`

Applied Data Analysis

PSC 200

Fall 2007

Note Set 4

Continuous Probability Distributions

Outline of Lecture

- Continuous Random Variables
- Expected Values
- Some Continuous Distributions

Continuous Random Variables

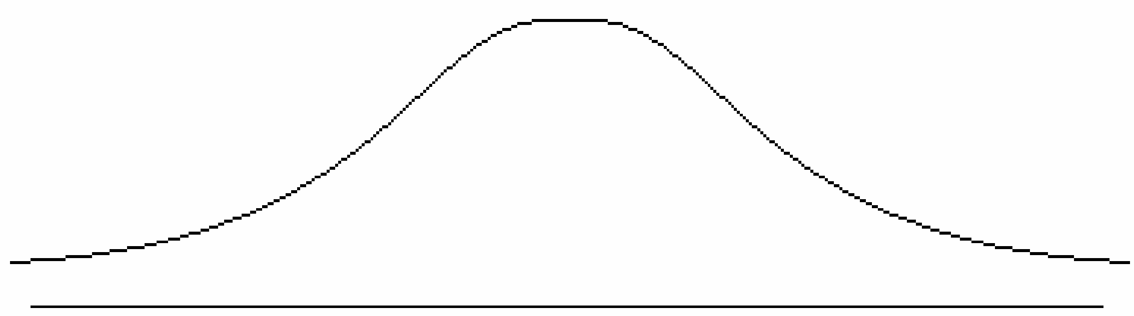
- Recall: A continuous random variable takes on an infinite continuum of values
- Examples:
  - Income
  - Weight of newborn babies
  - Sale price of a home
  - GDP of a country
  - Voter turnout in a district

Continuous Random Variables

- In note set 2, we explored ways of describing and summarizing the sample
- Like note set 3, note set 4 will explore ways of describing and summarizing the population
- Once again, we will use probability theory

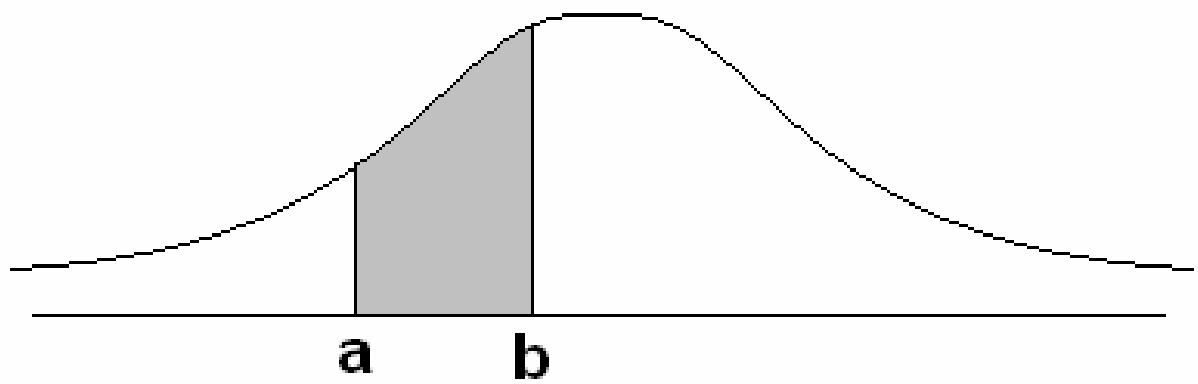
Continuous Random Variables

- We describe probability distributions using their probability density function
- We denote the probability density function of X using $f_X(x)$
- A probability density function is always positive, i.e. $f_X(x) \geq 0$
- The total area under the curve is equal to one



Continuous Random Variables

- The area between a and b indicates the probability that X is between a and b.
- We write the probability that $a \leq X \leq b$ as $\Pr(a \leq X \leq b)$



Expected Values

- In note set 2, we discussed the mean of a sample
- In note set 3, we discussed the expected value of a discrete random variable, $\mu_X = E[X] = \sum_x P(x)x$
- We will also symbolize the expected value of a continuous random variable, $\mu_X = E[X]$
- An expression for $E[X]$ can be developed, but this requires calculus

Expected Values

- Define the variance of a continuous random variable,
  $\sigma_X{}^2 = Var(X) = E[(X - E[X])^2] = E[(X - \mu_X)^2]$
- The population standard deviation is defined by
  $\sigma_X = \sqrt{\sigma_X{}^2}$
- As before, if $Y = a + bX$ we have
  - $E[Y] = a + bE[X]$
  - $Var(Y) = b^2 * Var(X)$
  - $StanDev(Y) = b * StanDev(X)$

Expected Values

- Examples:
  - Suppose that we know that the average heights of students in PSC 200 is 5.5 feet with a standard deviation of 0.25 feet. What are the mean, variance, and standard deviation in inches?

Expected Values

- Answer:
  - Let X denote the height of a student in inches and let Y denote the height of the same student in feet. We can write, $Y = 12X$. We have,

$$E[Y] = bE[X] = 12 * 5.5 = 66$$

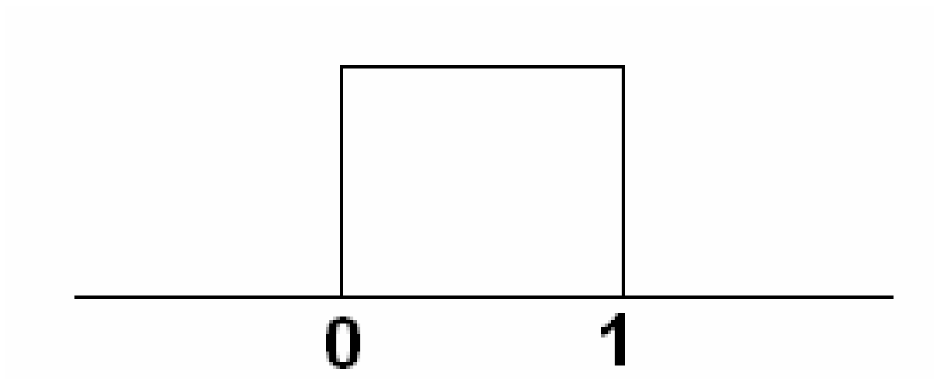
$$Var(Y) = b^2 Var(X) = 144 * 0.25^2 = 9$$

$$StanDev(Y) = bStanDev(X) = 12 * 0.25 = 3$$

  - Hence, the mean height of students is 66 inches, with a variance of 9 inches and a standard deviation of 3 inches

Some Continuous Distributions

- The Uniform(0,1) Distribution:
  - A random variable that is equally likely to take on any value between 0 and 1

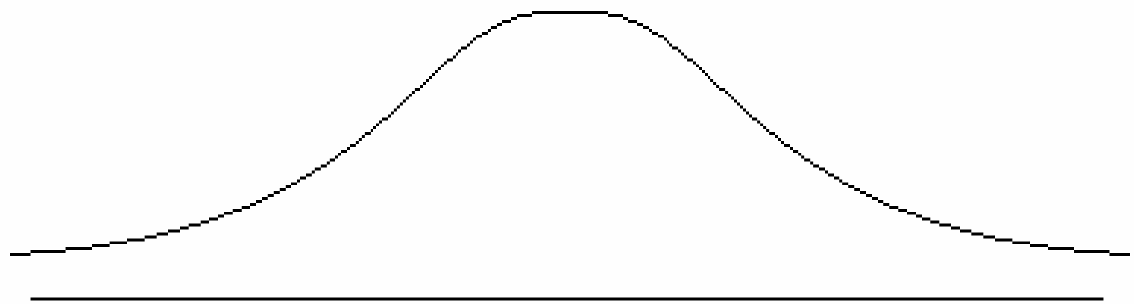


Some Continuous Distributions

- Question:
  - What is the probability that a uniform(0,1) random variable is between 0.45 and 0.7?
  - What is the probability that a uniform(0,1) random variables is greater than 0.56?

Some Continuous Distributions

- Normal Distribution:
  - The most important distribution in statistics
  - Characterized by two parameters, the mean (μ) and the standard deviation (σ)
  - Bell shaped, symmetric, peak centered at its mean, about 68% of area within one standard deviation of mean, about 95% within two standard deviations



Some Continuous Distributions

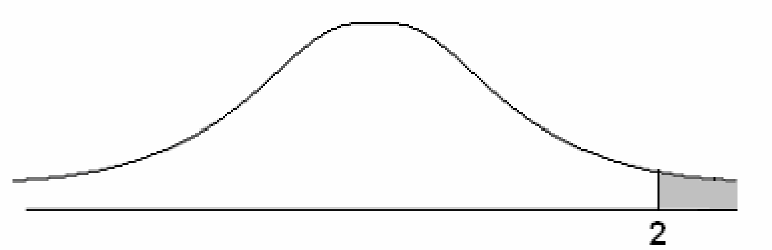
- If $X \sim N(\mu, \sigma^2)$, then $Z = \dfrac{X - \mu}{\sigma} \sim N(0,1)$
- $\Pr(X \leq x) = \Pr\left(\dfrac{X - \mu}{\sigma} \leq \dfrac{x - \mu}{\sigma} \right) = \Pr\left(Z \leq \dfrac{x - \mu}{\sigma} \right)$
- We can compute the probability that the probability above using a computer or table A in the textbook

Some Continuous Distributions

- Question:
  - Suppose that the birth-weight of babies is normally distributed with mean $\mu = 8$ and standard deviation $\sigma = 2$. What is the probability that a given baby weighs more than 12 pounds?

Some Continuous Distributions

- We have $X \sim N(8, 2^2)$
- We would like to compute $\Pr(X \geq 12)$
- We have $\Pr(X \geq 12) = \Pr(\frac{X-8}{2} \geq \frac{12-8}{2}) = \Pr(Z \geq 2)$



- From Table A, $\Pr(Z \geq 2) = 0.0228$

Some Continuous Distributions

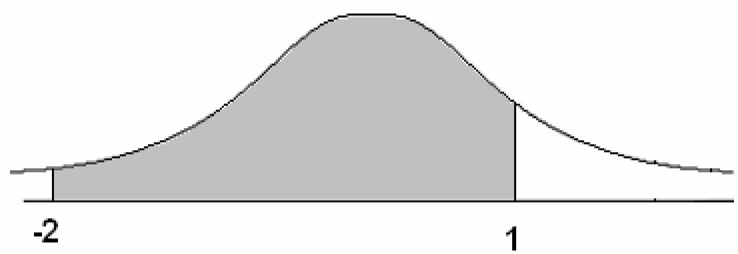
- Question:
  - What is the probability that bodyweight is between 4 and 10?

Some Continuous Distributions

- Once again, $X \sim N(8, 2^2)$
- We would like to compute $\Pr(4 \leq X \leq 10)$
- We have,

$$\Pr(4 \leq X \leq 10) = \Pr(\frac{4-8}{2} \leq \frac{X-8}{2} \leq \frac{10-8}{2})$$

$$= \Pr(-2 \leq Z \leq 1) = \Pr(Z \leq 1) - \Pr(Z \leq -2)$$



Some Continuous Distributions

- From Table A:
  - $\Pr(Z \leq 1) = 1 - \Pr(Z \geq 1) = 1 - 0.1587 = 0.8413$
  - $\Pr(Z \leq -2) = \Pr(Z \geq 2) = 0.0228$
  - $\Pr(-2 \leq Z \leq 1) = \Pr(Z \leq 1) - \Pr(Z \leq -2) = 0.8413 - 0.0228$
    $= 0.8185$

Some Continuous Distributions

- One More Example:
  - Suppose that grades on the midterm are normally distributed with a mean of 85 and a standard deviation of 10, and suppose I assign all grades above 90 an A. What is the probability that a random student will get an A on the midterm?
  - What is the probability that a random student scores between 70 and 95?

Some Continuous Distributions

- One More Example:
  - Let X be a random variable that indicates the grade of a student. The probability that a student receive a grade greater than 90 is given by,

$$\Pr(X \geq 90) = \Pr\left(\frac{X-\mu}{\sigma} \geq \frac{90-85}{10}\right) = \Pr\left(Z \geq 0.5\right)$$

  - Hence, we need to calculate the probability that a standard normal random variable takes on a value greater than 0.5

Some Continuous Distributions

- We can calculate this directly from Table A:

$$\Pr(Z \geq 0.5) = 0.3085$$

- If we want to calculate this using Excel, we have,

$$\Pr(Z \geq 0.5) = 1 - \Pr(Z \leq 0.5)$$

- We use “=1-NORMSDIST(0.5)” to obtain the same answer of 0.3085.

Some Continuous Distributions

- Why the Normal?
  - Normal random variables are mathematically convenient, i.e. sums of normal random variables are still normally distributed
  - Many variables of interest are approximately normal (for reasons we will learn later)
  - Sums of random variables converge to the normal distribution regardless of whether the original variables are normally distributed!
  - This fact provides the basis for much of statistics

What You Should Be Reading

- Agresti and Finlay, Chapters 4, Section 2