Chapter 6: Continuous Random Variables

UE 2.2.2 Probability Density Function (Section 6.1)

A continuous random variable, X, takes on all possible values in an interval. The probability distribution of a continuous random variable is described by a **probability density function** $f(x)$.

A curve (of function) is called a **probability density function** $f(x)$ if

1. It lies on or above the horizontal axis, i.e., $f(x) \geq 0$ for all x.

2. Total area under the curve is equal to 1.

The probability of an event is the **area under the curve** for the values of X that make up the event.

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Draw a picture:

In general, we need calculus to find the probability. In this class, we will use geometry and tables to find the probability for two special types of continuous r.v.

Example: Suppose X = checkout time at a store, which is a random variable that is **uniformly distributed** on values between 5 and 20 minutes.

a. What does the density look like?

b. What is the probability that a person will take more than 10 minutes to check out?

c. What is the probability that a person will take exactly 10 minutes to check out?

d. Given a person has already spent 10 minutes checking out, what is the probability he will take no more than 5 additional minutes to check out?

Note:

- The probability for any individual outcome of a continuous r.v. is always _____

A *uniform* r.v. on the interval $[a, b]$ has probability density function

UE 2.3.2 Means and Variances of Continuous R.V.

Recall that for a discrete r.v. with probability function $f(x)$

the **mean** μ_X is

the **variance**, σ_X^2, is

the **standard deviation**, σ_X, is

Definition: **Mean and standard deviation of a continuous r.v.**

The mean is the "center" of its probability distribution.

mean = balancing point of the density curve

$$E(X) = \int x f(x) dx$$

(need calculus/integration to find it – integral instead of sums)

Variance and standard deviation are defined in the same way as in the discrete case.

the **variance**, σ_X^2, is

the **standard deviation**, σ_X, is

Note: The same rules for means and variances work for both discrete and continuous r.v.

**Example:** X = checkout time at a store

What is the expected time to check out at this store?

Example: X is a continuos r.v. with mean 2.5 and s.d. 1.5.

If $Y = 1 + 2X$, what are the mean and s.d. of Y?

UE 2.6 The Normal Distribution (Section 6.2, 6.3)

The Normal Distribution is the most important family of distributions because many variables have this shape and form approximately and many statistics that we use in our inference methods are based on the sample mean $\overline{X}$ and sums or averages generally have (approximately) a normal distribution.
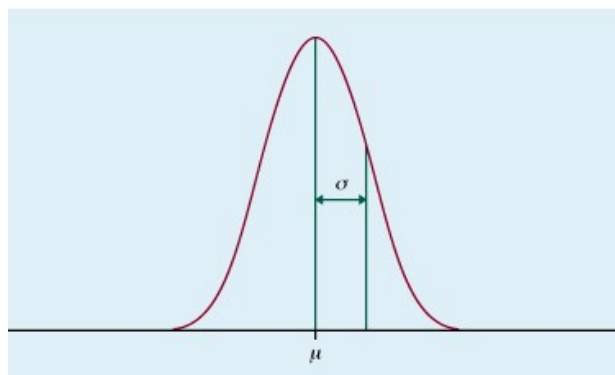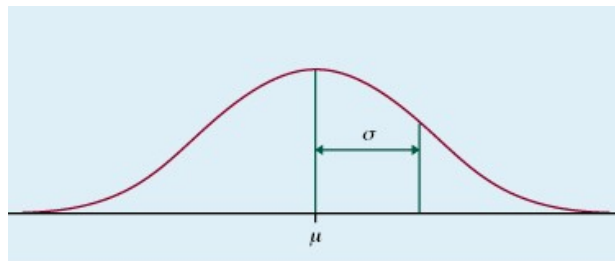
A Normal Curve is symmetric, unimodal, bell-shaped, centered at the mean μ and its spread is determined by the standard deviation σ. In fact, the change-of-curvature points on each side of the mean mark the values which are one standard deviation away from the mean.

A normal r.v. with mean μ and standard deviation σ has probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < \infty$$

Notation: The variable X is normally distributed with mean μ and standard deviation σ is denoted by $N(\mu, \sigma)$ (or $N(\mu, \sigma^2)$ as in UE).

Let's sketch two normal curves: $N(50, 10)$ and $N(80, 7)$

The Standard Normal Distribution

- is the normal distribution $N(0, 1)$ with mean 0 and standard deviation 1.

The **variable Z** is often used to denote the variable following the $N(0, 1)$ distribution.

Appendix A4 (pages 563-564) gives **areas under** the standard normal curve **to the left** of the points z.

Examples:

Table entry = 0.9192	Table entry = 0.0158 / Area = 0.9842
z = 1.40	z = −2.15
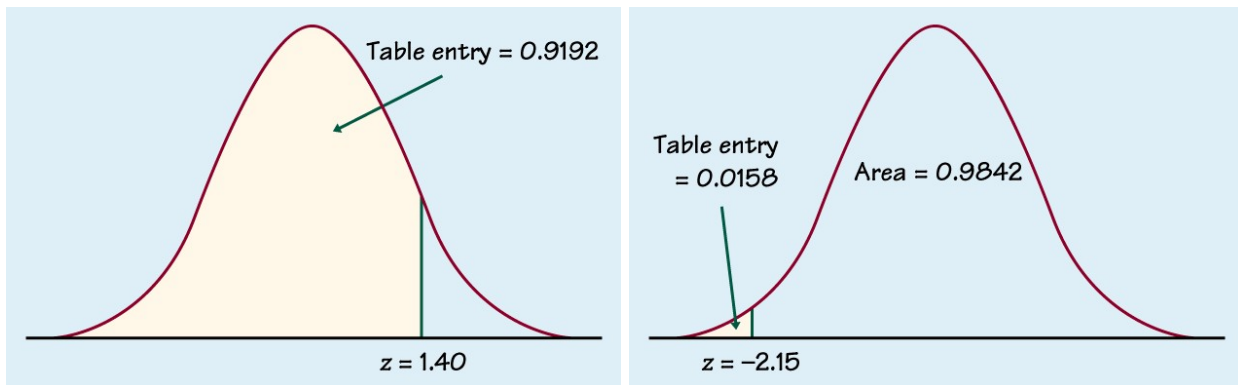
Example

1. What proportion of Z values are within 1 standard deviation of the mean?
$P(-1 \leq Z \leq 1)$?

2. What proportion of Z values are within 2 standard deviations of the mean?
$P(-2 \leq Z \leq 2)$?

3. What proportion of Z values are within 3 standard deviations of the mean?
$P(-3 \leq Z \leq 3)$?

4. $P(-1.5 < Z < 2.8)$

5. $P(Z > 4.8)$

6. What is the 90th percentile of the $N(0,1)$ distribution?

The first three results form what is called

The 68-95-99.7 Rule for any Normal Distribution

- **About 68% of the observations fall within 1 standard deviation of the mean**

- **About 95% of the observations fall within 2 standard deviations of the mean**

- **About 99.7% (nearly all) of the observations fall within 3 standard deviations of the mean**
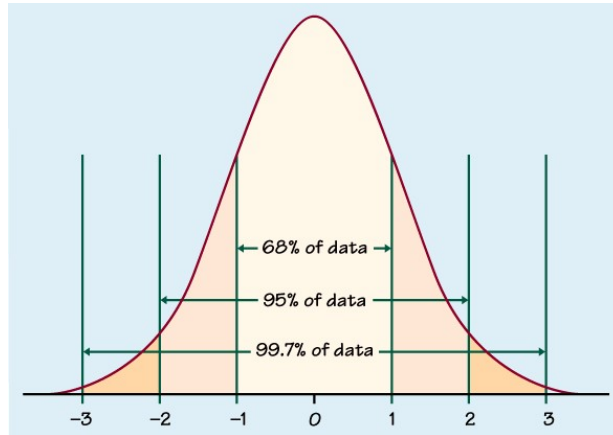
A useful result – can provide a **good frame of reference** without detailed calculation.

Example: Suppose the average test score is 70 with a standard deviation of 5 and you scored 85. If the distribution of scores is approximately normal then how good did you do?

We need one more idea to help us be able to find areas and thus proportions for any general $N(\mu, \sigma)$ distribution – the idea of standardization.

If the variable X has the $N(\mu, \sigma)$ distribution, then the standardized variable

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

will have the $N(0, 1)$ distribution.

The variable Z is sometimes referred to as the **Z-score** or standard score. Let's see how this standardization idea works through our next example.

Notation We use big Z or X to denote a **random variable** and little z or x to denote a **number**.

Example: Costs of Textbooks

The students at a local university spend a lot of money each term on textbooks. Suppose that the amount of money spent on textbooks for a term follows a normal distribution with a mean of $160 and a standard deviation of $20. Sketch this distribution below.

_____

a. The local bookstore will offer a t-shirt to any student who spends **more than $200** on textbooks. What proportion of students are eligible for the t-shirt?

b. Justin spent $170 on textbooks. **What percentile does this value of $170 correspond to?**

c. Complete the sentence: (show all work)

Based on the above model, **20% of the students spend less than** $_____ on textbooks.