

BIOS:4120 – Introduction to Biostatistics

Unit 6: Theoretical Probability Distributions

Knute D. Carter

Department of Biostatistics
The University of Iowa

September 16, 2025

Learning Objectives (1/2)

At the end of this unit, you should be able to:

- Define what a random variable is and distinguish between discrete and continuous random variables.
- Explain differences between probability mass and density functions and the properties of each.
- Describe situations under which data from either the binomial or Poisson distribution may arise.
- Calculate probabilities from binomial and Poisson distributions; and also the mean, variance, and standard deviation of each.

Learning Objectives (2/2)

At the end of this unit, you should be able to:

- List the criteria that characterize a Poisson process.
- Explain when the Poisson distribution approximates the binomial distribution.
- Apply the additive property of Poisson distributions.
- Describe the basic properties of the standard normal distribution, and list the common critical values and percentiles.
- Calculate standardized values and calculate probabilities from the general normal distribution.

Overview

- Theoretical Probability Distributions
- The Binomial (Bernoulli) Distribution
- The Poisson Distribution
- The Normal (Gaussian) Distribution

Variables

- A *Variable* is a generic term we use to describe any quantitative or qualitative measure we record.
- A *Random Variable* is a numeric variable that assumes a value based on the outcome of a random experiment.
- Formal definition:
Consider a random experiment with a sample space \mathcal{S} .
A function X , which assigns to each element $s \in \mathcal{S}$ one and only one number $X(s) = x$, is called a *Random Variable*.

Variables

- *Discrete Random Variable*: may assume only specific numeric values (often integers).
- *Continuous Random Variable*: may assume any value over some interval or continuum.

Examples of Random Variables

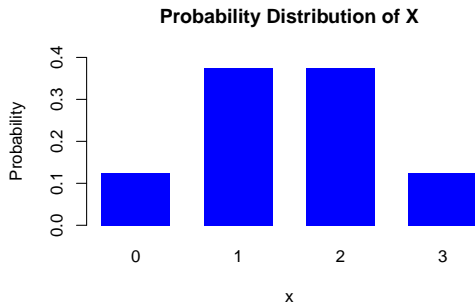
- $X =$ the # of heads that appear when a coin is tossed 3 times.
- X is
- Sample space for $X : \mathcal{S} = \{$
- $T =$ the high temperature in Iowa City (in Fahrenheit) on a randomly selected summer day.
- T is
- Sample space for $T : \mathcal{S} = \{$

Definitions

- The *Probability Distribution* of a discrete random variable X is a function $p(x)$ which assigns every possible value of X a probability.
- A *Probability Histogram* is a graphical representation of a probability distribution.

Example: Coin Toss

x	Probability, $p(x)$ or $P(X = x)$
0	$1/8 = 0.125$
1	$3/8 = 0.375$
2	$3/8 = 0.375$
3	$1/8 = 0.125$



Application of $p(x)$ to Example

- Let A denote the event that 2 heads are obtained
- Let B denote the event that at least 2 heads are obtained.
- $P(A) = P(X = 2) = 0.375$.
- $P(B) = P(X = 2 \cup X = 3) = P(X = 2) + P(X = 3) = 0.5$.
- Suppose we want to find $P(A|B)$:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{0.375}{0.5} = 0.75.$$

Probability Mass Function

- The probability function of a discrete random variable (RV):
 1. defines all possible values of the variable;
 2. displays the probabilities with which the random variable takes on those values;
 3. can sometimes be described using a formula.
- For discrete random variables the function $p(x)$ is often referred to as the *probability mass function* (or *pmf*).
- For each value of the RV, the pmf gives the probability of that value happening.

Factorials and Combinations

- Let n be a positive integer.
- The *Factorial* of n , denoted $n!$, is defined as

$$n! = n \times (n - 1) \times (n - 2) \times \cdots \times 2 \times 1.$$

- For example,

$$3! = 3 \times 2 \times 1 = 6.$$

$$6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720.$$

$$10! = 10 \times 9 \times 8 \times \cdots \times 2 \times 1 = 3,628,800.$$

- So what does, $0! = ?$.

Combination Examples

$$\binom{3}{2} = \frac{3!}{2!(3-2)!} = 3$$

$$\binom{6}{3} = \frac{6!}{3!(6-3)!} = 20$$

$$\binom{6}{6} = \frac{6!}{6!(6-6)!} = 1$$

Example

Consider a list of 8 patients.

- a) In how many ways can the names on this list be ordered?

- b) In how many ways can a sample of 4 of these patients be drawn (without regard to order)?

Bernoulli Trials

Consider a sequence of random experiments that satisfy the following criteria.

1. Each trial has only 2 possible outcomes. These outcomes are often referred to as 'successes' and 'failures.'
2. The trials are independent.
3. The probability of a success, p , remains the same from trial to trial.

Bernoulli Trials

Examples:

- 25 tosses of a coin.
- 12 tosses of a die, where the outcomes are classified as 'even numbers' or 'odd numbers.'
- 1000 ELISA test results for HIV status (positive or negative), conducted on 1000 blood samples randomly drawn from a large blood bank.

Properties of the Binomial Distribution

If Y is distributed $\text{Bin}(n, p)$, that is $Y \sim \text{Bin}(n, p)$, then:

- the mean of Y , denoted $E(Y)$ for the 'expectation of Y ,' is equal to np ;
- the variance of Y , denoted $\text{Var}(Y)$ or $V(Y)$, is $np(1 - p)$;
- the standard deviation of Y , usually denoted $sd(Y)$ or $SD(Y)$, is the square root of the variance and therefore equals $\sqrt{np(1 - p)}$.

The Bernoulli Distribution

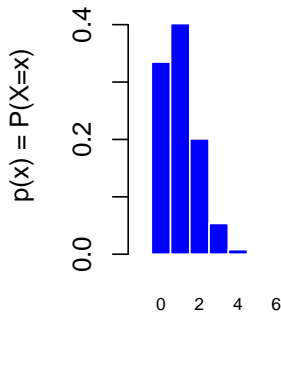
If $n = 1$ (one trial only), this distribution is also known as the *Bernoulli* distribution, for which:

- $E(Y) = \mu = p$
- $Var(Y) = \sigma^2 = p(1 - p)$
- $sd(Y) = \sigma = \sqrt{p(1 - p)}$

In other words, the Binomial random variable Y is just the count (sum) of successes in n independent Bernoulli trials.

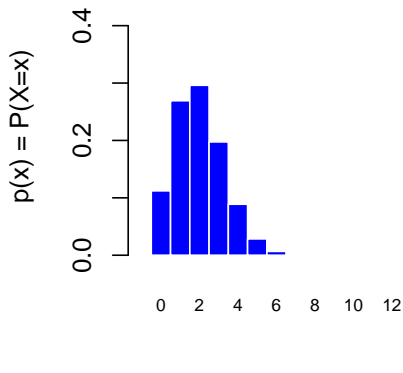
Probability Histograms for Some Binomial Distributions

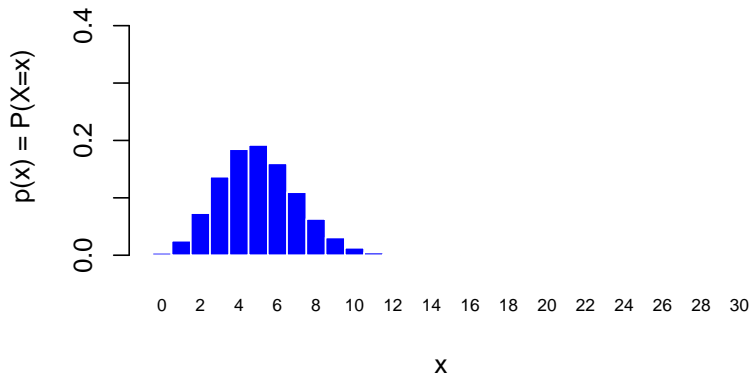
Bin(6, 1/6)



Probability Histograms for Some Binomial Distributions

Bin(12, 1/6)





Example: Blood Types

- Recall (from Chapter 6) the probability assignments for O, A, B, and AB blood types for a randomly chosen person in the United States.

Blood type	O	A	B	AB
U.S. probability	0.45	0.40	0.11	0.04

- Consider a random sample of 10 Americans. Let the random variable X count the number of individuals sampled with the blood type A.
- a) What is the mean of X ? What is the standard deviation of X ?

Example: Blue or Brown Eyes

- A newly married couple is planning to have 6 children. Assume that the probability the couple will conceive a child with blue eyes is 0.25, and the probability they will conceive a child with brown eyes is 0.75.
- Let X count the number of children conceived by the couple who will have blue eyes; let Y count the number of children who will have brown eyes.

a) What is the mean and standard deviation of X ?

Example: Blue or Brown Eyes

- b) Find the probability that the couple will have exactly one child with blue eyes.
- c) Find the probability that the couple will have at least one child with blue eyes.

Example: Blue or Brown Eyes

- d) Find the probability that the couple will have at least one child with brown eyes.

Estimation for the Binomial Distribution

- In the Binomial distribution, the parameter n (number of independent trials) will always be known, so there is no need to estimate it.
- The parameter p is often unknown as must be estimated from the sample data.

$$\hat{p} =$$

- Sometimes the Greek letter π is used to denote the underlying (often unknown) parameter value, and then either $\hat{\pi}$ or the lower case letter p to denote the sample estimate of that parameter (which can vary from sample to sample and thus is a random variable itself).

Estimation for the Binomial Distribution

Case 1:

- If the random variable of interest is the *proportion* of successes in a Binomial data situation.
- Estimated proportion of successes: \hat{p}
- Estimated variance of $\hat{p} = \frac{\hat{p}(1-\hat{p})}{n}$
- Estimated std. dev. of $\hat{p} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Estimation for the Binomial Distribution

Case 2:

- The random variable of interest is the *number* of successes.
- Estimated mean number of success: $n\hat{p}$
- Estimated variance of the # of successes = $n\hat{p}(1 - \hat{p})$
- Estimated std. dev. of the # of successes = $\sqrt{n\hat{p}(1 - \hat{p})}$

The Poisson Distribution

- Consider a process which results in a specific event that occurs periodically over time or space.
- The Poisson distribution is often used to assign probabilities to the number of events that take place over a certain interval of time or region of space.

The Poisson Distribution

A *Poisson Process* is characterized by the following criteria.

- An event occurs periodically over time (or space).
- The expected number of events in an interval is proportional to the length of the interval.
- Within a single interval, an infinite number of occurrences of the event are theoretically possible.
- The events occur independently, both within the same interval and between consecutive intervals.

The Poisson Distribution

- Given a Poisson process, let X count the number of occurrences of the event of interest over a certain interval. Assume that the mean number of occurrences over an interval of this length is denoted by λ .
- X is said to be a *Poisson Random Variable*.

The Poisson Distribution

- The probability distribution of X is the *Poisson Distribution*, and may be denoted as $\text{Poisson}(\lambda)$, $\text{Poi}(\lambda)$ or $P(\lambda)$.
- This distribution is characterized by

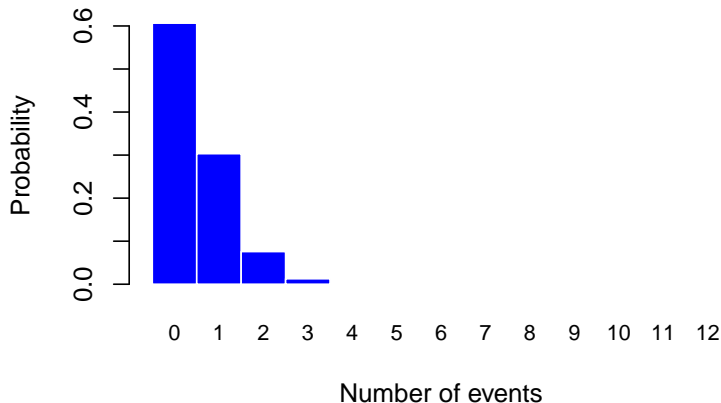
$$p(x) = P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

where the possible values of X are $0, 1, 2, \dots, \infty$.

- Here, $e = 2.71828$, the base of the natural logarithm.
- The mean of X is given by $\mu = \lambda$.
- The variance of X is given by $\sigma^2 = \lambda$.

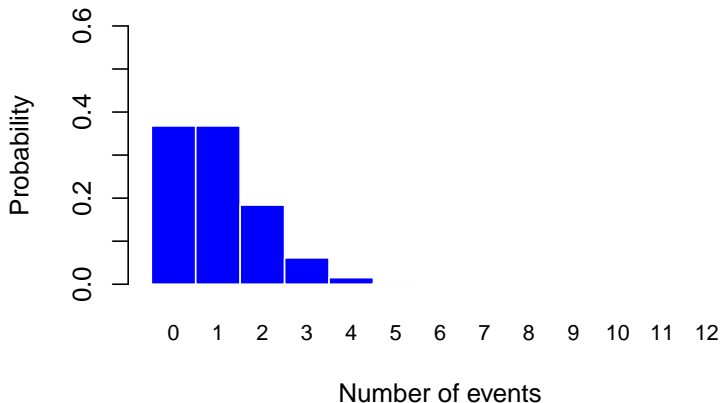
The Poisson Distribution

Poisson: Lambda = 0.5



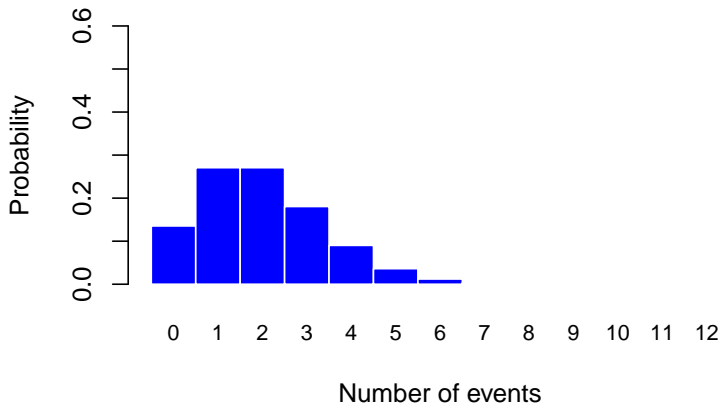
The Poisson Distribution

Poisson: Lambda = 1.0



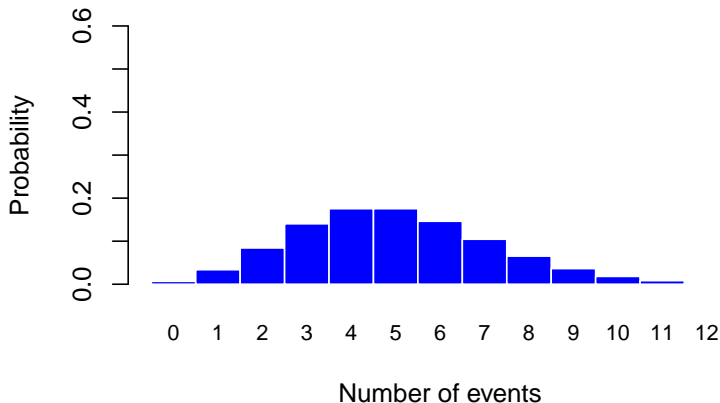
The Poisson Distribution

Poisson: Lambda = 2.0



The Poisson Distribution

Poisson: Lambda = 5.0



Examples of Poisson Random Variables

- The number of babies delivered in the maternity ward of UIHC on a particular day.
- The number of car accidents that occur at a busy intersection in Iowa City during a certain month.
- The number of calls received by Quitline Iowa (a statewide smoking cessation telephone counseling hotline) during a specific week.

The Poisson Distribution

- Let X denote the number of patients that arrive at the UIHC Emergency Treatment Center (ETC) during the early morning hours of a certain day (between midnight and 6:00 a.m.).
 - Assume that on average, 4.5 patients arrive at the UIHC ETC during the early morning hours.
- a) What is the probability that no patients will arrive at UIHC ETC during the given time period?

The Poisson Distribution

- b) What is the probability that at least one patient will arrive at UIHC ETC during the given time period?
- c) What is the probability that four or five patients will arrive at UIHC ETC during the given time period?

The Poisson Distribution

- Note: The Poisson distribution can be used to approximate probabilities for binomial random variables where n is large and p is small.
- In such applications, λ is set equal to the product np .
- Thus in such situations, if

$$X \sim \text{Bin}(n, p) \text{ then } X \overset{\text{approx}}{\sim} \text{Poisson}(np).$$

- An example of such an application will be covered in a lab session.

The Poisson Distribution

- Suppose X_1 and X_2 are independent Poisson random variables, such that $X_1 \sim \text{Poisson}(\lambda_1)$ and $X_2 \sim \text{Poisson}(\lambda_2)$ then

$$Y = X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$$

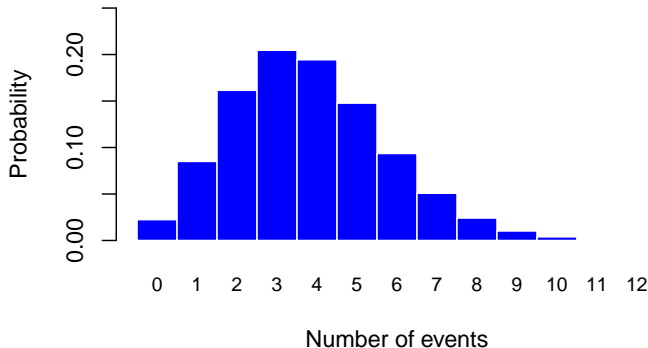
Probability Mass Function (Review)

- The probability function of a **discrete random variable**:
 1. defines all possible values of the variable;
 2. displays the probabilities with which the random variable takes on those values;
 3. can sometimes be described using a formula.
- For discrete random variables the function $p(x)$ is often referred to as the **probability mass function** (or **pmf**).
- For each value of the RV, the pmf gives the probability of that value happening.

Probability Mass Function: Example

If $X \sim \text{Poisson}(3.8)$ then $P(X = k) = \frac{e^{-3.8} 3.8^k}{k!}$ for $k = 0, 1, 2, \dots$

Poisson: Lambda = 3.8



Probabilities for Continuous Random Variables

- How do we calculate probabilities for continuous random variables?
- What is the probability of any specific value of a continuous random variable?
- What about the probability of being within a specified range of values?

Probabilities for Continuous Random Variables

- The *Probability Distribution* of a *continuous random variable* X is an unbroken curve or function $f(x)$ which is used to assign probabilities to X .
- Probabilities are assigned in the following manner: the area under the curve over an interval is equal to the probability the random variable will assume a value within that interval.
- $f(x)$ is often called a *Density Curve* or a *probability density function* (*pdf*).

Probability Density Function

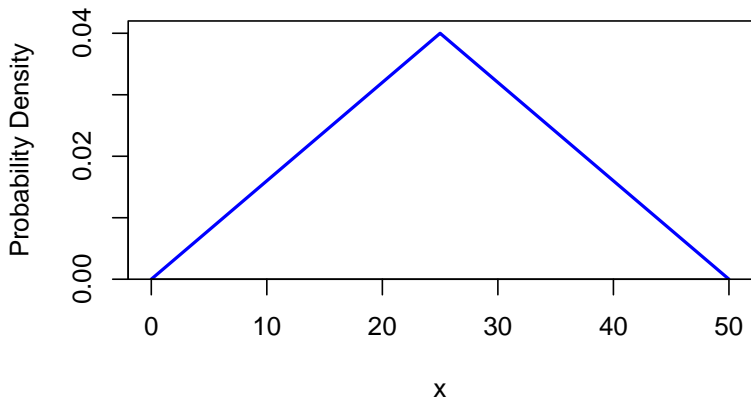
- A probability density function of a **continuous random variable** X is a function which satisfies:
 1. The probability that X falls between values a and b , that is $P(a \leq X \leq b)$, is equal to the area under the curve between a and b .
 2. The function always takes on values greater than or equal to zero.
 3. The total area under the curve is equal to 1.
- In general, calculation of probabilities for continuous random variables requires calculus.
- For a continuous random variable, the probability of the random variable assuming any specific value is zero.

Example: Study Hours

- Let X denote the time spent by a randomly selected college student studying each week.
- X is continuous.
- Suppose that the sample space for X is given by $\mathcal{S} = \{0 \leq X \leq 50\}$.
- Assume that the density curve for X is as illustrated in the next figure. (The vertical axis is scaled so that the entire area under the curve is equal to 1.)

Example: Study Hours

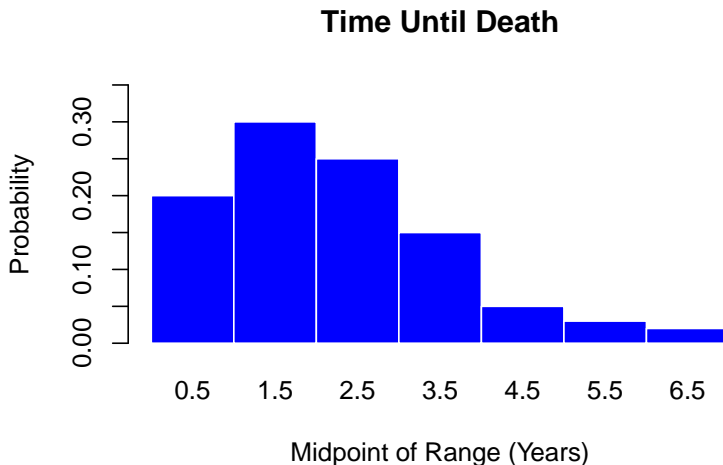
Probability Density Function of X



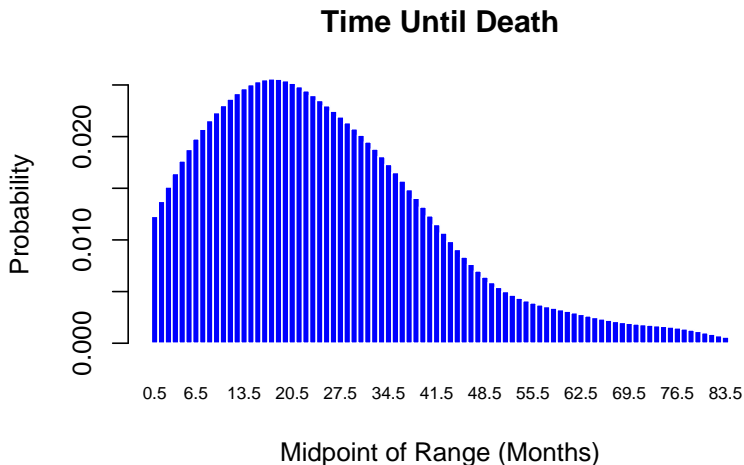
Example: Study Hours

- $P(X \leq 25) =$
- $P(37.5 \leq X \leq 50) =$
- $P(X \leq 37.5) =$
- $P(X = 37.5) =$
- $P(0 \leq X \leq 50) =$

Example: Time to Death from Diagnosis

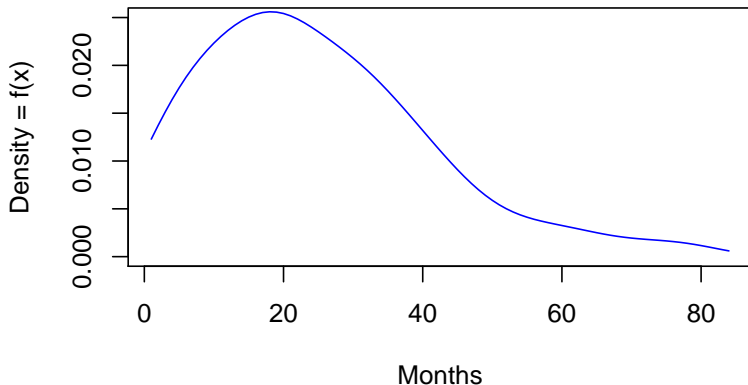


Example: Time to Death from Diagnosis



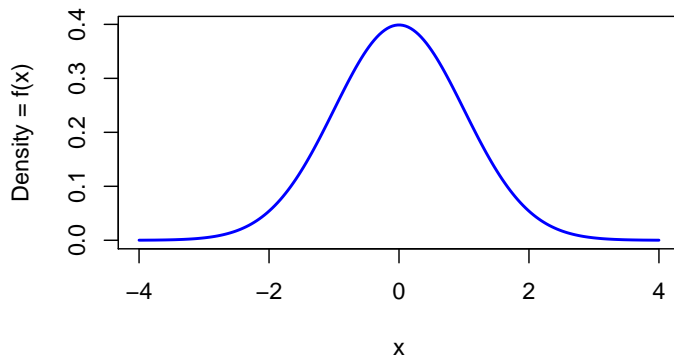
Example: Time to Death from Diagnosis

Time Until Death



The Normal Distribution

Standard Normal Probability Density Function



- The normal (or Gaussian) distribution is the most common probability distribution for continuous random variables.

The Normal Distribution

- The density curve for the normal distribution is called the **Normal Curve** or the **Bell Curve**.
- Has wide applicability to many different types of data, and often arises in nature.
- Widely used for theoretical properties of many statistical methods.
- The distribution depends on only two parameters:
the mean, μ ; and
the standard deviation, σ , (or variance, σ^2).

The Normal Distribution

- The formula for the normal density function is given by

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

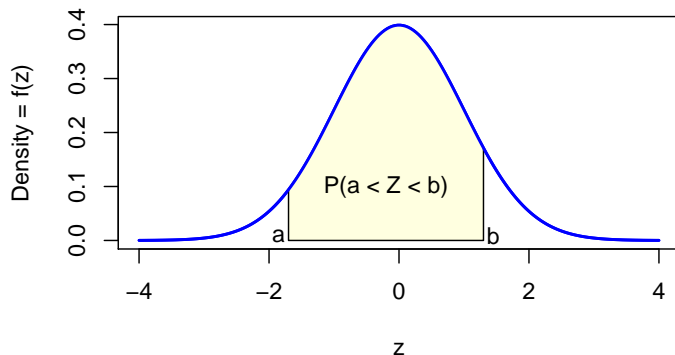
where $-\infty < x < \infty$, and $\pi = 3.14159\dots$

- This is often denoted as $X \sim N(\mu, \sigma^2)$.
- Caution: some authors use $X \sim N(\mu, \sigma)$.

The Standard Normal Distribution

- The *Standard Normal Distribution* is a normal distribution with $\mu = 0$, and $\sigma = 1$.
- A standard normal random variable is generally denoted as Z .
- That is, $Z \sim N(0, 1)$.

Computing Probabilities for Standard Normal Variables



- The area between a and b under the standard normal density curve provides the probability that Z will assume a value over the interval (a, b) : $P(a < Z < b)$.

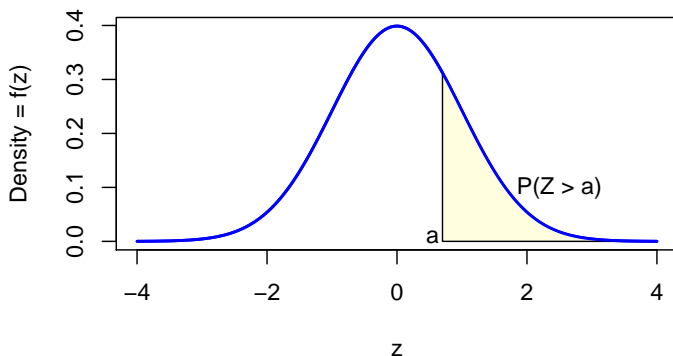
Computing Probabilities for Standard Normal Variables

- Note the following equivalences:

$$\begin{aligned}
 P(a < Z < b) &= P(a \leq Z < b) \\
 &= P(a < Z \leq b) \\
 &= P(a \leq Z \leq b)
 \end{aligned}$$

- This occurs because $P(Z = a) = P(Z = b) = 0$.

Computing Probabilities for Standard Normal Variables



- Our Normal Table provides upper tail probabilities for the standard normal distribution: i.e., $P(Z > a)$ where a is between 0.00 and 3.49.

Computing Probabilities for Standard Normal Variables

- For a number a having the format $a = w.xy$, the digits $w.x$ are looked up in the left-hand column of the table, and the remainder $0.0y$ is looked up in the top-most row of the table.

z (or a)	0.00	0.01	0.02	...
0.0	0.500	0.496	0.492	...
0.1	0.460	0.456	0.452	...
0.2	0.421	0.417	0.413	...
0.3	0.382	0.378	0.374	...
0.4	0.345	0.341	0.337	...
\vdots	\vdots	\vdots	\vdots	\ddots

Computing Probabilities for Standard Normal Variables

Examples:

- $P(Z > 0.21) =$
- $P(Z \geq 0.21) =$
- $P(Z > 0.42) =$
- $P(Z > 0.20) =$
- $P(Z > 0.00) =$

Further examples (using Normal Table)

- $P(Z > 1.96) = 0.025$
- $P(Z > 2.15) = 0.016$
- $P(Z > 3.00) = 0.001$

Computing Probabilities for Standard Normal Variables

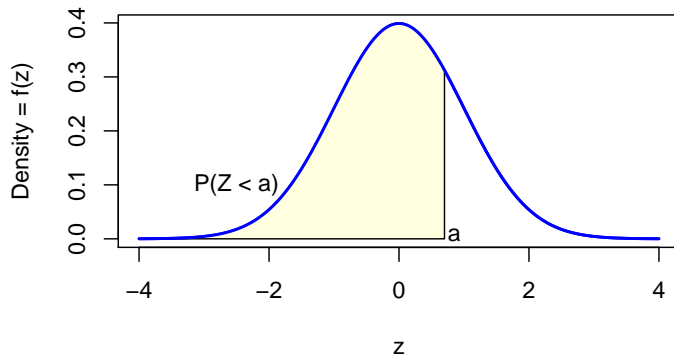
- Areas that do not correspond to upper tail areas can be suitably pieced together using the symmetry of the standard normal curve.
- For instance,

$$P(Z < a) = 1 - P(Z \geq a)$$

$$P(Z < -a) = P(Z > a) \text{ where } a > 0$$

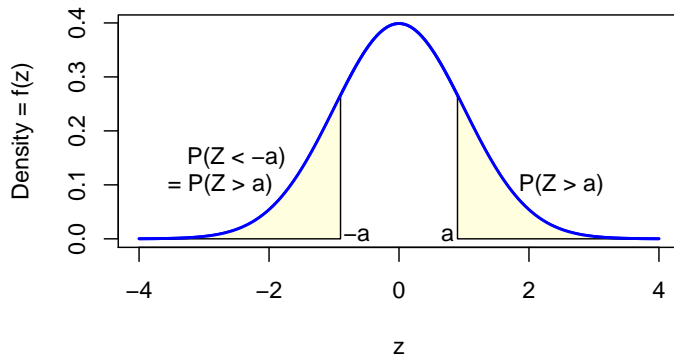
$$P(-a < Z < a) = 1 - 2P(Z \geq a) \text{ where } a > 0$$

Computing Probabilities for Standard Normal Variables



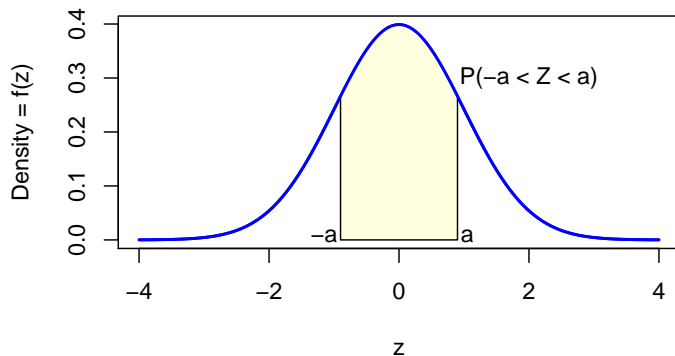
- $P(Z < a) = 1 - P(Z \geq a)$

- $P(Z < -a) = P(Z > a)$ where $a > 0$



- $P(Z < -a) = P(Z > a)$ where $a > 0$

Computing Probabilities for Standard Normal Variables



- $P(-a < Z < a) = 1 - P(Z \leq -a) - P(Z \geq a)$
 $= 1 - 2P(Z \geq a)$ where $a > 0$

Probabilities for Standard Normal Variables: Examples

d) $P(-5.13 \leq Z \leq 2.00)$

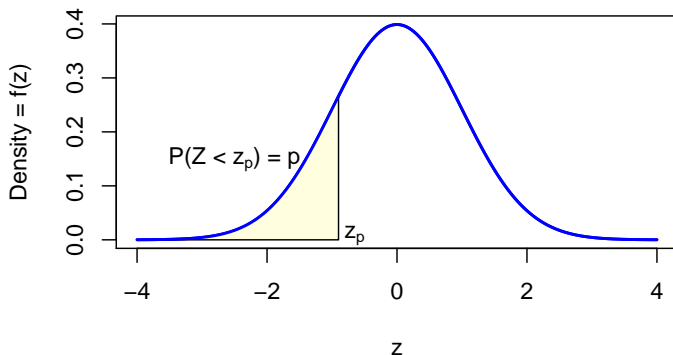
e) $P(Z = 1.71)$

- In a certain population, suppose that diastolic blood pressure is normally distributed with a mean of 75 mm Hg and a standard deviation of 10 mm Hg.
- a) What is the probability that a randomly selected individual from this population has a diastolic blood pressure in the range 73–77 mm Hg?

- b) Suppose 5 individuals are randomly selected from this population. What is the probability exactly 3 of these individuals have diastolic blood pressures in the range 73–77 mm Hg?
- c) What is the median of the distribution of diastolic blood pressure?

Percentiles of the Standard Normal Distribution

- The $100\% \times p^{\text{th}}$ percentile of the $N(0, 1)$ distribution is denoted by z_p .
- z_p is the value such that $P(Z \leq z_p) = p$.



Percentiles of the Standard Normal Distribution: Examples

- a) Find a value such that 30% of the area under the standard normal curve lies above (to the right of) z .
- b) Find a value such that 2.5% of the area under the standard normal curve lies below (to the left of) z .

Learning Objectives (1/2)

At the end of this unit, you should be able to:

- Define what a random variable is and distinguish between discrete and continuous random variables.
- Explain differences between probability mass and density functions and the properties of each.
- Describe situations under which data from either the binomial or Poisson distribution may arise.
- Calculate probabilities from binomial and Poisson distributions; and also the mean, variance, and standard deviation of each.

Learning Objectives (2/2)

At the end of this unit, you should be able to:

- List the criteria that characterize a Poisson process.
- Explain when the Poisson distribution approximates the binomial distribution.
- Apply the additive property of Poisson distributions.
- Describe the basic properties of the standard normal distribution, and list the common critical values and percentiles.
- Calculate standardized values and calculate probabilities from the general normal distribution.