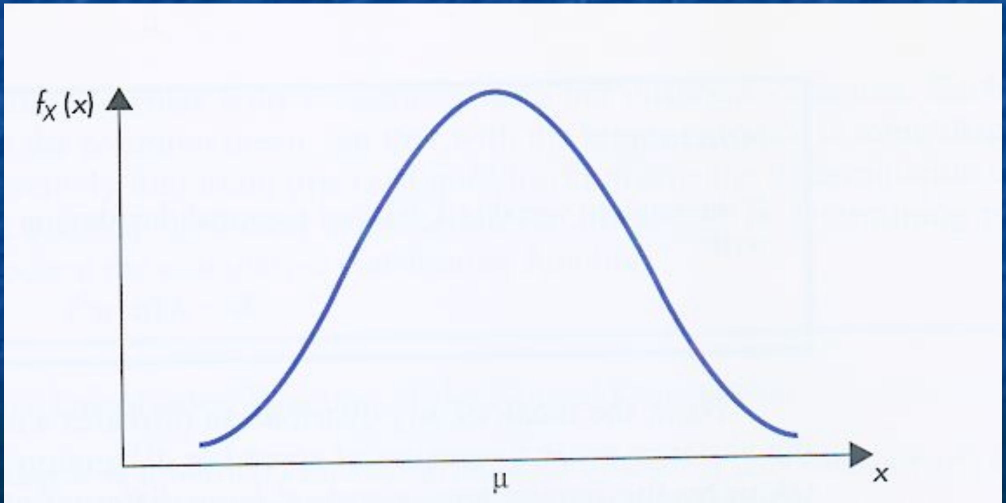
Some Properties of the Normal Distribution

- Suppose that the random variable X follows a normal distribution with parameters μ and σ^2. The following properties hold:
- (i) The mean of the random variable is μ ; that is $E(X) = \mu$
- (ii) The variance of the random variable is σ^2; that is

$$Var(X) = E\left[(X - \mu_X)^2\right] = \sigma^2$$

Some Properties of the Normal Distribution

- (iii) The shape of the probability density function is a symmetric bell-shaped curve centered on the mean μ .



Comments & Notation

- It follows from these properties that given the mean and variance of a normal random variable, an individual member of the family of normal distributions is specified.
- This allows use of a convenient notation.
- If the random variable X follows a normal distribution with mean μ and variance σ^2, we write

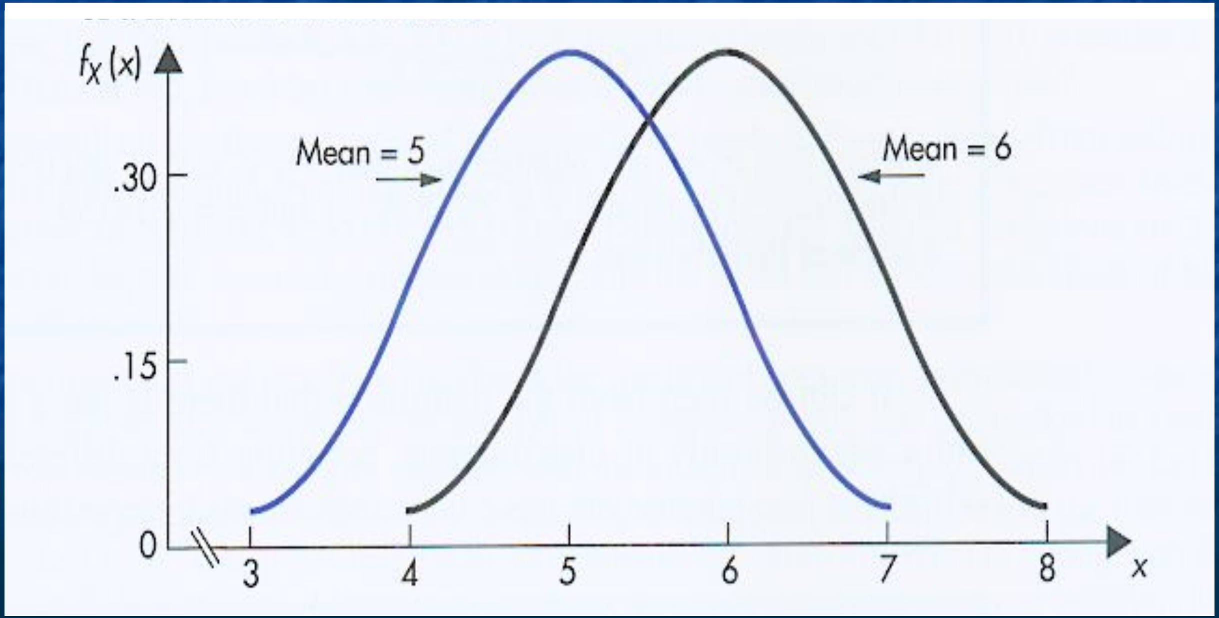
$$X \sim N\left(\mu, \sigma^2\right)$$

Comments

- Now,
- the mean of any distribution provides a measure of central location,
- while the variance gives a measure of spread or dispersion about the mean.
- Thus, the values taken by the parameters μ and σ^2 have different effects on the probability density function of a normal random variable.

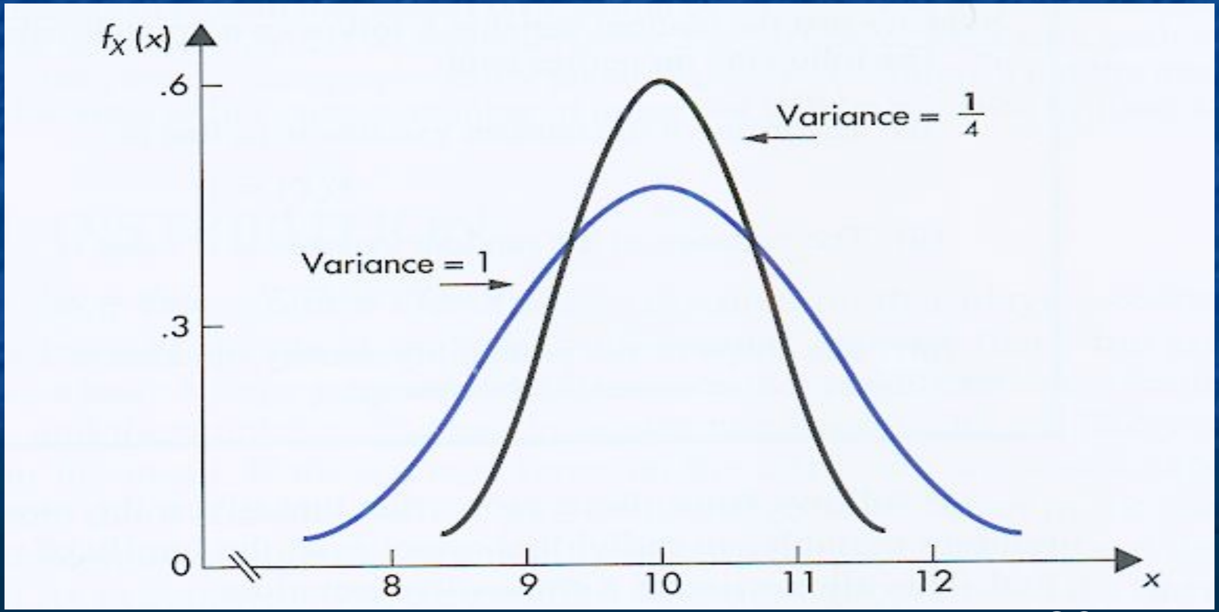
Comments

- We shows probability density functions for two normal distributions with a common variance but different means.
- It can be seen that increasing the mean while holding the variance fixed shifts the density function but does not alter its shape.



Comments

- The two density functions are of normal random variables with a common mean but different variances.
- Each is symmetric about the common mean, but that with the larger variance is more disperse



Cumulative Distribution Function of the Normal Distribution

- An extremely important practical question concerns the determination of probabilities from a specified normal distribution.
- As a first step in determining probabilities, we introduce the cumulative distribution function

Cumulative Distribution Function of the Normal Distribution

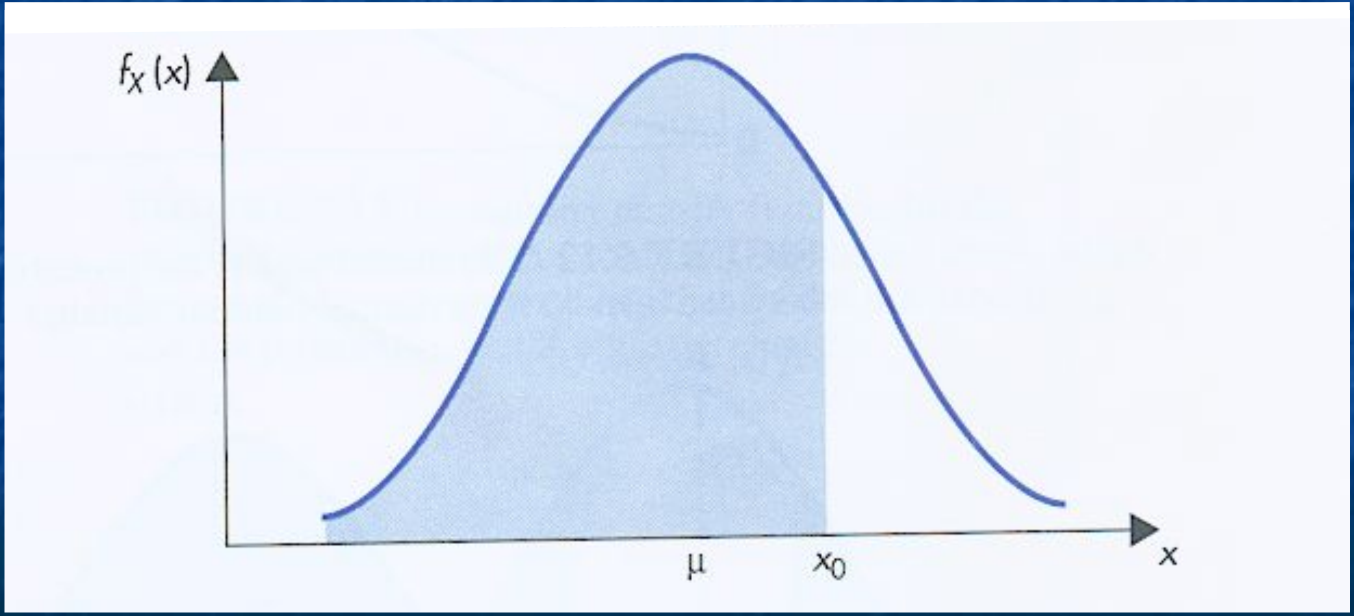
- Suppose that X is a normal random variable with mean μ and variance σ^2
- that is, $X \sim N\left(\mu, \sigma^2\right)$
- Then the **cumulative distribution function**

$$F_X\left(x_0\right) = P\left(X \leq x_0\right)$$

- This is the area under the probability density function to the left of x_0
- As for any proper density function, the total area under the curve is 1; that is $F_X\left(\infty\right) = 1$

Cumulative Distribution Function of the Normal Distribution

- The shaded area is the probability that X does not exceed x_0 for a normal random variable



Cumulative Distribution Function of the Normal Distribution

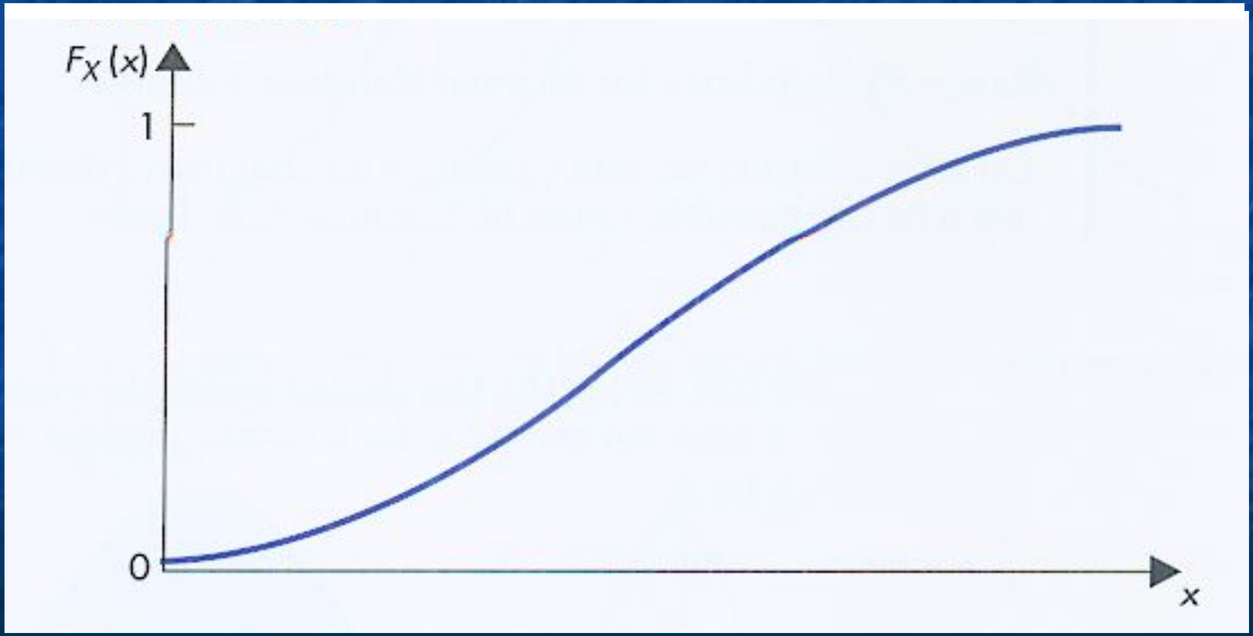
- There is no simple algebraic expression for calculating the cumulative distribution function of a normally distributed random variable.
- That is to say that the integral

$$F_X\left(x_0\right) = \int_{-\infty}^{x_0} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2 / 2\sigma^2} dx$$

does not have a simple algebraic form

Cumulative Distribution Function of the Normal Distribution

- The general shape of the cumulative distribution function is shown below



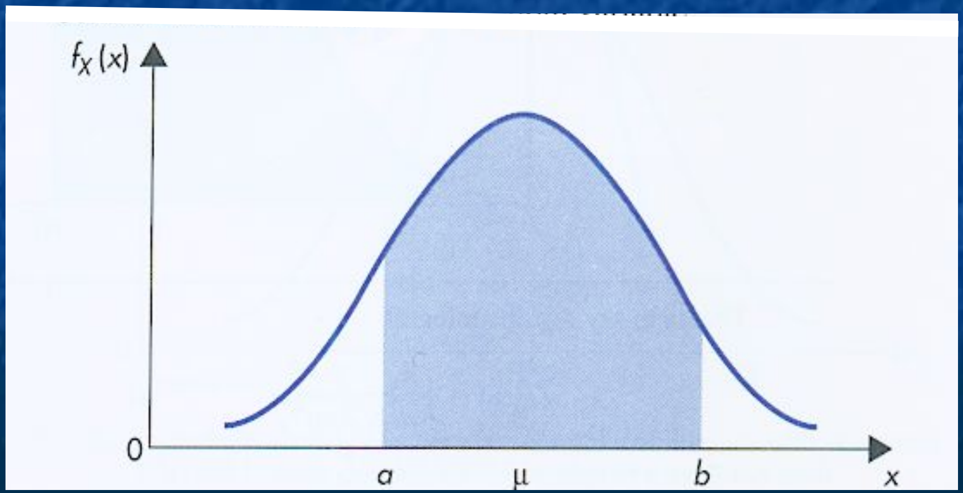
Range Probabilities for Normal Random Variables

- We have already seen
- that for *any* continuous random variable, probabilities can be expressed in terms of the cumulative distribution function

Range Probabilities for Normal Random Variables

- Let X be a normal random variable with cumulative distribution function $F_X(x)$, and let a and b be two possible values of X, with $a < b$.

  Then $$P(a < X < b) = F_X(b) - F_X(a)$$



Range Probabilities for Normal Random Variables

- Any required probability can be obtained from the cumulative distribution function.
- However, a crucial difficulty remains because there does not exist a convenient formula for determining the cumulative distribution function.
- In principle, for any specific normal distribution, probabilities could be obtained by numerical methods using an electronic computer.

Range Probabilities for Normal Random Variables

- However, it would be enormously tedious if we had to carry out such an operation for every normal distribution we encountered.
- Fortunately, probabilities for *any* normal distribution can always be expressed in terms of probabilities for a *single* normal distribution for which the cumulative distribution function has been evaluated and tabulated.

The Standard Normal Distribution

- We now introduce the particular distribution that is used for this purpose

- Let Z be a normal random variable with mean 0 and variance 1; that is

$$X \sim N\left(0, 1^2\right)$$

- Then Z is said to follow the **standard normal distribution.**

The Standard Normal Distribution

- If the cumulative distribution function of this random variable is denoted $F_Z(z)$,
  and *a** and *b** are two numbers with *a** < *b**,
  then

$$P\left(a^* < X < b^*\right) = F_Z\left(b^*\right) - F_Z\left(a^*\right)$$

  The cumulative distribution function of the standard normal distribution is tabulated.

The table of Normal Distribution

- This table gives values of $F_Z(z) = P(Z \le z)$

  for nonnegative values of z.
- For example $F_Z(1.25) = 0.8944$

- Thus, the probability is 0.8944 that the standard normal random variable takes a value less than 1.25

NORMAL DISTRIBUTION

The function tabulated is the cumulative distribution function of a standard $N(0,1)$ random variable, namely

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt.$$

If X is distributed $N(0,1)$ then $\Phi(x) = Pr.(X \leq x)$.

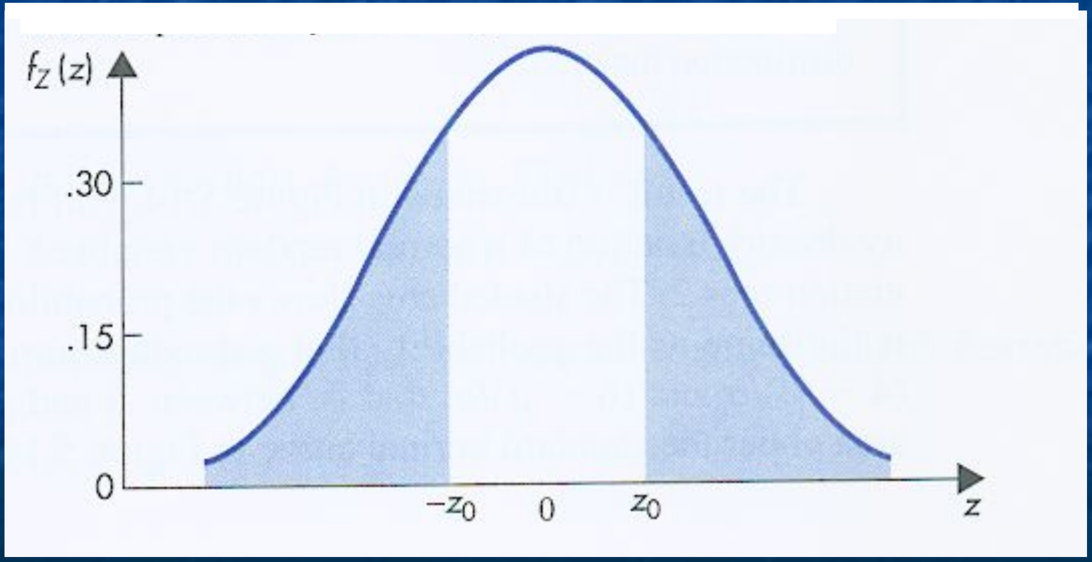
x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936

Values of the cumulative distribution function for negative values of z can be inferred from the symmetry of the probability density function

- Let z_0 be any positive number, and suppose that we require $F_Z(-z_0) = P(Z \leq -z_0)$
- The density function of the standard normal random variable is symmetric about its mean, 0, the area under the curve to the left of $-z_0$ is the same as the area under the curve to the right of z_0; that is $P(Z \leq -z_0) = P(Z \geq z_0)$

Probability density function for the standard normal random variable Z;

- the shaded areas, which are equal, show the probability that Z does not exceed $-z_0$ and the probability that Z is greater than z_0



Moreover, since the total area under the curve is 1:

$$P\left(Z \geq z_0\right) = 1 - P\left(Z \leq z_0\right) = 1 - F_Z\left(z_0\right)$$

- Hence, it follows that $F_Z\left(-z_0\right) = 1 - F\left(z_0\right)$

- For example

$$P\left(Z \leq -1.25\right) = F_Z\left(-1.25\right) = 1 - F_Z\left(1.25\right) =$$

$$= 1 - 0.8944 = 0.1056$$

Example

- If Z is a standard normal random variable, find

$$P\left(-0.50 < Z < 0.75\right)$$

- The required probability is

$$P\left(-0.50 < Z < 0.75\right) = F_Z\left(0.75\right) - F_Z\left(-0.50\right) =$$

$$= F_Z\left(0.75\right) - \left(1 - F_Z\left(0.50\right)\right)$$

- Then, using Table we obtain

$$P\left(-0.50 < Z < 0.75\right) =$$

$$= 0.7734 - \left(1 - 0.6915\right) = 0.4649$$

How can probabilities for any normal random variable be expressed in terms of those for the standard normal random variable?

- Let the random variable X be normally distributed with mean μ and variance σ^2.
- We know that subtracting the mean and dividing by the standard deviation yields a random variable Z that has mean 0 and variance 1.
- It can also be shown that if X is normally distributed, so is Z.
- Hence, Z has a standard normal distribution

How can probabilities for any normal random variable be expressed in terms of those for the standard normal random variable?

- Suppose, then, that we require the probability that *X* lies between the numbers *a* and *b.*
- This is equivalent to $(X-\mu)/\sigma$ lying between $(a-\mu)/\sigma$ and , $(b-\mu)/\sigma$
- so that the probability of interest is

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) =$$

$$= P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

Finding Range Probabilities for Normal Random Variables

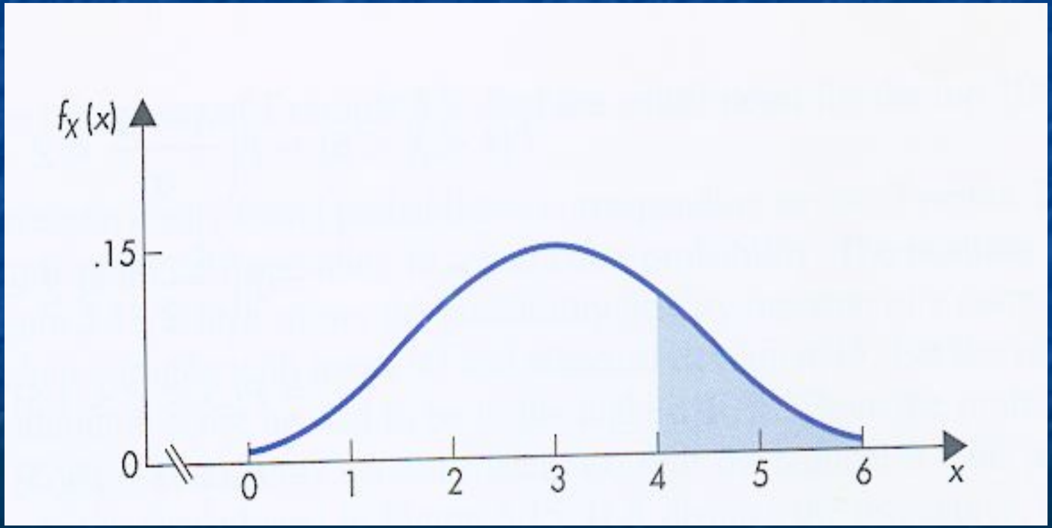
- Let X be a normal random variable with mean μ and variance σ^2
- Then the random variable $Z = (X - \mu)/\sigma$ has a standard normal distribution; that is, $Z \sim N(0, 1^2)$
- It follows that if *a* and *b* are any numbers with *a* < *b*, then $P(a < X < b) =$

$$= P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = F_Z\left(\frac{b-\mu}{\sigma}\right) - F_Z\left(\frac{a-\mu}{\sigma}\right)$$

  - where Z is the standard normal random variable and $F_Z(z)$ denotes its cumulative distribution function

Probability density function for normal random variable X with mean 3 and standard deviation 2; shaded area is probability that X lies between 4 and 6

- Figure shows the probability density function of a normal random variable *X* with mean $\mu = 3$ and standard deviation $\sigma = 2$



Probability density function for normal random variable X with mean 3 and standard deviation 2; shaded area is probability that X lies between 4 and 6

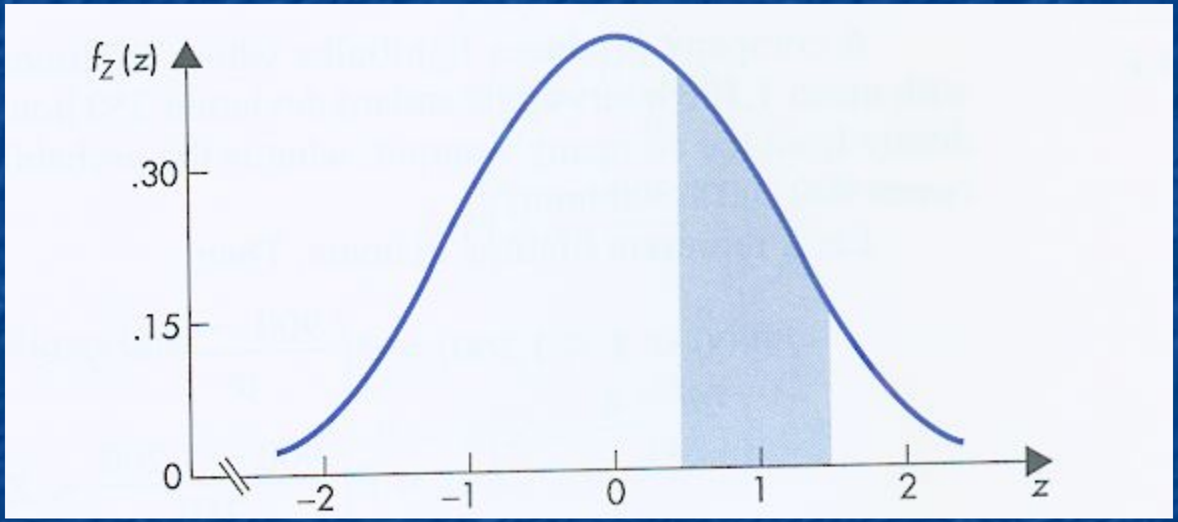
- The shaded area shows the probability that *X* lies between 4 and 6.
- This is the same as the probability that a standard normal random variable lies between

  $(4-\mu)/\sigma$ and $(6-\mu)/\sigma$

  that is, between 0.5 and 1.5.

  This probability is the shaded area under the standard normal curve

Probability density function for standard normal random variable Z; shaded area is probability that Z lies between 0.5 and 1.5 and is equal to shaded area in the previous slide



Example

- If X ~ *N(15,* 16), find the probability that X is larger than 18.
- This probability is $P(X > 18) =$

$$P\left(Z > \frac{18-\mu}{\sigma}\right) = P\left(Z > \frac{18-15}{4}\right) =$$

$$P(Z > 0.75) = 1 - P(Z < 0.75) = 1 - F_Z(0.75)$$

- From the Table we have $F_Z(0.75) = 0.7734$
- so

$$P(X > 18) = 1 - 0.7734 = 0.2266$$

Example

- If X is normally distributed with mean 3 and standard deviation 2, find P(4 < X < 6).
- We have $P(4 < X < 6) = P\left(\frac{4-\mu}{\sigma} < Z < \frac{6-\mu}{\sigma}\right)$

$$= P\left(\frac{4-3}{2} < Z < \frac{6-3}{2}\right) = P(0.5 < Z < 1.5) =$$

$$= F_Z(1.5) - F_Z(0.5) = 0.9332 - 0.6915 = 0.2417$$

Example

- A company produces lightbulbs whose lifetimes follow a normal distribution with mean 1,200 hours and standard deviation 250 hours.
- If a lightbulb is chosen randomly from the company's output, what is the probability that its lifetime will be between 900 and 1,300 hours?
- Let X represent lifetime in hours

Example.

- Then $$P(900 < X < 1300) =$$

$$= P\left(\frac{900-\mu}{\sigma} < Z < \frac{1300-\mu}{\sigma}\right) =$$

$$= P\left(\frac{900-1200}{250} < Z < \frac{1300-1200}{250}\right) =$$

$$= P(-1.2 < Z < 0.4) = F_Z(0.4) - F_Z(-1.2) =$$

$$= 0.6554 - (1 - 0.8849) = 0.5403$$

- Hence, the probability is approximately 0.54 that a lightbulb will last between 900 and 1,300 hours

Example

- A very large group of students obtains test scores that are normally distributed with mean 60 and standard deviation 15.
- What proportion of the students obtained scores between 85 and 95?
- Let X denote the test score.

Example

- Then we have

$$P(85 < X < 95) = P\left(\frac{85-\mu}{\sigma} < Z < \frac{95-\mu}{\sigma}\right) =$$

$$= P\left(\frac{85-60}{15} < Z < \frac{95-60}{15}\right) = P(1.67 < Z < 2.33) =$$

$$= F_Z(2.33) - F_Z(1.67) = 0.9901 - 0.9525 = 0.0376$$

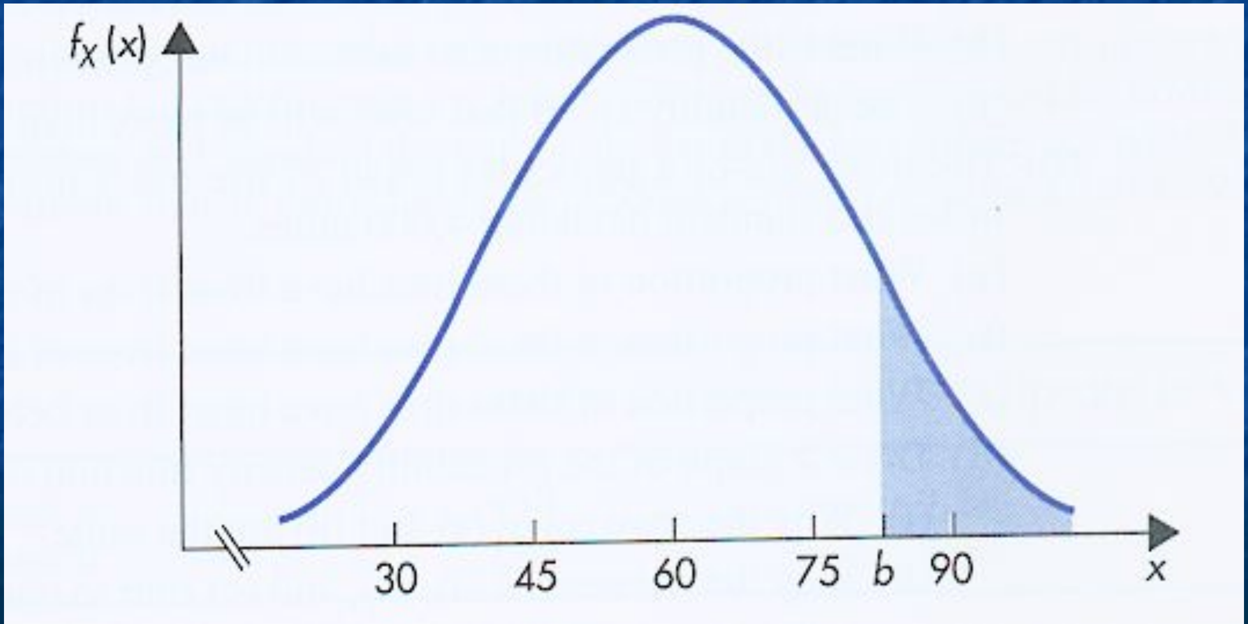
- That is, 3.76% of the students obtained scores in the range 85 to 95.

Example

- For the test scores of the previous Example, find the cutoff point for the top 10% of all students.
- We have previously found probabilities corresponding to cutoff points. Here we need the cutoff point corresponding to a particular probability.
- The position is illustrated in Figure (next slide), which shows the probability density function of a normally distributed random variable with mean 60 and standard deviation 15.

Example

- The probability is 0.10 that the random variable X exceeds the number b;
- Here X is normally distributed, with mean 60 and standard deviation 15



Example

- Let the number b denote the minimum score needed to be in the highest 10%. Then, the probability is 0.10 that the score of a randomly chosen student exceeds the number b.
- This probability is shown as the shaded area in Figure.
- If X denotes the test scores, then the probability that X exceeds b is 0.1,

Example

- So

$$0.1 = P\left(X > b\right) = P\left(Z > \frac{b-\mu}{\sigma}\right) = P\left(Z > \frac{b-60}{15}\right)$$

- Hence, it follows that

$$0.9 = P\left(Z < \frac{b-60}{15}\right) = F_Z\left(\frac{b-60}{15}\right)$$

- Now, from Table, if $F_Z\left(z\right) = 0.9$ then $z = 1.28$.

Example

- Therefore, we have $$\frac{b-60}{15}=1.28$$
- So $b=79.2$
- The conclusion is that 10% of the students obtain scores higher than 79.2

Comments

- In Examples, if the scores awarded on the test were integers, the distribution of scores would be inherently discrete.
- Nevertheless, the normal distribution can typically provide an adequate approximation in such circumstances.
- We will see later that the normal distribution can often be employed as an approximation to discrete distributions.
- As a preliminary, we introduce in the next lecture a result that provides strong justification for the emphasis given to the normal distribution.

Thank you for your attention!