# Chapter 4

# Continuous Random Variables and Probability Distributions

- A **discrete random variable** is a random variable that takes either a finite number of possible values or at most a countably infinite number of possible values.

- A **continuous random variable** is a random variable that takes an infinite number of possible values that is not countable. For any possible value $c$, $P(X = c) = 0$.

In this chapter, we introduce basic properties of continuous random variables and discuss important examples of continuous random variables and their probability distributions.

## 4.1 Probability Density Functions

Let $X$ be a continuous random variable (rv), then the **probability distribution** or **probability density function (pdf)** of $X$ is a function $f(x)$ such that for any two numbers $a$ and $b$ with $a \leq b$,

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

That is, the probability that $X$ falls in the interval $[a,b]$ is the area under the density function between $a$ and $b$.

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

The graph of $f(x)$ is open referred to as the **density curve**.

Any legitimate probability density function (pdf) $f(x)$, should satisfy the following conditions:

1. $f(x) \geq 0$, for all $x$
2. $\int_{-\infty}^{\infty} f(x)dx =$ area under the entire density curve $= 1$
   The rule of total probability holds.

---

For any possible value $c$, $P(X = c) = 0$

- $P(X = c) = P(c \leq X \leq c) = \int_c^c f(x)dx = 0$
- $P(c - \dfrac{\epsilon}{2} \leq X \leq c + \dfrac{\epsilon}{2}) = \int_{c-\frac{\epsilon}{2}}^{c+\frac{\epsilon}{2}} f(x)dx \approx \epsilon f(c)$

So the probability that $X$ falls in an interval of length $\epsilon$ around the point $c$ is approximately $\epsilon f(c)$. Thus $f(c)$ is a measure of how likely it is that the random variable will be near $c$.

Exercise 4.3 The error involved in making a certain measurement is a continuous random variable $X$ with pdf

$$f(x) = \begin{cases} .09375(4 - x^2) & -2 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Sketch the graph of $f(x)$.

(b) Compute $P(X > 0)$.

(c) Compute $P(-1 < X < 1)$.

(d) Compute $P(X < -.5$ or $X > .5)$.

# 4.2 Cumulative Distribution Functions and Expected Values

The **cumulative distribution function (cdf)** $F(x)$ for a continuous random variable $X$ is defined for every number $x$ by

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

Thus, $F(x)$ is the area under the density curve to the left of $x$. $F(x)$ is often referred to as the left-tail probability.

From $F(x)$ to $f(x)$ by the fundamental theorem of calculus

If $X$ is a continuous rv with pdf $f(x)$ and cdf $F(x)$, then at every $x$ at which the derivative $F'(x)$ exists,

$$F'(x) = f(x)$$

---

Let $X$ be a continuous rv with pdf $f(x)$ and cdf $F(x)$.

- $F(x) \to 0$, as $x \to -\infty$
- $F(x) \to 1$, as $x \to \infty$
- $F(x)$ is a monotonic non-decreasing functions of $x$
- $F(x)$ does not need to be smooth, but **is continuous**

- For any number $b$, $P(X > b) = 1 - F(b)$

- For any two number $a$ and $b$ with $a \le b$,

$$\begin{aligned} P(a \le X \le b) &= P(a < X \le b) = P(a \le X < b) = P(a < X < b) \\ &= \int_a^b f(x)dx = F(b) - F(a) \end{aligned}$$

Example $X$ is a **continuous** rv. Complete the following table.

| | Using pdf $f(x)$ | Using cdf $F(x)$ |
|---|---|---|
| $P(X \le a)$<br>$P(X < a)$ | | |
| $P(a \le X \le b)$<br>$P(a < X \le b)$<br>$P(a \le X < b)$<br>$P(a < X < b)$ | | |
| $P(X \ge b)$<br>$P(X > b)$ | | |

## Percentiles of a Continuous Distribution

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

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

The median of a continuous distribution, denoted by $\widetilde{\mu}$, is the 50th percentile, so $\widetilde{\mu}$ satisfies

$$F(\widetilde{\mu}) = .5$$
$$\eta(.5) = \widetilde{\mu}$$

That is, half the area under the density curve is to the left of $\widetilde{\mu}$ and half is to the right of $\widetilde{\mu}$.

If the pdf is symmetric, the median is the point of symmetry. That is $\widetilde{\mu} = \mu$.

Exercise 4.12 The cdf for $X$ = measurement error in Exercise 4.3 is

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

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

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

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

(d) Verify that $f(x)$ is as given in Exercise 4.3 by obtaining $F'(x)$.

(e) Verify that $\widetilde{\mu} = 0$.

## Expected Values

If $X$ is a continuous random variable with pdf $f(x)$, then its **expected value** or **mean value** is

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

For any function $h(X)$,

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

If $X$ is a continuous random variable with mean value $\mu$, then the **variance** of $X$, denoted by $V(X)$, is defined by

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

The **standard deviation (SD)** of $X$ is $\sigma_X = \sqrt{V(X)}$.

## Properties

- $E(aX + b) = aE(X) + b$
- $V(aX + b) = a^2 V(X)$.
- $E(X + Y) = E(X) + E(Y)$
- $V(X \pm Y) = V(X) + V(Y)$, if $X$ and $Y$ are independent.

Exercise 4.13 "Time headway" in traffic flow is the elapsed time between the time that one car finishes passing a fixed point and the instant that the next car begins to pass that point. Let $X$ = the time headway between two randomly chosen consecutive cars (sec.) in a traffic environment. The distribution of time headway has the form

$$f(x) = \begin{cases} \frac{k}{x^4} & x > 1 \\ 0 & x \leq 1 \end{cases}$$

(a) Determine the value of $k$ for which $f(x)$ is a legitimate pdf.

(b) Obtain the cumulative distribution function cdf $F(x)$.

(c) Use the cdf $F(x)$ to determine the probability that headway exceeds 2 sec and also the probability that headway is between 2 and 3 sec.

(d) Obtain the mean value and the standard deviation of headway.

(e) What is the probability that headway is within 1 standard deviation of the mean value?

## Uniform Distribution

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

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

We write $X \sim U(A, B)$.

Its mean and variance are

$$E(X) = \frac{A+B}{2} \qquad\qquad V(X) = \frac{(B-A)^2}{12}$$

Example Suppose buses arrive at a specific bus stop at 15-minute intervals. If a passenger arrives at the bus stop at a random time, then $X$ = waiting time in minutes, is uniformly distributed between 0 and 15.

(a) Compute $P(X = 5)$, $P(3 < X < 8)$, and $P(-1 < X < 5)$.

(b) For $a$ satisfying $0 < a < a + 5 < 15$, compute $P(a < X < a + 5)$.

(c) Compute $E(X)$ and $V(X)$.

(d) For the next 10 passengers, what's the probability that exactly five of them need to wait less than five minutes?

# 4.3 The Normal Distribution

The normal or "bell-shaped" distribution is the cornerstone of most methods of estimation and hypothesis testing developed in the rest of this course.

A continuous random variable $X$ is said to have a **normal distribution** with parameters $\mu$ and $\sigma$, where $-\infty < \mu < \infty$ and $\sigma > 0$, if the pdf of $X$ is given by

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

We write $X \sim N(\mu, \sigma^2)$.

Its mean and variance are

$$E(X) = \mu \qquad\qquad V(X) = \sigma^2$$

If $X \sim N(\mu, \sigma^2)$, then

- $X$ has a bell-shaped probability distribution.
- The probability distribution is perfectly symmetric and centered at its mean $\mu$.
- Its spread is determined by $\sigma$.
- There are an infinitely large number of normal curves, one for each pair of $\mu$ and $\sigma$.
- The total area under any normal curve is 1.

Specially, $N(\mu = 0, \sigma^2 = 1)$ is called **standard normal distribution**.

Let $Z$ denote a **standard normal rv**. That is, $Z \sim N(0, 1)$.
The pdf of $Z$ is

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

The cdf of $Z$ is denoted by

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} f(y; 0, 1) dy$$

Appendix Table A.3 gives $\Phi(z)$, the area under the standard normal density curve to the left of $z$.

If $Z \sim N(0, 1)$, then

- $P(Z \leq -z) = P(Z \geq z)$
- $\Phi(-z) = 1 - \Phi(z)$
- If $X = \mu + \sigma Z$, then $X \sim N(\mu, \sigma^2)$

## Standardizing Normal Distributions

If $X \sim N(\mu, \sigma^2)$ and $Z = \frac{X-\mu}{\sigma}$, then $Z \sim N(0, 1)$.

- $F(X) = \Phi(Z)$
- $P(a \leq X \leq b) = F(b) - F(a) = \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$
- [(100$p$)th percentile of $X$] = $\mu + \sigma \times$ [(100$p$)th percentile of $Z$]

# 4 Typical Problems about Normal Distribution

- ► Compute $P($about $Z)$
- ► Find $z$-values
- ► Compute $P($about $X)$
- ► Find $x$-values

► Compute $P($about $Z)$ – Use Table A.3 to find $\Phi(z)$

- $P(Z \leq a) = \Phi(a)$
- $P(a \leq Z \leq b) = \Phi(b) - \Phi(a)$
- $P(Z \geq b) = 1 - \Phi(b)$

Example Suppose $Z \sim N(0, 1)$, find the following probabilities

(a) $P(Z < 1)$

(b) $P(0.38 \leq Z \leq 1.25)$

(c) $P(Z \geq -3)$

## ► Find $z$-values

1. Find the $z$-value's left-tail probability $\Phi(z)$
2. Search for the $z$-value in Table A.3

Example: Find a value of the standard normal random variable $Z$, call it $z_0$, such that

(a) $P(Z \geq z_0) = 0.2090$

(b) $P(Z < z_0) = 0.7090$

(c) $P(-z_0 \leq Z < z_0) = 0.8472$

► Compute $P(\text{about } X)$, where $X \sim N(\mu, \sigma^2)$

1. $Z = \frac{X-\mu}{\sigma}$, i.e., standardizing $X$
2. Compute $P(\text{about } Z)$

Example Suppose $X \sim N(3, 9)$, Compute

(a) $P(3 < X < 6)$
(b) $P(X > 0)$
(c) $P(|X - 3| > 6)$

Example Suppose $X \sim N(\mu, \sigma^2)$, find $P(\mu - c\,\sigma \le X \le \mu + c\,\sigma)$ where

(a) $c = 1$
(b) $c = 2$
(c) $c = 3$

► Find $x$-values – Think Percentiles!

1. Find the desired $z$-value such at $F(x) = \Phi(z)$
2. Convert $z$ to $x$ using $x = \mu + \sigma z$.

Example Suppose $X \sim N(9, 4)$, find

(a) $\eta(.75)$

(b) $\eta(.08)$

Example Suppose $X \sim N(6, 16)$, find a value of the normal random variable $X$, call it $x_0$, such that

(a) $P(X \leq x_0) = 0.8531$

(b) $P(X > x_0) = 0.95$

Exercise 4.35 Suppose the diameter at breast height (in.) of trees of a certain type is normally distributed with $\mu$ = 8.8 and $\sigma$ = 2.8.

(a) What is the probability that the diameter of a randomly selected tree will be at least 10 in.? Will exceed 10 in.?

(b) What is the probability that the diameter of a randomly selected tree will exceed 20 in.?

(c) What is the probability that the diameter of a randomly selected tree will be between 5 and 10 in.?

(d) What value $c$ is such that the interval $(8.8 - c, 8.8 + c)$ includes 98% of all diameter values?

(e) If four trees are independently selected, what is the probability that at least one has a diameter exceeding 10 in.?

The normal distribution is often used as an approximation to the distribution of values in a discrete population. In such situations, extra care should be taken to ensure that probabilities are computed in an accurate manner.

## Normal Approximation of Binomial Distribution

Let $X \sim Bin(n, p)$ be a binomial rv based on $n$ trials with success probability $p$. Then if the binomial probability histogram is not too skewed, $X$ has approximately a normal distribution

$$N\left(\mu = np, \sigma^2 = np(1-p)\right)$$

In practice, the approximation is adequate provided that both $np \geq 10$ and $n(1-p) \geq 10$, since there is then enough symmetry in the underlying binomial distribution.

Example Assume that $X \sim Bin(n, p)$. In which of the following cases would it be appropriate to use the normal approximation to the binomial?

(a) $n$ = 100, $p$ = .01

(b) $n$ = 25, $p$ = .6

(c) $n$ = 10, $p$ = .4

Since we are using a continuous probability distribution to approximate probabilities for a discrete probability distribution and for $X \sim N(\mu, \sigma^2)$,

$$P(X = x) = 0, \quad -\infty < x < \infty,$$

and for $X \sim Bin(n, p)$,

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \cdots, n,$$

we must do **continuity correction**.

If $X \sim Bin(n, p)$ can be approximated by $N(\mu, \sigma^2)$, where $\mu = np$ and $\sigma = \sqrt{np(1-p)}$, we have

$$Z = \frac{X - np}{\sqrt{np(1-p)}} \overset{approx.}{\sim} N(0, 1)$$

In particular, for $x$ = a possible value of $X$,

$$\begin{aligned} P(X \le x) = B(x; n, p) &\approx \begin{pmatrix} \text{area under the normal curve} \\ \text{to the the left of } x + 0.5 \end{pmatrix} \\ &= P(X' \le x + 0.5) \\ &= P\left(\frac{X' - \mu}{\sigma} \le \frac{x + 0.5 - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{x + 0.5 - \mu}{\sigma}\right) = \Phi\left(\frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

Example Suppose $X \sim Bin(n, p)$, with $np \geq 10$ and $n(1-p) \geq 10$.
Show how to do continuity correction when calculating the following probabilities using normal approximation.

(a) $P(X = 10)$

(b) $P(X \leq 12)$

(c) $P(X < 9)$

(d) $P(X \geq 19)$

(e) $P(X > 28)$

(f) $P(2 \leq X \leq 17)$

(g) $P(3 \leq X < 27)$

(h) $P(6 < X \leq 40)$

(i) $P(5 < X < 8)$

Exercise 4.54 Suppose that 10% of all steel shafts produced by a certain process are nonconforming but can be reworked (rather than having to be scrapped). Consider a random sample of 200 shafts, and let $X$ denote the number among these that are nonconforming and can be reworked. What is the (approximate) probability that $X$ is

(a) At most 30?

(b) Less than 30?

(c) Between 15 and 25 (inclusive)?

# 4.4 The Exponential Distribution

A continuous rv $X$ is said to have an **exponential distribution** with parameter $\lambda > 0$, if the pdf of $X$ is

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

The cdf of the exponential rv $X$ is

$$F(x;\lambda) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

And its mean and variance are

$$E(X) = \frac{1}{\lambda} \qquad\qquad V(X) = \frac{1}{\lambda^2}$$

Relationship between Poisson and Exponential distributions

Suppose that the number of events occurring in any time interval of length $t$ has a **Poisson** distribution with parameter $at$ (where $a$, the rate of the event process, is the expected number of events occurring in 1 unit of time) and that numbers of occurrences in non-overlapping intervals are independent of one another. Then the distribution of elapsed time between the occurrence of two successive events is **exponential** with parameter $\lambda = a$.

The exponential distribution is frequently used as a model for the distribution of elapsed time between the occurrence of successive events, such as customers arriving at a service facility or calls coming in to a switchboard.

Exercise 4.59 Let $X$ = the time between two successive arrivals at the drive-up window of a local bank. If $X$ has an exponential distribution with $\lambda = 1$ , compute the following:

(a) The expected time between two successive arrivals

(b) The standard deviation of the time between successive arrivals

(c) $P(X \leq 4)$

(d) $P(2 \leq X \leq 5)$