Section 7.1: Continuous Random Variables
A random variable $X$ is *continuous* if and only if its set of possible values $\mathcal{X}$ is a *continuum*.

A continuous random variable $X$ is uniquely determined by

- Its set of possible values $\mathcal{X}$
- Its *probability density function* (pdf):
  A real-valued function $f(\cdot)$ defined for each $x \in \mathcal{X}$
  
  $$
  \int_{a}^{b} f(x) \, dx = \Pr(a \leq X \leq b)
  $$

  By definition,
  
  $$
  \int_{\mathcal{X}} f(x) \, dx = 1
  $$
Example 7.1.1

- $X$ is $\text{Uniform}(a, b)$
  - $\mathcal{X} = (a, b)$ and all values in this interval are equally likely
  
  $f(x) = \frac{1}{b - a}$ \hspace{1cm} a < x < b

- In the continuous case,
  - $\Pr(X = x) = 0$ for any $x \in \mathcal{X}$
  - If $[a, b] \subseteq \mathcal{X}$,

  $\int_{a}^{b} f(x)\,dx = \Pr(a \leq X \leq b) = \Pr(a < X \leq b) = \Pr(a \leq X < b) = \Pr(a < X < b)$
The cumulative distribution function (cdf) of the continuous random variable $X$ is the real-valued function $F(\cdot)$ for each $x \in \mathcal{X}$ as

$$F(x) = \Pr(X \leq x) = \int_{t \leq x} f(t)dt$$

**Example 7.1.2:** If $X$ is $\text{Uniform}(a, b)$, the cdf is

$$F(x) = \int_{t=a}^{x} \frac{1}{b - a} dt = \frac{x - a}{b - a} \quad a < x < b$$

In special case where $U$ is $\text{Uniform}(0, 1)$, the cdf is

$$F(u) = \Pr(U \leq u) = u \quad 0 \leq u \leq 1$$
Continuous Random Variables

Relationship between pdfs and cdfs

- Shaded area in pdf graph equals $F(x_0)$
More on cdfs

- The cdf is strictly monotone increasing:
  if \( x_1 < x_2 \), then \( F(x_1) < F(x_2) \)
- The cdf is bounded between 0.0 and 1.0
- The cdf can be obtained from the pdf by integration
  The pdf can be obtained from the cdf by differentiation as
  \[
  f(x) = \frac{d}{dx} F(x) \quad x \in \mathcal{X}
  \]
- A continuous random variable model can be specified by \( \mathcal{X} \) and either the pdf or the cdf
Example 7.1.3: Exponential($\mu$)

- $X = -\mu \ln(1 - U)$ where $U$ is Uniform(0, 1)
- The cdf of $X$ is

$$F(x) = \Pr(X \leq x) = \Pr(-\mu \ln(1 - U) \leq x)$$
$$= \Pr(1 - U \geq \exp(-x/\mu))$$
$$= \Pr(U \leq 1 - \exp(-x/\mu))$$
$$= 1 - \exp(-x/\mu)$$

- The pdf of $X$ is

$$f(x) = \frac{d}{dx} F(x) = \frac{d}{dx} (1 - \exp(-x/\mu)) = \frac{1}{\mu} \exp(-x/\mu) \quad x > 0$$
The mean $\mu$ of the continuous random variable $X$ is

$$\mu = \int x f(x) \, dx$$

The corresponding standard deviation $\sigma$ is

$$\sigma = \sqrt{\int (x - \mu)^2 f(x) \, dx} \quad \text{or} \quad \sigma = \sqrt{\left(\int x^2 f(x) \, dx\right) - \mu^2}$$

The variance is $\sigma^2$
Examples

- If $X$ is $\text{Uniform}(a, b)$
  \[ \mu = \frac{a + b}{2} \quad \text{and} \quad \sigma = \frac{b - a}{\sqrt{12}} \]

- If $X$ is $\text{Exponential}(\mu)$,
  \[ \int x f(x) \, dx = \int_0^\infty \frac{x}{\mu} \exp(-x/\mu) \, dx = \mu \int_0^\infty t \exp(-t) \, dt = \cdots = \mu \]
  \[ \sigma^2 = \left( \int_0^\infty \frac{x^2}{\mu} \exp(-x/\mu) \, dx \right) - \mu^2 = \cdots = \mu^2 \]
Expected Value

- The mean of a continuous random variable is also known as the *expected value*
- The expected value of the continuous random variable $X$ is

$$
\mu = E[X] = \int_x xf(x) \, dx
$$

- The variance is the expected value of $(X - \mu)^2$

$$
\sigma^2 = E[(X - \mu)^2] = \int_x (x - \mu)^2 f(x) \, dx
$$

- In general, if $Y = g(X)$, the expected value of $Y$ is

$$
E[Y] = E[g(X)] = \int_x g(x)f(x) \, dx
$$
Example 7.1.6

- A circle of radius $r$ and a fixed point $Q$ on the circumference
- $P$ is selected at random on the circumference
- Let the random variable $Y$ be the distance of the line segment joining $P$ and $Q$

$$Y = 2r \sin\left(\frac{\Theta}{2}\right)$$
Example 7.1.6 ctd.

- If $\Theta$ is $\text{Uniform}(0, 2\pi)$, the pdf of $\Theta$ is $f(\theta) = 1/2\pi$
- The expected length of $Y$ is

$$E[Y] = \int_0^{2\pi} 2r \sin(\theta/2) f(\theta) d\theta = \int_0^{2\pi} \frac{2r \sin(\theta/2)}{2\pi} d\theta = \cdots = \frac{4r}{\pi}$$

- $Y$ is not $\text{Uniform}(0, 2r)$; otherwise, $E[Y]$ would be $r$.

Example 7.1.7

- If continuous random variable $Y = aX + b$ for constants $a$ and $b$,

$$E[Y] = E[aX + b] = aE[X] + b$$
**Standard Normal Random Variable**

$Z$ is $\text{Normal}(0, 1)$ if and only if the set of all possible values is $Z = (-\infty, \infty)$ and the pdf is

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2) \quad -\infty < z < \infty$$
If $Z$ is $\text{Normal}(0, 1)$, $Z$ is “standardized”

The mean is

$$\mu = \int_{-\infty}^{\infty} zf(z)dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z \exp(-z^2/2)dz = \cdots = 0$$

The variance is

$$\sigma^2 = \int_{-\infty}^{\infty} (z-\mu)^2 f(z)dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 \exp(-z^2/2)dz = \cdots = 1$$

The cdf is

$$F(z) = \int_{-\infty}^{z} f(t)dt = \Phi(z) \quad -\infty < z < \infty$$
Standard Normal cdf

- \( \Phi(\cdot) \) is defined as
  
  \[
  \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-t^2/2) \, dt \quad -\infty < z < \infty
  \]

- No closed-form expression for \( \Phi(z) \)
  
  \[
  \Phi(z) = \begin{cases} 
  \frac{1 + P(1/2, z^2/2)}{2} & z \geq 0 \\
  1 - \Phi(z) & z < 0 
  \end{cases}
  \]

  \( P(a, x) \) is an incomplete gamma function (see Appendix D)

- Function \( \Phi(z) \) is available in \texttt{rvms} as \texttt{cdfNormal(0.0, 1.0, z)}
Suppose $X$ is a random variable with mean $\mu$ and standard deviation $\sigma$.

Define random variable $X' = aX + b$ for constants $a, b$.

The mean $\mu'$ and standard deviation $\sigma'$ of $X'$ are

$$
\mu' = E[X'] = E[aX + b] = aE[X] + b = a\mu + b
$$

$$(\sigma')^2 = E[(X' - \mu')^2] = E[(aX - a\mu)^2] = a^2 E[(X - \mu)^2] = a^2 \sigma^2
$$

Therefore,

$$
\mu' = a\mu + b \quad \text{and} \quad \sigma' = |a|\sigma
$$
Example 7.1.8

- Suppose $Z$ is a random variable with mean 0 and standard deviation 1.
- Construct a new random variable $X$ with specified mean $\mu$ and standard deviation $\sigma$.
- Define $X = \sigma Z + \mu$.
- $E[X] = \sigma E[Z] + \mu = \mu$.
- $E[(X - \mu)^2] = E[\sigma^2 Z^2] = \sigma^2 E[Z^2] = \sigma^2$. 
The continuous random variable $X$ is $\text{Normal}(\mu, \sigma)$ if and only if

$$X = \sigma Z + \mu$$

where $\sigma > 0$ and $Z$ is $\text{Normal}(0, 1)$.

- The mean of $X$ is $\mu$ and the standard deviation is $\sigma$.
- $\text{Normal}(\mu, \sigma)$ is constructed from $\text{Normal}(0, 1)$ by “shifting” the mean from 0 to $\mu$ via the addition of $\mu$,
  - by “scaling” the standard deviation from 1 to $\sigma$ via multiplication by $\sigma$. 
The cdf of a $\text{Normal}(\mu, \sigma)$

$$F(x) = \Pr(X \leq x) = \Pr(\sigma Z + \mu \leq x) = \Pr(Z \leq (x - \mu)/\sigma)$$

so that

$$F(x) = \Phi \left( \frac{x - \mu}{\sigma} \right) \quad -\infty < x < \infty$$

where $\Phi(\cdot)$ is the cdf of $\text{Normal}(0, 1)$
Because

\[
\frac{d}{dz} \Phi(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2) \quad -\infty < z < \infty
\]

the pdf of Normal\((\mu, \sigma)\) is

\[
f(x) = \frac{dF(x)}{dx} = \frac{d}{dx} \Phi \left( \frac{x - \mu}{\sigma} \right) = \cdots = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\left(\frac{x - \mu}{\sigma}\right)^2/2\right)
\]
Sums of iid random variables approach the normal distribution

*Normal*(μ, σ) is sometimes called a *Gaussian* random variable

The 68-95-99.73 rule

- Area under pdf between μ − σ and μ + σ is about 0.68
- Area under pdf between μ − 2σ and μ + 2σ is about 0.95
- Area under pdf between μ − 3σ and μ + 3σ is about 0.9973

The pdf has inflection points at μ ± σ

Common notation for Normal(μ, σ) is *N*(μ, σ²)

Support is \( \mathcal{X} = \{ x | -\infty \leq x \leq \infty \} \)

Usually not appropriate for simulation unless modified to produce only positive values
The continuous random variable $X$ is $Lognormal(a, b)$ if and only if

$$X = \exp(a + bZ)$$

where $Z$ is $Normal(0, 1)$ and $b > 0$

$Lognormal(a, b)$ is also based on transforming $Normal(0, 1)$

- The transformation is non-linear
The cdf of a $\text{Lognormal}(a, b)$

$$F(x) = \Pr(X \leq x) = \Pr(\exp(a+bZ) \leq x) = \Pr(a+bZ \leq \ln(x))$$

so that

$$F(x) = \Pr(Z \leq (\ln(x) - a)/b) = \Phi\left(\frac{\ln(x) - a}{b}\right) \quad x > 0$$

where $\Phi(\cdot)$ is the cdf of $\text{Normal}(0, 1)$
The pdf of Lognormal \((a, b)\) is

\[
f(x) = \frac{dF(x)}{dx} = \cdots = \frac{1}{bx\sqrt{2\pi}} \exp\left(-\left(\ln(x) - a\right)^2 / 2b^2\right) \quad x > 0
\]

\((a, b) = (-0.5, 1.0)\)

\[
\mu = \exp(a + b^2/2) \quad \text{Above, } \mu = 1.0
\]

\[
\sigma = \exp(a + b^2/2) \sqrt{\exp(b^2) - 1} \quad \text{Above, } \sigma \approx 1.31
\]
Continuous Random Variables

Erlang Random Variable

- $Uniform(a, b)$ is the continuous analog of $Equilikely(a, b)$
- $Exponential(\mu)$ is the continuous analog of $Geometric(p)$
- $Pascal(n, p)$ is the sum of $n$ iid $Geometric(p)$
- What is the continuous analog of $Pascal(n, p)$?

The continuous random variable $X$ is $Erlang(n, b)$ if and only if

$$X = X_1 + X_2 + \cdots + X_n$$

where $X_1, X_2, \cdots, X_n$ are iid $Exponential(b)$ random variables
pdf of Erlang Random Variable

The pdf of $Erlang(n, b)$ is

$$f(x) = \frac{1}{b(n-1)!} \left(\frac{x}{b}\right)^{n-1} \exp\left(-\frac{x}{b}\right) \quad x > 0$$

$$(n, b) = (3, 1.0)$$

For $(n, b) = (3, 1.0)$, $\mu = 3.0$ and $\sigma \approx 1.732$
The corresponding cdf is

\[ F(x) = \int_0^x f(t) dt = P(n, x/b) \quad x > 0 \]

Incomplete gamma function (see Appendix D)

- \( \mu = nb \)
- \( \sigma = \sqrt{nb} \)

**Chisquare And Student Random Variables**

- \( \text{Chisquare}(n) \) and \( \text{Student}(n) \) are commonly used for statistical inference
- Defined in section 7.2