4.7 The Beta Probability Distribution

- Some r.v.'s have support on the interval \([0, 1]\).

Examples:
- The proportion of a chemical product that is pure.
- The proportion of a school's students who pass a standardized exam.
- The proportion of a hospital's patients infected with a certain virus.
- The beta distribution is a flexible model for such r.v.'s.

Defn: A r.v. \(Y\) has a beta distribution [Shorthand: \(Y \sim \text{Beta}(\alpha, \beta)\)] if its pdf is:

- Note: The beta pdf has support on \([0, 1]\). If a r.v. \(Y\) of interest has support on some other interval \([c, d]\) of finite length, then we could model the transformed r.v. \(Y^* = \frac{Y - c}{d - c}\) with a beta distribution.
- The beta pdf is very flexible, changing shape dramatically for various $\alpha$ and $\beta$ values:

  * Graph of Beta (2,2) pdf
  * Graph of Beta (0.5,0.5) pdf
  * Graph of Beta (2,3) pdf
  * Graph of Beta (3,2) pdf
  * Graph of Beta (2,1) pdf
  * Graph of Beta (1,1) pdf
- If $\alpha = \beta$, the Beta($\alpha, \beta$) pdf is _______.
- If $\alpha < \beta$, the Beta($\alpha, \beta$) pdf is _______.
- If $\alpha > \beta$, the Beta($\alpha, \beta$) pdf is _______.
- The Beta(1,1) distribution is simply a _______ distribution.

Note: The beta mgf exists, but not in closed form. The moments can be found directly.

Theorem (Beta mean and variance):
If $Y \sim$ Beta($\alpha, \beta$), then

Proof:
\[ E(y^2) = \]

\[ V(y) = \]
Finding Beta Probabilities

- When $\alpha$ and $\beta$ are integers, beta probabilities can be found via direct integration or via the formula (established using repeated integration by parts):

which is a sum of probabilities.

- If $\alpha$ and $\beta$ are not both integers, we can use R to find beta probabilities (see examples on course web page).

**Example 1:** Define the downtime rate as the proportion of time a machine is under repair. Suppose a factory produces machines whose downtime rate follows a $\text{beta}(3,18)$ distribution.

- What is the pdf of the downtime rate $y$?
- For a randomly selected machine, what is the expected downtime rate?

- What is the probability that a randomly selected machine is under repair less than 5% of the time?

- If machines act independently, in a shipment of 25 machines, what is the probability that at least 3 have downtime rates greater than 0.20?
4.10 Tchebysheff's Theorem
- The spread of a probability distribution may be characterized by either its variance or its standard deviation.
- An advantage of using the standard deviation is that (like μ) σ is measured in the same units as the r.v. Y.
- This allows us to make certain probability statements involving μ and σ.

Lemma (Markov's Inequality): Let W be a nonnegative r.v. with pdf (or pmf) f(w), and let c > 0 be a constant. Then

Proof: - We will prove this in the continuous case. Consider the set
- The proof is similar in the discrete case.

**Theorem (Tchebycheff’s Inequality):**
Let \( Y \) be a r.v. (continuous or discrete) with mean \( \mu \) and variance \( \sigma^2 < \infty \). For any \( k > 0 \):

**Proof:**

- Thus, for any r.v. \( Y \), the probability that \( Y \) falls within \( k \) standard deviations of its mean is at least
Example 1: Suppose house prices in an area follow a distribution with mean $93,000 and standard deviation $20,000. What is a lower bound for the probability that a randomly selected price is between $43,000 and $143,000?

Example 2: Tchebycheff’s theorem guarantees that any r.v. with mean $\mu$ and standard deviation $\sigma$ will have probability at least of falling between $\mu - 2\sigma$ and $\mu + 2\sigma$.

- For distributions with “mound-shaped” densities, such a probability may be substantially larger than this lower bound:
- In Section 4.6, we found that if $Y \sim \text{Gamma}(5,3)$, then $Y$ had probability 0.9588 of falling within 2 standard deviations of its mean.
- If $Y \sim \text{Pois}(6.25)$, then