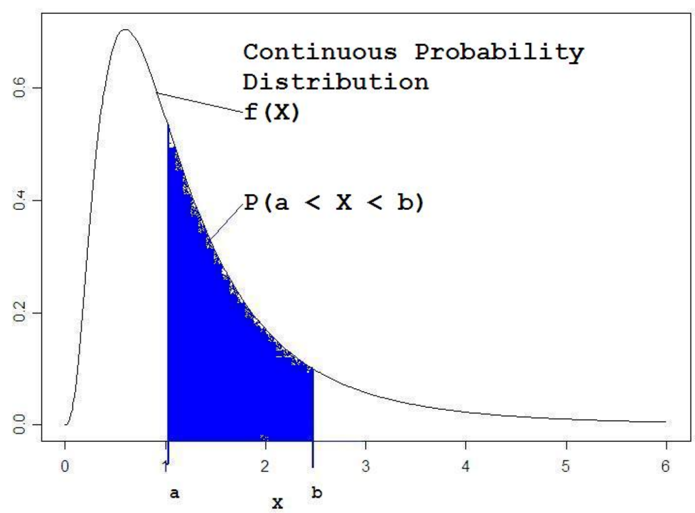
Continuous Random Variables I (4.1 - 4.3)

1. **Continuous Random Variables:** Random variable X is continuous if its set of all possible values is an entire interval of numbers. For $A < B$, any real number x between A and B is a possible value.

   (a) **Properties of Continuous Random Variables:** Important properties of random variables are their probability distributions, mean, and variance or their shape, center, and spread.

   (b) **Probability Density Function(PDF):** The probability distribution or probability density function (pdf) for continuous random variable X is a function $f(x)$ such that for any two numbers a and b with $a \leq b$, the probability that $a \leq X \leq b$ is equal to the area under $f(x)$ between a and b.

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



   - For $f(x)$ to be a legitimate pdf: $f(x) > 0$ for all values of x.
   - The total area under the density function curve from $-\infty$ to $+\infty$ is equal to 1.
   - For any number c, $P(x = c) = 0$
   - For any two numbers a and b with $a < b$,

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$$

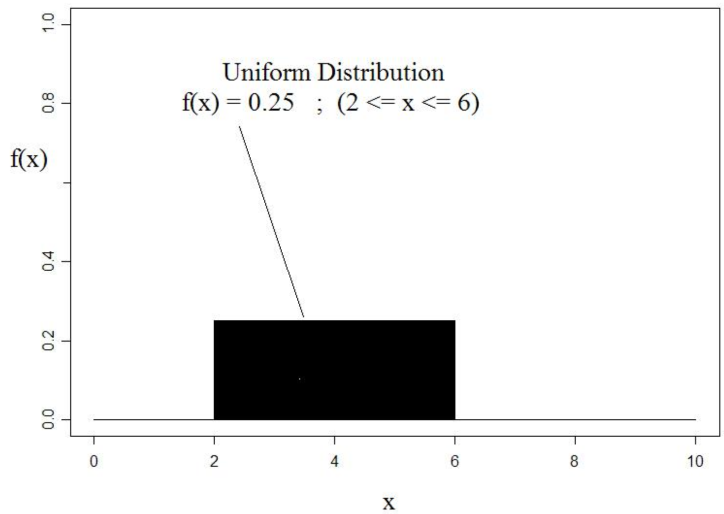
   - The probability of continuous rv X being on an interval does not depend on whether interval end points are included.

   (c) **Uniform Distribution:** A continuous random variable X is said to have a uniform distribution on the interval $[A, B]$ if the probability density function (pdf) of X is:

$$f(x; A, B) = \begin{cases} \frac{1}{B-A} & A \leq x \leq B \\ 0 & Otherwise \end{cases}$$

   - All values of X on interval $[A, B]$ are equally likely.

- As a pdf, the total area under $f(x;A,B)$ between A and B must equal 1 (area in black on figure below).



(d) Example: Suppose the reaction temperature, X (in ^oC), for a chemical process has a uniform distribution with $A = -5$ and $B = 5$.

a.) Compute $P(X < 0)$.

- $f(x;A,B) = \frac{1}{B-A} = \frac{1}{5-(-5)} = 1/10$
- $P(X < 0) = \int_{-\infty}^{0} f(x)dx = \int_{-5}^{0} \frac{1}{10}dx = \frac{1}{10}\left(x\right]_{-5}^{0} = 0 - (\frac{-5}{10}) = 0.5$

b.) Compute $P(-2.5 < X < 2.5)$.

- $P(-2.5 < X < 2.5) = \int_{-2.5}^{2.5} \frac{1}{10}dx = \frac{1}{10}\left(x\right]_{-2.5}^{2}.5 = \frac{1}{10}(2.5 - (-2.5)) = 0.5$

c.) Compute $P(-2 \leq X \leq 3)$.

- $P(-2 \leq X \leq 3) = \int_{-2}^{3} \frac{1}{10}dx = \frac{1}{10}\left(x\right]_{-2}^{3} = \frac{1}{10}(3 - (-2)) = 0.5$

d.) Compute $P(k < X < (k+4))$.

- $P(k < X < k+4) = \int_{k}^{k+4} \frac{1}{10}dx = \frac{1}{10}\left(x\right]_{k}^{k+4} = \frac{1}{10}((k+4) - k) = 0.4$

(e) Example: Let X denote the vibratory stress (psi), on the blade of a wind turbine rotating at constant speed in a wind tunnel. If X follows the Rayleigh distribution with pdf given below, answer the following:

$$f(x;\theta) = \begin{cases} \frac{x}{\theta^2} \cdot e^{\frac{-x^2}{2\theta^2}} & x > 0 \\ 0 & Otherwise \end{cases}$$

a.) Verify that $f(x;\theta)$ is a legitimate pdf.

- Must show that $\int_{-\infty}^{\infty} f(x)dx = 1$ and that $f(x) \geq 0$
- Note the general form of this integral:

$$\int e^{a \cdot u} du = \frac{e^{a \cdot u}}{a}$$

- let: $a = -1$ and $u = \frac{x^2}{2\theta^2}$, then $du = \frac{2x}{2\theta^2}dx = \frac{x}{\theta^2}dx$, $u = 0$ when $x = 0$, and $u = \infty$ when $x = \infty$

$$\int_{-\infty}^{\infty} f(x;\theta)dx = \int_{0}^{\infty} \frac{x}{\theta^2} \cdot e^{\frac{-x^2}{2\theta^2}} dx$$

$$\int_{0}^{\infty} e^{\frac{-x^2}{2\theta^2}} \cdot \frac{x}{\theta^2} dx = \int_{0}^{\infty} e^{a \cdot u} du = \frac{e^{a \cdot u}}{a}]_0^\infty = \frac{e^{-u}}{-1}]_0^\infty = \left[-e^{-\infty} - (-1)\right] = 0-(-1) = 1$$

- since $e^{-\infty} = 0$ and $e^0 = 1$

b.) For $\theta = 200$, what is the probability that X is at most 200?, Less than 200?, At least 200?

$$P(X \leq 200) = \int_{-\infty}^{200} f(x;\theta)dx = \int_{0}^{200} \frac{x}{\theta^2} \cdot e^{\frac{-x^2}{2\theta^2}} dx = -e^{\frac{-x^2}{2\theta^2}}]_0^{200} \approx -0.1353 + 1 = 0.8647$$

- Since X is continuous, $P(X < 200) = P(X \leq 200) \approx 0.8647$
- $P(X \geq 200) = 1 - P(X \leq 200) \approx 0.1353$

c.) Again, for $\theta = 200$, what is the probability that X is between 100 and 200?

$$P(100 \leq X \leq 200) = \int_{100}^{200} f(x;\theta)dx = \int_{100}^{200} \frac{x}{\theta^2} \cdot e^{\frac{-x^2}{2\theta^2}} dx = -e^{\frac{-x^2}{2\theta^2}}]_{100}^{200} \approx 0.4712$$

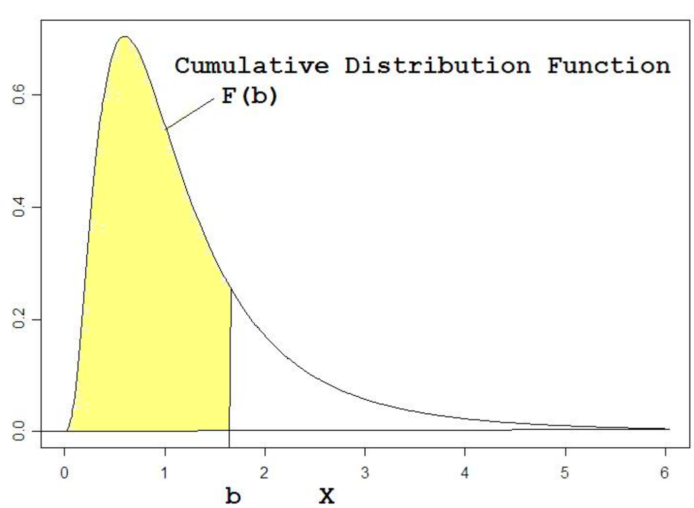
d.) Give an expression for $P(X \leq x)$.

for $x > 0$

$$P(X \leq x) = \int_{-\infty}^{x} f(y;\theta)dy = \int_{0}^{x} \frac{y}{\theta^2} \cdot e^{\frac{-y^2}{2\theta^2}} dy = -e^{\frac{-y^2}{2\theta^2}}]_0^x = 1 - e^{\frac{-x^2}{2\theta^2}}$$

2. **Cumulative Distribution Function (CDF):** The cumulative distribution function(cdf), $F(x)$, for continuous random variable X, is defined for every number x by,

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$$



(a) The CDF for continuous random variable X, $F(x)$, is the area under the density function, $f(x)$, to the left of x.

- $F(a) = P(X \leq a) = \int_{-\infty}^{a} f(y)dy$
- $P(X \geq a) = 1 - F(a) = \int_{a}^{+\infty} f(y)dy$
- $1 - F(a)$ is area under the density function to the right of a.
- $P(a \leq X \leq b) = F(b) - F(a) = \int_{b}^{a} f(y)dy$
- If continuous rv X has pdf $f(x)$ and cdf $F(x)$, then at every number x that derivitive $\frac{dF(x)}{dx}$ exists, $\frac{dF(x)}{dx} = f(x)$.

(b) Percentiles of a Continuous Distribution

- Let continuous rv $X = x_0$, with a cdf value $F(x_0) = 0.65$, (area under the pdf to the left of x_0), then x_0 is the 65th percentile value of X.
- Example: If your test score was at the 85th percentile of the population:
  - 85% of all population scores were below your score.
  - 15% of all population scores were above your score.
- The **median** of continuous rv X, is it's 50th percentile value.
- If $F(x_0) = 0.50$, then x_0 is the median value of X.

(c) Example: The cdf for continuous random variable X is given as:

$$F(x) = \begin{cases} 0 & x < -2 \\ \frac{1}{2} + \frac{3}{32}\left(4x - \frac{x^3}{3}\right) & -2 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

a.) Compute $P(X < 0)$.

- $P(X < 0) = P(X \leq 0) = F(0) = \frac{1}{2} + \frac{3}{32}(4 \cdot 0 - \frac{0^3}{3}) = \frac{1}{2} = 0.5$

b.) Compute $P(-1 < X < 0)$.

- $P(-1 < X < 1) = F(1) - F(-1) = (\frac{1}{2} + \frac{11}{32}) - (\frac{1}{2}) - \frac{11}{32}) = \frac{22}{32} = 0.6875$

c.) Compute $P(.5 < X)$.

- $P(X > 0.5) = 1 - P(X \leq 0.5) = 1 - F(0.5) = 1 - (0.5 + 0.1836) = 0.3164$

d.) Verify that:

$$f(x) = \begin{cases} 0.9375(4 - x^2) & -2 \leq x \leq 2 \\ 0 & Otherwise \end{cases}$$

- Since $F'(x) = f(x)$, differentiate $F(x)$ with respect to x and compare.
- $\frac{dF(x)}{dx} = \frac{d}{dx}\left(\frac{1}{2} + \frac{3}{32}\left(4x - \frac{x^3}{3}\right)\right) = 0 + \frac{3}{32}\left(4 - \frac{3x^2}{3}\right) = 0.09375(4 - x^2) = f(x)$

e.) Verify that the median of X equals 0.

- $X = 0$ is the median only if $F(0) = 0.5$, but this was shown to be true above in part a.

3. **Expected Value (Mean):** The mean or expected value, $E[X]$, of continuous random variable X is a measure of the center (or location) of it's distribution and calculated as

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

(a) If $h(x)$ is any function of continuous random variable X with pdf $f(x)$, then

$$E[h(x)] = \mu_{h(x)} = \int_{-\infty}^{+\infty} h(x) \cdot f(x) dx$$

(b) When $h(x) = aX + b$, where a and b are constants, then

$$E[h(x)] = E[aX + b] = aE[X] + b$$

(c) **Variance:** The variance, $Var[X]$, of continuous random variable X is a measure of spread of X about its mean, μ_X, or expected value.

$$Var(X) = \sigma_X^2 = \int_{-\infty}^{+\infty} (x - \mu_X)^2 \cdot f(x) dx = E[(X - \mu_X)^2]$$

(d) Short-cut formula for calculating variance is:

$$Var[X] = E[X^2] - (E[X])^2$$

(e) The standard deviation, $SD[X]$, of continuous random variable X is the positive square root of its variance, and has the same units as X.

$$SD[X] = \sigma_X = \sqrt{Var[X]}$$

(f) When $h(x) = aX + b$, where a and b are constants, then

$$Var[h(x)] = Var[aX + b] = a^2 Var[X]$$

$$SD[h(x)] = \sqrt{Var[h(x)]} = |a| \cdot SD[X]$$

(g) Example: Let continuous rv X denote weekly gravel sales (tons of gravel sold per week) by a construction supply company. The pdf of X is given below:

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \le x \le 1 \\ 0 & Otherwise \end{cases}$$

a.) What is the expected value, $E(X)$, of weekly gravel sales?

- $E[X] = \mu_X = \int_{-\infty}^{+\infty} x \cdot f(x) dx = \int_0^1 x \cdot \frac{3}{2}(1 - x^2) dx = \frac{3}{2} \int_0^1 (x - x^3) dx$
- Note that: $\frac{d}{dx}(\frac{1}{2}x^2 - \frac{1}{4}x^4) = (x - x^3)$
- so that $E[X] = \frac{3}{2} \int_0^1 (x - x^3) dx = \frac{3}{2} \left(\frac{x^2}{2} - \frac{x^4}{4} \right) |_{x=0}^{x=1} = \frac{3}{8} = 0.375$

b.) What is the variance, $Var(X)$, and standard deviation, $SD(X)$, of weekly gravel sales?

- $Var[X] = E[X^2] - (E[X])^2$
- $E[X^2] = \int_{-\infty}^{+\infty} x^2 \cdot f(x)dx = \int_0^1 x^2 \cdot \frac{3}{2}(1 - x^2)dx = \frac{3}{2}\int_0^1 (x^2 - x^4)dx$
- Note that: $\frac{d}{dx}(\frac{1}{3}x^3 - \frac{1}{5}x^5) = (x^2 - x^4)$
- so that $E[X^2] = \frac{3}{2}\int_0^1 (x^2 - x^4)dx = \frac{3}{2}\left(\frac{x^3}{3} - \frac{x^5}{5}\right)|_{x=0}^{x=1} = \frac{1}{5} = 0.200$
- $Var[X] = \frac{1}{5} - (\frac{3}{8})^2 = 19/320 = 0.059$
- $SD[X] = \sigma_X = \sqrt{Var[X]} = \sqrt{0.059} = 0.244$

4. **Normally Distributed Random Variables:** A continuous random variable, X, is said to have a normal distribution with parameters μ and σ, when its distribution function has the following form:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

here $-\infty < x < +\infty$, $-\infty < \mu < +\infty$, and $\sigma > 0$

(a) When X is a normally distributed random variable with mean,μ, and variance, σ^2, it is said to be distributed as $N(\mu, \sigma^2)$.

- For $X \sim N(\mu, \sigma^2)$, $E(X) = \mu$ and $Var(X) = \sigma^2$
- $N(\mu, \sigma^2)$ is symmetric about its mean value μ.
- As a pdf, the total area under $N(\mu, \sigma^2)$ is equal to 1.

(b) **Standard Normal Distribution:** A special case of the normal distribution where $\mu = 0$ and $\sigma^2 = 1$, it is written as $N(0, 1)$.

- Let X be a normally distributed random variable with mean,μ, and variance, σ^2, then

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

- any normally distributed random variable, X, can be converted to its equivalent Z form by subtracting its mean and dividing by it's standard deviation.
- Z is called the standard normal random variable, which ranges in value from $(-\infty < z < \infty)$; its pdf is given as follows:

$$f(z; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

- The CDF for Z, the $P(Z \leq z)$ is denoted $\Phi(z)$ and is the area under $f(z; 0, 1)$ to the left of z.
- These CDF areas are tabulated for varied values of z, so we don't have to integrate to determine probabilities. See Table A.3 of the text (p. 740).
- Normal probabilities are typically calculated, by transforming a problem in X to it's equivalent problem in Z and then using the table for Standard Normal Curve Areas.

(c) Example: If X is a normally distributed random variable with mean 80 and a standard deviation of 10, compute the following probabilities using the table of Standard Normal Curve Areas.

i. $P(X \leq 70) = P(Z \leq \frac{70-80}{10}) = P(Z \leq -1.0) = \Phi(-1.0) =$

ii. $P(65 \leq X \leq 78) = P(\frac{65-80}{10} \leq Z \leq \frac{78-80}{10}) = P(-1.5 \leq Z \leq -0.2) = \Phi(1.5) - \Phi(0.2) = 0.4332 - 0.0793 = 0.3539$

iii. $P(X \geq 47) = 1 - P(X \leq 47) = 1 - P(Z \leq \frac{47-80}{10}) = 1 - \Phi(-3.3) = 1 - 0.0001 = 1.0$

(d) **Critical Values of Z:** A critical value, z_α, refers to the value of Z such that the area under the standard normal curve to the right of z_α equals the value α, or stated as a probability:

$$P(Z \geq z_\alpha) = \alpha$$

(e) **Normal Approximation for Binomial**

i. Binomial random variable, X, has the following mean and variance:

$$\mu_X = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{(n-x)} = np$$

$$\sigma_X^2 = \sum_{x=0}^{n} (x - np)^2 \binom{n}{x} p^x (1-p)^{(n-x)} = np(1-p)$$

ii. When both np and $n(1-p)$ are greater than 10, binomial random variable X is approximately normally distributed with:

$$X \sim N\left(np, np(1-p)\right)$$

iii. Example: 35% of drivers fail to come to a complete stop at a stop sign before proceeding. If 50 drivers are watched at an intersection, what is the probability that fewer than 20 come to a complete stop?

- X is $b(x; 50, .35)$
- $np = 17.5$ and $n(1-p) = 32.5$ are both greater than 10
- $np(1-p) = 11.375$
- X is approximately $N(17.5, 11.375)$

- **Continuity Correction:** The error in approximating a binomial (discrete) probability using the normal distribution (continuous) is greatly reduced by use of a *continuity correction*.
- $P(X = x)$ is approximated by the normal probability between $x - .5$ and $x + .5$
- so $P(X < 20) = P(X < 19.5) = P(Z < \frac{19.5-17.5}{\sqrt{11.375}}) = P(Z < 0.593) = 0.7224$
- $P(X \leq 20) = P(X < 20.5) = P(Z < \frac{20.5-17.5}{\sqrt{11.375}}) = P(Z < 0.889) = 0.8133$
- $P(15 \leq X \leq 20) = P(\frac{14.5-17.5}{\sqrt{11.375}} < Z < \frac{20.5-17.5}{\sqrt{11.375}}) = P(-0.889 < Z < 0.889) = 0.6265$