STAT230: Principles of Probability

Unit 5b: Continuous Random Variables

Department of Statistics

United Arab Emirates University

April 10, 2022

Continuous Random Variables

- A random variable X is called continuous if it can take any value within a finite or infinite interval of the real line $(-\infty, \infty)$. Thus, the sample space of a continuous random variable has uncountably infinite values.
- Examples of experiments resulting in continuous variables include measurements of time required to assemble a computer, weight of new born baby, lifetime of a TV, proportion of used space in a hard disk, salary of an engineer.
- Note that, although a continuous variable can take any possible value in an interval, its measured value cannot. This is because no measuring device has infinite precision. Thus, continuous variables do not really exist in real life; they are only ideal versions of the discrete variables which are measured. Nevertheless, the study of continuous random variables is meaningful as it provides useful, and quite accurate, approximations to probabilities pertaining to their discrete versions.

Continuous Random Variables

- We say that X is a continuous random variable if there exists a nonnegative function f , defined for all real $x \in (-\infty, +\infty)$, having the property that, for any set $\mathcal{B}$ of real numbers,

$$P(X \in \mathcal{B}) = \int_{\mathcal{B}} f(x)dx \tag{1}$$

- The function f is called the *probability density function* of the random variable X.

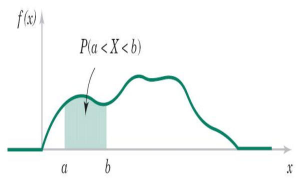
Probability Density Function

A probability density function (pdf), $f(x)$, can be used to describe the probability distribution of a continuous random variable X.

Definition (Probability Density Function)

A random variable X is said to be continuous if there is a function $f(x)$, called the probability density function (pdf), such that

1. $f(x) \geq 0, \qquad$ for all x.
2. $\int\limits_{-\infty}^{\infty} f(x)dx = 1.$
3. $P(a \leq X \leq b) = \int\limits_{a}^{b} f(x)dx.$



Note:

- $P(X = c) = 0 \qquad$ for any real number c
- $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$

Example

Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} C(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{Otherwise} \end{cases}$$

(a) What is the value of C?

(b) Find $P[X > 1]$.

<u>Solution:</u>

(a) Since f is a pdf, we must have $\int_0^2 f(x)dx = 1$.

$$\int_0^2 f(x)dx = C \int_0^2 (4x - 2x^2)dx = 1. \tag{2}$$

Then, $C = 3/8$.

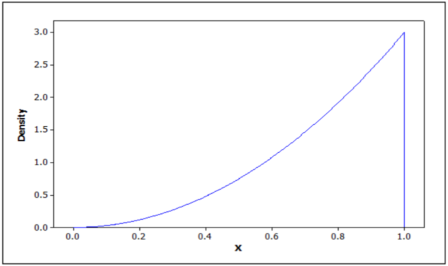
(b) $P[X > 1] = 3/8 \int_1^2 (4x - 2x^2)dx = 1/2$.

Example

For a given IT technician in a support center, let X denote the percentage of time, out of a 40-hour work week, that he is directly serving customers. Suppose that X has a probability density function given by

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

(a) Graph $f(x)$.



Example

(b) Find the probability that the technician will spend less than 30% of his workweek serving customers.

$$P(X < 0.3) = \int_0^{0.3} 3x^2 dx = x^3 \Big|_0^{0.3} = 0.3^3 - 0^3 = 0.027$$

(c) Find the probability that the technician will spend 20 to 70% of his workweek serving customers.

$$P(0.2 < X < 0.7) = \int_{0.2}^{0.7} 3x^2 dx = x^3 \Big|_{0.2}^{0.7} = 0.7^3 - 0.2^3 = 0.337$$

(1) The amount of time in hours that a computer functions before breaking down is a continuous random variable with probability density function given by

$$f(x) = \begin{cases} \lambda e^{-x/100} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

What is the probability that

(a) a computer will function between 50 and 150 hours before breaking down?

(b) it will function for fewer than 100 hours?

(2) The lifetime in hours of a certain kind of radio tube is a random variable having a probability density function given by

$$f(x) = \begin{cases} \frac{100}{x^2} & x > 100 \\ 0 & x \leq 100 \end{cases}$$

What is the probability that exactly 2 of 5 such tubes in a radio set will have to be replaced within the first 150 hours of operation? Assume that the events E_i, $i = 1, 2, 3, 4, 5$, that the i^{th} such tube will have to be replaced within this time are independent.

Cumulative Distribution Function (cdf)

Definition (The Cumulative Distribution Function)

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

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

Properties of $F(x)$

1. $P(a \leq X \leq b) = F(b) - F(a)$.
2. $P(X \geq a) = P(X > a) = 1 - P(X \leq a) = 1 - F(a)$.
3. If X is a continuous random variable with cdf $F(x)$ then at every x at which $F'(x)$ exists:

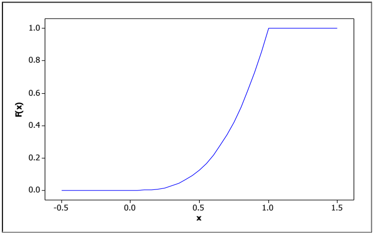
$$f(x) = F'(x) = \frac{dF(x)}{dx}$$

Example

Refer to the previous example, obtain and graph the cdf of X. Use $F(x)$ to compute $P(0.5 \le X \le 0.8)$.

$$F(x) = \int_0^x 3t^2 dt = t^3 \Big|_0^x = x^3 \implies F(x) = \begin{cases} 0 & x < 0 \\ x^3 & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

The function is shown graphically below:



$$P(0.5 \le X \le 0.8) = F(0.8) - F(0.5) = 0.8^3 - 0.5^3 = 0.387$$

Example

If X is continuous with distribution function F_X and density function f_X, find the density function of $Y = 2X$.

Solution: By writing first the definition of the cdf of Y, we obtain

$$\begin{aligned} F_Y(a) &= P(Y \leq a) \\ &= P(X \leq a/2) \\ &= F_X(a/2) \end{aligned}$$

Differentiating F_Y over a: $f_Y(a) = 1/2 f_X(a/2)$.

Percentiles

Definition (Percentiles)

Let p be a number between 0 and 1. The $(100p)^{th}$ percentile of the distribution of a continuous random variable X, we shall denote by c, is that value for which

$$p = F(c) = \int_{-\infty}^{c} f(x)dx$$

i.e.

$$c = F^{-1}(p)$$

where $F^{-1}(.)$ is the inverse cumulative distribution function.

Special percentiles:

- The median of a continuous distribution, denoted by m, is the 50^{th} percentile. So m satisfies $0.5 = F(m)$ or $m = F^{-1}(0.5)$.
- The first and the third quartiles can be computed as $Q_1 = F^{-1}(0.25)$ and $Q_3 = F^{-1}(0.75)$, respectively.

Example

Refer to the previous example, find the median and the interquartile range of the distribution.
The cdf of X in the previous example is given by

$$F(x) = \begin{cases} 0 & x < 0 \\ x^3 & 0 \leq x < 1 \\ 1 & x \geq 1 \end{cases}$$

- The inverse cumulative distribution function is given by

$$x = F^{-1}(p) = \sqrt[3]{p}$$

- The median of the distribution is $m = \sqrt[3]{0.5} = 0.794$.
- Similarly, the quartiles are

$$Q_1 = \sqrt[3]{0.25} = 0.630 \qquad \text{and} \qquad Q_3 = \sqrt[3]{0.75} = 0.909$$

- Thus, $IQR = Q_3 - Q_1 = 0.909 - 0.630 = 0.279$

Definition (The The Expected Value (Mean))

If X is a random variable with pdf $f(X)$, then the expected value (the mean) of X denoted by $E(X)$ or μ is given by

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

The Expected Value of a Function

Let $h(x)$ be any function and let Y denote the random variable $Y = h(X)$, then the expected value of $Y = h(X)$ is computed by

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

The Variance of a Continuous Random Variable

The Variance of a Continuous Random Variable

The variance of X, denoted by $V(X)$ or σ^2 is given by

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

Expectation of Linear Functions

If the function $h(x)$ is linear, i.e. $y = h(x) = ax + b$, where a is a multiplicative constant and b is an additive constant, then

1. $E(Y) = E(aX + b) = aE(X) + b$
2. $V(Y) = V(aX + b) = a^2 V(X)$
3. $\sigma_Y = |a|\sigma_X$

Example

Refer to the previous example, find the expected value and variance of percentage of time the technician spends serving customers.

- The expected value of X is

$$E(X) = \mu = \int_0^1 (x)(3x^2)dx = \frac{3}{4}x^4 \Big|_0^1 = 0.75$$

- To compute the variance we need to find $E(X^2)$ first

$$E(X^2) = \int_0^1 (x^2)(3x^2)dx = \frac{3}{5}x^5 \Big|_0^1 = 0.60$$

Hence, the variance of X is

$$V(X) = E(X^2) - \mu^2 = 0.6 - 0.75^2 = 0.0375$$

Examples

(1) Find $E[X]$ and $V(X)$ when the density function of X is

$$f(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & \text{Otherwise} \end{cases}$$

Solution: $E[X] = \int_0^1 x f(x) dx = \int_0^1 2x^2 dx = 2/3$.
$E[X^2] = \int_0^1 x^2 f(x) dx = \int_0^1 2x^3 dx = 1/2$ and then
$V(X) = 1/2 - 4/9 = 10/9$.

(2) The density function of X is given by

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{Otherwise} \end{cases}$$

Find $E[e^X]$.

Solution: $E[e^X] = \int_0^1 e^x dx = e - 1$.

Exercises

1. Let X denote the resistance of a randomly chosen resistor, and suppose that its pdf is given by

$$f(x) = \begin{cases} \frac{x}{18} & \text{if } 8 \le x \le 10 \\ 0 & \text{otherwise} \end{cases}$$

   (a) Find and graph the cdf of X.
   (b) Find $P(8.6 \le X \le 9.8)$.
   (c) Find the median of the resistance of such resistors.
   (d) Find the mean and variance of X.

2. The length of time to failure (in hundreds of hours) for a transistor is a random variable X with distribution function given by

$$F(x) = 1 - e^{-x^2} \qquad \text{for } x \ge 0$$

   (a) Find $f(x)$.
   (b) Find the probability that the transistor operates for at least 200 hours.
   (c) Find the 30^{th} percentile of X.

3. Weekly CPU time used by an accounting firm has probability density function (measured in hours) given by

$$f(x) = \begin{cases} \frac{3}{64}x^2(4-x) & 0 \le x \le 4 \\ 0 & \text{Otherwise} \end{cases}$$

(a) Find the $F(x)$ for weekly CPU time.
(b) Find the probability that the of weekly CPU time will exceed two hours for a selected week.
(c) Find the expected value and variance of weekly CPU time.
(d) Find the probability that the of weekly CPU time will be within half an hour of the expected weekly CPU time.
(e) The CPU time costs the firm $200 per hour. Find the expected value and variance of the weekly cost for CPU time.

4. The length of time required by students to complete a one-hour exam is a random variable with a density function given by

$$f(y) = \begin{cases} cy^2 + y & 0 \leq y \leq 1 \\ 0 & \text{Otherwise} \end{cases}$$

(a) Find c that makes this function a valid probability density function.
(b) Find $F(y)$.
(c) Graph $f(y)$ and $F(y)$.
(d) Find the probability that a randomly selected student will finish in less than half an hour.
(e) Find the time that 95% of the students finish before it.
(f) Given that a particular student needs at least 15 minutes to complete the exam, find the probability that she will require at least 30 minutes to finish.

Uniform Distribution $U(a, b)$

- The uniform random variable is used to model the behavior of a continuous random variable whose values are uniformly or evenly distributed over a given interval.

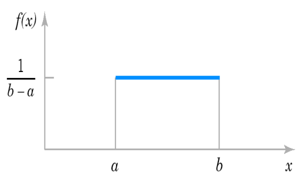
Uniform Distribution

- A random variable X is said to be uniformly distributed over the interval $[a, b]$, denoted by $X \sim U(a, b)$, if its density function is

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{Otherwise} \end{cases}$$

- If $X \sim U(a, b)$, then:

$$E(X) = \frac{a+b}{2}$$
$$V(X) = \frac{(b-a)^2}{12}$$

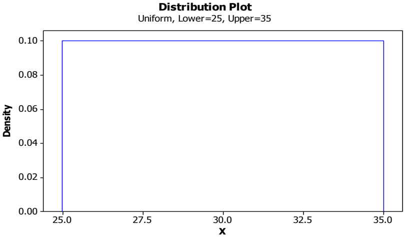


Example

The time (in min) for a lab assistant to prepare the equipment for a certain experiment is believed to have a uniform distribution with $a = 25$ and $b = 35$.

(a) Write the pdf of X and sketch its graph.

$$f(x) = \begin{cases} \frac{1}{10} & 25 \leq x \leq 35 \\ 0 & \text{Otherwise} \end{cases}$$



Example

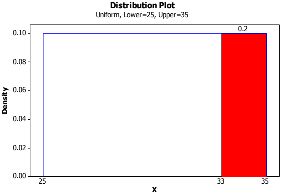
(b) What is the probability that preparation time exceeds 33 min?

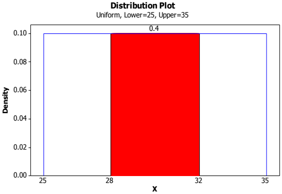
$$P(X > 33) = \frac{35 - 33}{10} = 0.2$$

(c) What is the probability that preparation time is within 2 min of the mean time?

$$E(X) = \frac{25 + 35}{2} = 30$$

$$P(30 - 2 < X < 30 + 2) = P(28 < X < 32) = \frac{32 - 28}{10} = 0.40$$





Exercises

1. Upon studying low bids for shipping contracts, a microcomputer company finds that intrastate contracts have low bids that are uniformly distributed between 20 and 25, in units of thousands of dollars.
   - (a) Find the probability that the low bid on the next intrastate shipping contract is below $22,000.
   - (b) Find the probability that the low bid on the next intrastate shipping contract is in excess of $24,000.
   - (c) Find the expected value and standard deviation of low bids on contracts of the type described above.
2. A grocery store receives delivery each morning at a time that varies uniformly between 5:00 and 7:00 AM.
   - (a) Write and sketch the pdf of the delivery arrival.
   - (b) Find the probability that the delivery on a given morning will occur between 5:30 and 5:45 A.M.
   - (c) Find the probability that the time of delivery will be within one-half standard deviation of the expected time.

Normal Distribution

The normal distribution is one of the most commonly used probability distribution for applications:

1. When we repeat an experiment numerous times and average our results, the random variable representing the average or total tends to have a normal distribution as the number of experiments becomes large.
2. The previous fact, which is known as the central limit theorem, is fundamental to many of the statistical techniques we will discuss later.
3. Many physical characteristics (Heights, weights, etc.) tend to follow a normal distribution.
4. Errors in measurement or production processes can often be approximated by a normal distribution.
5. Under certain conditions, many probability distributions can be approximated by a normal distribution.

Normal Distribution $\mathcal{N}(\mu, \sigma^2)$

- The Normal Distribution denoted by $\mathcal{N}(\mu, \sigma^2)$ has two parameters associated with it; the mean μ and the standard deviation σ.

Normal Distribution

- A random variable X is said to be normally distributed if its density function is

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

- If $X \sim \mathcal{N}(\mu, \sigma^2)$, then:

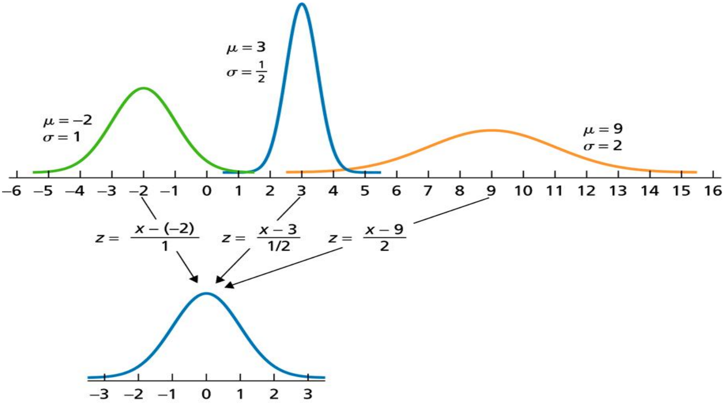
$$E(X) = \mu \qquad \text{and} \qquad V(X) = \sigma^2$$

- The standard normal distribution, $\mathcal{N}(0, 1)$ is defined by

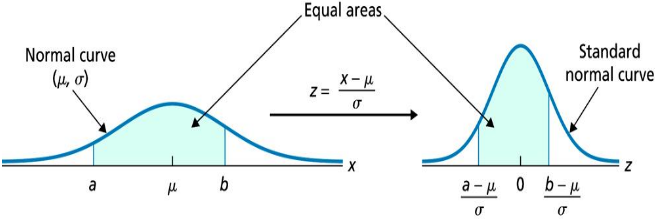
$$Z = \frac{X - \mu}{\sigma}$$

Properties of Normal Distribution

- The normal distribution has a bell-shaped curve:
  - measurements concentrate near the mean μ.
  - symmetric distribution (Mean=Median).
  - the variance σ^2 summarizes the variability - larger variance $\Longrightarrow$ measurements are more variable



- The cdf of the normal distribution, $F(x)$, does not have a closed form expression and the corresponding integral is hard to evaluate. The same is true for the cdf of the standard normal distribution (commonly denoted by $\Phi(z)$). However, $\Phi(z)$ is usually tabulated for values of z from -3.49 to 3.49 in increments of 0.01 and can be used to calculate any normal probability by standardizing it first.



Example

The times of first failure of a unit of a brand of ink jet printers are approximately normally distributed with a mean of 1,500 hours and a standard deviation of 200 hours. Use the statistical calculator.

(a) What fraction of these printers will fail before 1,000 hours?
The failure time (X) is $\mathcal{N}(1500, 200^2)$.

$$\begin{aligned} P(X < 1000) &= P\left(Z < \frac{1000 - 1500}{200}\right) = P(Z < -2.5) \\ &= \Phi(-2.5) = 0.0062 \end{aligned}$$

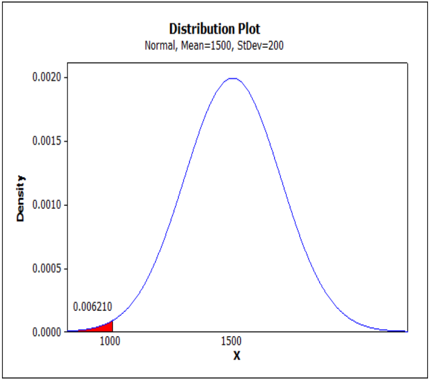
Example

(b) What is the probability that the first failure time of a selected printer will fail be between 1,300 and 1700 hours?

$$\begin{aligned} P(1300 < X < 1700) &= P\left(\frac{1300 - 1500}{200} < Z < \frac{1700 - 1500}{200}\right) \\ &= P(-1 < Z < 1) = \Phi(1) - \Phi(-1) \\ &= 0.8413 - 0.1587 = 0.6826 \end{aligned}$$

$$P(1300 < X < 1700) = F(1700) - F(1300) = 0.6826$$

Example



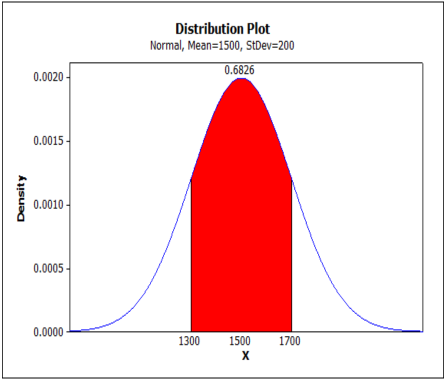
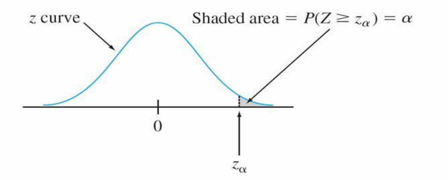


Figure: $P(X < 1000)$ and $P(1300 < X < 1700)$

- We could find the observed value (x) of a given proportion or percentile in $\mathcal{N}(\mu, \sigma^2)$ by unstandardizing the z-value as follows:
  1. Find the z-value corresponding to the lower tail probability using Φ^{-1}.
  2. Unstandardize $x = \mu + z\sigma$
- In the standard normal distribution, z_α will denote the z-value for which of the area under the standard normal curve lies to the right of z_α, i.e.

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

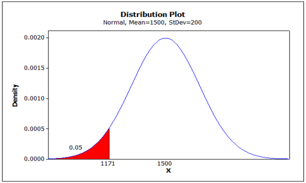


Example

In the previous example, what should be the guarantee time for these printers if the manufacturer wants only 5% to fail within the guarantee period?

Recall that the failure time (X) is $\mathcal{N}(1500, 200^2)$. Now we want to find a such that $P(X < a) = 0.05$, so

$$P(X < a) = 0.05 \ \Rightarrow\ F(a) = 0.05 \ \Rightarrow\ a = F^{-1}(0.05) = 1171 \text{ years}$$



Exercises

1. An engineer working for a manufacturer of electronic components takes a large number of measurements of a particular dimension of components from the production line. She finds that the distribution of dimensions is normal, with a mean of 2.340 cm and a standard deviation of 0.06 cm.
   - (a) What percentage of measurements will be less than 2.45 cm?
   - (b) What percentage of dimensions will be between 2.25 cm and 2.45 cm?
   - (c) What value of the dimension will be exceeded by 98% of the components?
2. Wires manufactured for use in a computer system are specified to have resistances between 0.12 and 0.14 ohms. The actual measured resistances of the wires produced by company A have a normal probability distribution with mean 0.13 ohm and standard deviation 0.005 ohm.
   - (a) What is the probability that a randomly selected wire from company A's production will meet the specifications?
   - (b) If four of these wires are used in each computer system and all are selected from company A, what is the probability that all four in a randomly selected system will meet the specifications?

3. The SAT and ACT college entrance exams are taken by thousands of students each year. The mathematics portions of each of these exams produce scores that are approximately normally distributed. In recent years, SAT mathematics exam scores have averaged 480 with standard deviation 100. The average and standard deviation for ACT mathematics scores are 18 and 6, respectively.
   - (a) An engineering school sets 550 as the minimum SAT math score for new students. What percentage of students will score below 550 in a typical year?
   - (b) What score should the engineering school set as a comparable standard on the ACT math test?
4. Of the Type A electrical resistors produced by a factory, 85% have resistance greater than 41 ohms, and 3.7% of them have resistance greater than 45 ohms. The resistances follow a normal distribution. What percentage of these resistors have resistance greater than 44 ohms?

Normal Approximation to the Binomial

The normal distribution can be used to approximate binomial probabilities when there is a very large number of trials and when np and $n(1-p)$ are both large ($np \geq 10$ and $n(1-p) \geq 10$). We have the following theorem.

The DeMoivre-Laplace limit theorem

If S_n denotes the number of successes that occur when n independent trials, each resulting in a success with probability p, are performed, then, for any $a < b$,

$$P\left\{a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right\} \rightarrow \Phi(b) - \Phi(a)$$

as $n \rightarrow \infty$

Normal Approximation to the Binomial

Note: Normal Approximation to the Binomial

If X is distributed as Binomial with parameter n and p, then the binomial probability distribution can be approximated by using a normal distribution with $\mu = np$ and $\sigma = \sqrt{np(1-p)}$.

To improve the accuracy of the approximation, we usually use a correction factor to take into account that the binomial random variable is discrete while the normal is continuous (called continuity correction). The basic idea is to treat the discrete value k as the continuous interval from $k - 0.5$ to $k + 0.5$ giving the following adjustments:

- $P(a \leq X \leq b) = P(a - 0.5 \leq X \leq b + 0.5)$
- $P(X = k) = P(k - 0.5 \leq X \leq k + 0.5)$

Example

- A new computer virus attacks a folder consisting of 200 files. Each file gets damaged with probability 0.2 independently of other files. What is the probability that fewer than 50 files get damaged?
- Let X be the number of damaged files, then X is binomial random variable with $n = 200$ and $p = 0.2$. Then $\mu = (200)(0.2) = 40$ and $\sigma = \sqrt{(200)(0.2)(0.8)} = 5.657$. Since $np = (200)(0.2) = 40 > 10$ and $n(1-p) = (200)(0.8) = 160 > 10$, the approximation can safely be used. The probability that fewer than 50 files get damaged is

$$\begin{aligned} P(X < 50) &= P(X \leq 49) \approx P(X \leq 49 + 0.5) \\ &= F(49.5) = 0.9535 \end{aligned}$$

Exercises

1. Assume that 12% of the memory cards made at a certain factory are defective. If a sample of 150 cards is selected randomly, use the normal approximation to the binomial distribution to calculate the probability that the sample contains:
   (a) at most 20 defective cards;
   (b) between 15 and 23 defective cards;
   (c) exactly 17 defective cards.

2. The probability a visitor to the homepage of a Web site views another page on the site is 0.15. Assume that 300 visitors arrive at the home page and that they behave independently. Approximate the probabilities for the following events:
   (a) More than 40 visitors view another page.
   (b) At least 30 visitors view another page.
   (c) Fewer than 20 visitors view another page.

Exponential Distribution

- The exponential distribution is often used to model time (waiting time, interarrival time, hardware lifetime, failure time, etc.).
- When the number of occurrences of an event follows Poisson distribution, the time between occurrences follows exponential distribution.

Exponential Distribution

- The exponential probability distribution with parameter λ (called the failure rate) is

$$f(x) = \lambda e^{-\lambda x}, \qquad x \geq 0$$

- The mean and the variance of the distribution are:

$$\mu = \frac{1}{\lambda} \qquad \text{and} \qquad \sigma^2 = \frac{1}{\lambda^2}$$

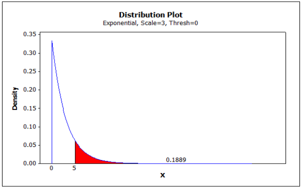
- The cdf of the exponential distribution is

$$F(x) = 1 - e^{-\lambda x}$$

Example

Suppose that a study of a certain computer system reveals that the response time, in seconds, has an exponential distribution with a mean of 3 seconds. What is the probability that response time exceeds 5 seconds?

$$P(X > 5) = 1 - F(5) = 1 - (1 - e^{-\frac{5}{3}}) = 0.1889$$



Exercises

1. The failure rate for a type of electric light bulb is 0.002 per hour. Under the exponential model,
   (a) Find the probability that a randomly selected light bulb will fail in less than 1000 hours.
   (b) Find the probability that a randomly selected light bulb will last 2000 hours before failing.
   (c) Find the mean and the variance of time until failure.
   (d) Find the median time until failure.
   (e) Find the time where 95% of these bulbs are expected to fail before it.

2. An engineer thinks that the best model for time between breakdowns of a generator is the exponential distribution with a mean of 15 days.
   (a) If the generator has just broken down, what is the probability that it will break down in the next 21 days?
   (b) What is the probability that the generator will operate for 30 days without a breakdown?
   (c) If the generator has been operating for the last 20 days, what is the probability that it will operate for another 30 days without a breakdown?
   (d) Comment on the results of parts (b) and (c).

Continuous Probability Distributions: Web Resources

- Empirical Rule
- Normal Distribution
- Normal Distribution Calculator
- Normal Approximation to Binomial Distributions
- Continuous Distributions Probabilities
- Exponential Distribution Calculator
- Interactive Distributions Resources
- Abraham de Moivre
- Carl Friedrich Gauss

Beta Distribution $\mathcal{B}(\alpha, \beta)$

- Probability Density Function:

$$f(y) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} \qquad 0 < y < 1$$

where $\Gamma(\alpha)$ is defined by

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

- Moments:

$$E(Y) = \frac{\alpha}{\alpha+\beta}$$

$$Var(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Gamma Distribution $Gamma(\alpha, \beta)$

- Probability Density Function:

$$f(y) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} y^{\alpha-1} e^{-y/\beta}, \qquad 0 < y < \infty$$

- Moments:

$$E(Y) = \alpha\beta$$
$$Var(Y) = \alpha\beta^2$$
$$M_Y(t) = (1 - \beta t)^{-\alpha}$$

- Relation to Other Distributions
  - Exponential distribution is a special case of Gamma distribution with $\alpha = 1$. That is $Exp(\beta) = \Gamma(1, \beta)$.
  - $\chi^2(\nu) = Gamma(\frac{\nu}{2}, \frac{1}{2})$.
  - If X and Y are independent, $X \sim \Gamma(\alpha_1, \beta)$ and $Y \sim \Gamma(\alpha_2, \beta)$ distributed, then $X/(X + Y) \sim Beta(\alpha_1, \alpha_2)$.

Exercises

- Suppose the time spent by a randomly selected student who uses a terminal connected to a local time-sharing computer facility has a gamma distribution with mean 20 min and variance 80 min^2.
  1. What are the values of α and β?
  2. What is the probability that a student uses the terminal for at most 24 min?
  3. What is the probability that a student spends between 20 and 40 min using the terminal?
- A pumping station operator observes that the demand for water at a certain hour of the day can be modeled as an exponential random variable with a mean of 100 cfs (cubic feet per second).
  1. Find the probability that the demand will exceed 200 cfs on a randomly selected day.
  2. What is the maximum water producing capacity that the station should keep on line for this hour so that the demand will have a probability of only 0.01 of exceeding this production capacity?

  beginitemize

Compute probabilities

For continuous random variables, the syntax is also broken down into two pieces: the root and the prefix. The root determines which random variable that we are talking about, and here are the names of the ones that we have covered so far:

1. unif is uniform
2. exp is exponential
3. norm is normal

The available prefixes are

1. p computes the cumulative distribution
2. d computes pdf or pmf
3. r samples from the rv
4. q quantile function

ggplot for graphs

You need the package ggplot2 first If you have never used ggplot before. The basic idea is that the call of ggplot() sets up the default data that will be plotted and the basic aesthetic mappings. The first argument of ggplot will be a data frame that we wish to visualize, and the second argument will be aes(x = , y =). You will need to tell ggplot which variable you want to be considering as the independent variable (x) and which is the dependent variable (y).

ggplot(dataframe..., aes(x = ..., y = ...)) + Ext

- Ext: $geom\_line(aes(y = .), colour = "."")$ for pdf, cdf
  Example: Plot the pdf and cdf of a normal rv with mean 1 and standard deviation 2.
  To plot more than one graph in the same plot, you must add more than one Ext with different coulours.
- Ext: $geom\_bar(stat = "identity")$ for pmf
  Example: Plot the pmf of a geometric random variable with probability of success 1/3.

Chebyshev inequality

Theorem

If X is a random variable with mean μ and standard deviation σ, then for any positive number k

$$P\left[|X-\mu| < k\sigma\right] \geq 1 - \frac{1}{k^2}$$

Example: Let X be a random variable with density function

$$f(x) = \begin{cases} |x-1| & for\ 0 \leq x \leq 2 \\ 0 & otherwise \end{cases}$$

a. Find the exact value of $P\left[|X-\mu| > 1.5 \times \sigma\right]$.
b. Use Chebyshev's inequality to find a bound for the above probability.

Moment generating function

Definition (MGF)

The moment generating function of a random variable X is given by $M_X(t) = E\left[e^{tX}\right]$, which is:

$$M_X(t) = \sum_i e^{ti} P\left[X = i\right]$$

for discrete random variables and

$$M_X(t) = \int_{-\infty}^{+\infty} e^{tx} f(x)\, dx$$

for continuous random variables.

Example: Let X be a cont. r.v. with a Gamma(1,4) density function. Find the moment generating function of X.

Properties

We define the rth non-centered moment of the r.v. X by

$$\mu'_r = E\left[X^r\right].$$

Theorem

Let X be a r.v. with rth non-centered moment μ'_r and m.g.f. $M_X(t)$, then

$$\mu'_r = M_X^{(r)}(0) = \left.\frac{\mathrm{d}^r M_X(t)}{\mathrm{d}t^r}\right|_{t=0}.$$

Theorem

Let $Y = a + bX$ then

$$M_Y(t) = e^{at} M_X(bt).$$

Example

Example: Suppose X and Y are r.v. with prob. dist.

X	1	2	3	4	5	6
p(x)	0	1/4	1/2	0	0	1/4

Y	1	2	3	4	5	6
p(y)	1/4	0	0	1/2	1/4	0

1. Show that $E(X) = E(Y) = 7/2$ and $Var(X) = Var(Y) = 9/4$.
2. Compare the m.g.f. of X and Y.