Studio 3
18.05 Spring 2014
Jeremy Orloff and Jonathan Bloom

**Frequency and Density Histograms**

- **Frequency Chart**
  - x-axis: 0.5, 1.5, 2.5, 3.5, 4.5
  - y-axis: 1, 2, 3

- **Density Chart**
  - x-axis: 0.5, 1.5, 2.5, 3.5, 4.5
  - y-axis: 0.1, 0.2, 0.3, 0.4
Suppose $X$ is a continuous random variable.

a) What is $P(a \leq X \leq a)$?

b) What is $P(X = 0)$?

c) Does $P(X = 2) = 0$ mean $X$ never equals 2?

**answer:**

a) 0

b) 0

b) 0

For a continuous distribution any single value has probability 0. Only a range of values has non-zero probability.
Concept question

Which of the following are graphs of valid cumulative distribution functions?

![Graphs of cumulative distribution functions]

Add the numbers of the valid cdf’s and click that number.

answer: Test 2 and Test 3.
Test 1 is not a cdf: it takes negative values, but probabilities are positive.
Test 2 is a cdf: it increases from 0 to 1.
Test 3 is a cdf: it increases from 0 to 1.
Test 4 is not a cdf: it decreases. A cdf must be non-decreasing since it represents \textit{accumulated} probability.
Exponential Random Variables

Parameter: $\lambda$ (called the rate parameter).

Range: $[0, \infty)$.

Notation: exponential($\lambda$) or exp($\lambda$).

Density: $f(x) = \lambda e^{-\lambda x}$ for $0 \leq x$.

Models: Waiting time

Continuous analogue of geometric distribution – memoryless!
Uniform and Normal Random Variables

**Uniform:** $U(a, b)$ or uniform$(a, b)$
Range: $[a, b]$
PDF: $f(x) = \frac{1}{b - a}$

**Normal:** $N(\mu, \sigma^2)$
Range: $(-\infty, \infty]$ 
PDF: $f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$

http://mathlets.org/mathlets/probability-distributions/
Table questions

Open the applet

http://mathlets.org/mathlets/probability-distributions/

1. For the **standard normal** distribution $N(0, 1)$ how much probability is within 1 of the mean? Within 2? Within 3?

2. For $N(0, 3^2)$ how much probability is within $\sigma$ of the mean? Within $2\sigma$? Within $3\sigma$.

3. Does changing $\mu$ change your answer to problem 2?
Normal probabilities

Rules of thumb:

\[ P(-1 \leq Z \leq 1) \approx 0.68, \]
\[ P(-2 \leq Z \leq 2) \approx 0.95, \]
\[ P(-3 \leq Z \leq 3) \approx 0.997 \]
Download R script

Download studio3.zip and unzip it into your 18.05 working directory.

Open studio3.r in RStudio.
Histograms

Will discuss in more detail in class 6.

Made by ‘binning’ data.

**Frequency**: height of bar over bin = \# of data points in bin.

**Density**: area of bar over bin is proportional to \# of data points in bin. Total area of a density histogram is 1.
Histograms of averages of $\exp(1)$

1. Generate a frequency histogram of 1000 samples from an $\exp(1)$ random variable.

2. Generate a density histogram for the average of 2 independent $\exp(1)$ random variable.

3. Using `rexp()`, `matrix()` and `colMeans()` generate a density histogram for the average of 50 independent $\exp(1)$ random variables. Make 10000 sample averages and use a binwidth of .1 for this. Look at the spread of the histogram.

4. Superimpose a graph of the pdf of $N(1, 1/50)$ on your plot in problem 3. (Remember the second parameter in $N$ is $\sigma^2$.)

Code for the solutions is at

18.05 Introduction to Probability and Statistics
Spring 2014

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.