Chapter 4 - Continuous Random Variables and Probability Distributions
Outline

1. Continuous random variables and probability density functions
2. Cumulative distribution function and expected values
3. The normal distribution
4. The Gamma distribution and its relatives
5. Probability plots
Outline

1. Continuous random variables and probability density functions
2. Cumulative distribution function and expected values
3. The normal distribution
4. The Gamma distribution and its relatives
5. Probability plots
Outline

1. Continuous random variables and probability density functions
2. Cumulative distribution function and expected values
3. The normal distribution
4. The Gamma distribution and its relatives
5. Probability plots
Outline

1. Continuous random variables and probability density functions
2. Cumulative distribution function and expected values
3. The normal distribution
4. The Gamma distribution and its relatives
5. Probability plots
Outline

1. Continuous random variables and probability density functions
2. Cumulative distribution function and expected values
3. The normal distribution
4. The Gamma distribution and its relatives
5. Probability plots
Continuous Random Variables

Definition  A random variable that can (theoretically) assume any value in a finite or infinite interval is said to be *continuous*.

Measurements  Let $X$ be the depth measurement at a randomly chosen locations of a lake. Then $X$ is a continuous random variable.

Time to failure  The result is potentially any positive number.

Round-off error  Round-off error in calculations is generally modeled as a uniform continuous distribution.
Probability density function

Definition A probability density function (pdf) of a continuous random variable $X$ is a function $f(x)$ such that for any two numbers $a \leq b$,

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

That is, the probability that $X$ takes on a value in the interval $[a, b]$ is the area under the graph of the density function. For any number $c$, $P(X = c) = 0$.

Conditions for pdf:

$$f(x) \geq 0 \text{ for all } x$$

$$\int_{-\infty}^{\infty} f(x) \, dx = 1.$$
Uniform distribution

A continuous random variable $X$ is said to have a *uniform distribution* on the interval $[A, B]$ if the pdf of $X$ is

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

Example: Assume the waiting time at a bus stop is uniformly distributed on the interval $[0, 5]$. The probability that it is between 1 and 3 minutes is

$$P(1 \leq X \leq 3) = \int_1^3 f(x)\,dx = \int_1^3 \frac{1}{5} \,dx = \frac{2}{5}.$$ 

> `punif(3, min = 0, max = 5) - punif(1, min = 0, max = 5)`
> [1] 0.4

> `diff(punif(c(1, 3), min = 0, max = 5))`
> [1] 0.4
Example 4.4

Let $X$ be the time headway for two randomly chosen consecutive cars on a freeway during a period of heavy flow. The pdf of $X$ can be approximated by

$$f(x) = \begin{cases} 0.15e^{-0.15(x-0.5)} & x \geq 0.5 \\ 0 & \text{otherwise} \end{cases}$$

This is a density function since it is non-negative, and

$$\int_{-\infty}^{\infty} f(x) \, dx = \int_{0.5}^{\infty} 0.15e^{-0.15(x-0.5)} \, dx = -e^{-0.15(x-0.5)} \bigg|_{0.5}^{\infty} = 1.$$ 

The probability that headway time is at most 5 sec is

$$P(X \leq 5) = \int_{0.5}^{5} 0.15e^{-0.15(x-0.5)} \, dx = 0.491$$
The cumulative distribution function $F(x)$ for a continuous random variable $X$ is defined for every number $x$ by

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

For each $x$, $F(x)$ is the area under the density curve to the left of $x$. From this we see that $f(x) = F'(x)$ at every $x$ at which $F'(x)$ exists.

Example: Let $X$ have a uniform distribution on $[A, B]$. For $A \leq x \leq B$,

$$F(x) = \int_{-\infty}^{x} f(y) \, dy = \int_{A}^{x} \frac{1}{B-A} \, dy = \frac{x - A}{B - A}.$$  

For $x < A$, $F(x) = 0$. For $x \geq B$, $F(x) = 1$. If we are given this $F(x)$ to begin with, we can get $f(x)$ by taking the derivative.
Computing probabilities using the cdf

Let $X$ be a continuous rv with pdf $f(x)$ and cdf $F(x)$. Then for any number $a$,

$$P(X > a) = 1 - F(a)$$

and for any two numbers $a$ and $b$ with $a < b$,

$$P(a \leq X \leq b) = F(b) - F(a)$$
Example 4.6

Suppose the cdf of the magnitude $X$ of a dynamic load on a bridge is given by

$$F(x) = \begin{cases} 
\frac{x}{8} + \frac{3}{16}x^2 & 0 \leq x \leq 2 \\
1 & 2 < x 
\end{cases}$$

The probability that the load is between 1 and 1.5 is

$$P(1 \leq X \leq 1.5) = F(1.5) - F(1) = 0.297$$

The probability that the load exceeds 1 is

$$P(X > 1) = 1 - F(1) = 0.688$$
Percentiles of a continuous distribution

- Let $p$ be a number between 0 and 1. The \((100p)th\) percentile of the distribution of a continuous random variable $X$, denoted by $\eta(p)$, is defined by

$$p = F(\eta(p)) = P(X \leq \eta(p))$$

- That is, $\eta(p)$ is that value such that $100p\%$ of the area under the graph of $f(x)$ lies to the left of $\eta(p)$.

- The median $\tilde{\mu}$ is the 50th percentile. So half the area under the density curve is to the left of $\tilde{\mu}$. The median is one measure of the “central” value of the distribution.
Mean of a continuous random variable

- A different characterization of the center of the distribution is the expected value or mean of $X$
  \[ \mu_X = E(X) = \int_{-\infty}^{\infty} x \cdot f(x) \, dx. \]

- A symmetric continuous distribution — which means that the density curve to the left of some point is a mirror image of the density curve to the right of that point — has both median $\tilde{\mu}$ and mean $\mu_X$ equal to the point of symmetry.
Example 4.9

The distribution of the amount of gravel sold by a particular construction supply company in a given week is a continuous rv $X$ with pdf

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

The cdf is then, for $0 < x < 1$,

$$F(x) = \int_0^x \frac{3}{2} \left(1 - y^2\right) \, dy = \frac{3}{2} \left(x - \frac{x^3}{3}\right)$$

The $(100p)$th percentile satisfies

$$p = F(\eta(p)) = \frac{3}{2} \left(\eta(p) - \frac{(\eta(p))^3}{3}\right)$$
For the median, \( p = 0.5 \), and the equation is \( \tilde{\mu}^3 - 3\tilde{\mu} + 1 = 0 \). The solution is \( \tilde{\mu} = 0.347 \).

The mean is

\[
E(X) = \int_{-\infty}^{\infty} x f(x) \, dx = \int_{0}^{1} x \cdot \frac{3}{2} (1 - x^2) \, dx = \frac{3}{8}.
\]
Expected value of a function of a rv

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

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

For $h(X)$ a linear function, $E(aX + b) = aE(X) + b$. 
Example 4.10

Two species are competing for control of a certain resource. Let $X$ be the proportion controlled by species 1 and suppose $X$ has a uniform distribution on $[0, 1]$. Then the species that controls the majority of this resource controls the amount

$$h(X) = \max(X, 1 - X)$$

$$E[h(X)] = \int_{-\infty}^{\infty} \max(x, 1 - x) \cdot f(x) \, dx$$

$$= \int_0^1 \max(x, 1 - x) \cdot 1 \, dx$$

$$= \int_0^{1/2} (1 - x) \, dx + \int_{1/2}^1 x \, dx = \frac{3}{4}$$
Variance of a continuous rv

The **variance** of a continuous random variable $X$ with pdf $f(X)$ and mean value $\mu$ is

$$\sigma_X^2 = V(X) = E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) \, dx$$

The **standard deviation** (SD) of $X$ is $\sigma_X = \sqrt{V(X)}$. The variance or standard deviation tell us how “spread out” the distribution is.
A shortcut formula for variance

\[ V(X) = E(X^2) - [E(X)]^2 \]

For the \( X \) in Example 4.9,

\[ E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) \, dx = \int_{0}^{1} x^2 \cdot \frac{3}{2} (1 - x^2) \, dx = \frac{1}{5}. \]

\[ V(X) = E(X^2) - [E(X)]^2 = \frac{1}{5} - \left( \frac{3}{8} \right)^2 = 0.59. \]
A continuous rv $X$ is said to have a **normal distribution** with parameters $\mu$ and $\sigma$, where $-\infty < \mu < \infty$ and $0 < \sigma$, if the pdf of $X$ is

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

A shorthand notation is $X \sim \mathcal{N}(\mu, \sigma^2)$.

It can be shown that $E(X) = \mu$, and $V(X) = \sigma^2$. 
Normal density curves

N(0,1)  N(0,0.6^2)  N(0,2^2)  N(3,0.8^2)
The standard normal distribution

- $\mathcal{N}(0, 1)$ is called the standard normal distribution. A standard normal random variable will be denoted by $Z$. The cdf of $Z$ will be denoted by $\Phi(z) = P(Z < z)$.

- Appendix Table A.3 (reproduced on the inside front cover) can be used to obtain $\Phi(z)$ and the $(100p)$th percentile of $\mathcal{N}(0, 1)$.

- Example: Find $P(-.38 \leq Z \leq 1.25)$ and the 99th percentile of the standard normal distribution.

- Notation: $z_\alpha$ denotes the value for which $\alpha$ of the area under the standard normal density curve lies to the right of $z_\alpha$. That is $z_\alpha$ is the $[100(1 - \alpha)]$th percentile of $\mathcal{N}(0, 1)$. For example, $z_{0.05} = 1.645$. 
Nonstandard normal distribution

If \( X \sim \mathcal{N}(\mu, \sigma^2) \), then \( Z = \frac{X-\mu}{\sigma} \) has a standard normal distribution. Thus

\[
P(a \leq X \leq b) = P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)
\]

Example 4.15: The reaction time for an in-traffic response to a brake signal from standard brake lights can be modeled as \( X \sim \mathcal{N}(1.25, 0.46^2) \). Then

\[
P(1 \leq X \leq 1.75) = P\left(\frac{1-1.25}{0.46} \leq Z \leq \frac{1.75-1.25}{0.46}\right) = P\left(-0.54 \leq Z \leq 1.09\right) = 0.5675
\]

\[
> \text{diff(pnorm(c(1, 1.75), mean = 1.25, sd = 0.46))}
\]

\[
[1] \quad 0.5680717
\]
The Gamma distribution

For $\alpha > 0$, the Gamma function $\Gamma(\alpha)$ is defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \, dx$$

A continuous random variable $X$ is said to have a gamma distribution if the pdf of $X$ is

$$f(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\alpha > 0$ and $\beta > 0$. The standard gamma distribution has $\beta = 1$. $E(X) = \alpha \beta$ and $V(X) = \alpha \beta^2$. 
Gamma density curves
Example 4.21: Suppose the survival time $X$ in weeks of a randomly selected male mouse exposed to 240 rads of gamma radiation has a gamma distribution with $\alpha = 8$ and $\beta = 15$. What is the probability that a mouse survives between 60 and 120 weeks?

\[
> \text{diff(pgamma(c(60, 120), shape = 8, scale = 15))}
\]

[1] 0.4959056
The exponential distribution

A gamma distribution with $\alpha = 1$ and $\beta = 1/\lambda$ is also called an exponential distribution with parameter $\lambda$. The exponential distribution pdf is

$$f(x; \lambda) = \begin{cases} 
\lambda e^{-\lambda x} & x \geq 0 \\
0 & \text{otherwise}
\end{cases}$$

If $X$ is an exponential random variable with parameter $\lambda$, then

$$E(X) = \frac{1}{\lambda} \quad \text{and} \quad V(X) = \frac{1}{\lambda^2}$$
Exponential density curves

![Exponential density curves](image_url)
Let \( \nu \) be a positive integer. Then a random variable \( X \) is said to have a chi-squared distribution with parameter \( \nu \) if the pdf of \( X \) is the gamma density with \( \alpha = \nu/2 \) and \( \beta = 2 \). The pdf of a chi-squared rv \( \chi^2(\nu) \) is thus

\[
f(x; \nu) = \begin{cases} 
\frac{1}{2^{\nu/2} \Gamma(\nu/2)} x^{(\nu/2)-1} e^{-x/2} & x \geq 0 \\
0 & \text{otherwise}
\end{cases}
\]

The parameter \( \nu \) is called the number of degrees of freedom (df) of \( X \).
Chi-squared densities