Simulation and Modelling

Spring 2023
CS4056

Muhammad Shahid Ashraf

Notes

Statistics

Notes

Continuous Distribution

- For all continuous variables, the probability mass function (pmf) is always equal to zero,1 $P(x) = 0 \quad \forall x$.
- we can use the cumulative distribution function (cdf) $F(x)$

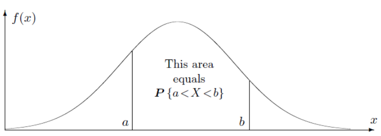
$$F(x) = P\{X \leq x\} = P\{X < x\}$$

- Probability density function (pdf, density) is the derivative of the cdf, $f(x) = F'(x)$. The distribution is called continuous if it has a density.

Notes

Continuous Distributions

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



Notes

Continuous Distributions

- The lifetime, in years, of some electronic component is a continuous random variable with the density $f(x) = \begin{cases} \frac{k}{x^3} & \text{for} \quad x \geq 1 \\ 0 & \text{for} \quad x < 1. \end{cases}$
- Find k, draw a graph of the cdf $F(x)$, and compute the probability for the lifetime to exceed 5 years

Notes

Continuous Distributions

- Find k from the condition $\int f(x)dx = 1$

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{1}^{+\infty} \frac{k}{x^3}dx = -\frac{k}{2x^2}\bigg|_{x=1}^{+\infty} = \frac{k}{2} = 1.$$

$$F(x) = \int_{-\infty}^{x} f(y)dy = \int_{1}^{x} \frac{2}{y^3}dy = -\frac{1}{y^2}\bigg|_{y=1}^{x} = 1 - \frac{1}{x^2}$$

- Hence, $k = 2$. Integrating the density, we get the cdf
- Next, compute the probability for the lifetime to exceed 5 years,

$$P\{X > 5\} = 1 - F(5) = 1 - (1 - \frac{1}{5^2} = 0.04$$

- We can also obtain this probability by integrating the density

$$\boldsymbol{P}\{X > 5\} = \int_{5}^{+\infty} f(x)dx = \int_{5}^{+\infty} \frac{2}{x^3}dx = -\frac{1}{x^2}\bigg|_{x=5}^{+\infty} = \frac{1}{25} = 0.04.$$

Notes

Continuous Distributions

Distribution	Discrete	Continuous
Definition	$P(x) = P\{X = x\}$ (pmf)	$f(x) = F'(x)$ (pdf)
Computing probabilities	$\boldsymbol{P}\{X \in A\} = \sum_{x \in A} P(x)$	$\boldsymbol{P}\{X \in A\} = \int_A f(x)dx$
Cumulative distribution function	$F(x) = \boldsymbol{P}\{X \leq x\} = \sum_{y \leq x} P(y)$	$F(x) = \boldsymbol{P}\{X \leq x\} = \int_{-\infty}^{x} f(y)dy$
Total probability	$\sum_x P(x) = 1$	$\int_{-\infty}^{\infty} f(x)dx = 1$

Notes

Moments for discrete and continuous distributions

Discrete	Continuous
$\mathbf{E}(X) = \sum_x xP(x)$	$\mathbf{E}(X) = \int xf(x)dx$
$\mathrm{Var}(X) = \mathbf{E}(X-\mu)^2$ $= \sum_x (x-\mu)^2 P(x)$ $= \sum_x x^2 P(x) - \mu^2$	$\mathrm{Var}(X) = \mathbf{E}(X-\mu)^2$ $= \int (x-\mu)^2 f(x)dx$ $= \int x^2 f(x)dx - \mu^2$
$\mathrm{Cov}(X,Y) = \mathbf{E}(X-\mu_X)(Y-\mu_Y)$ $= \sum_x \sum_y (x-\mu_X)(y-\mu_Y)P(x,y)$ $= \sum_x \sum_y (xy)P(x,y) - \mu_x \mu_y$	$\mathrm{Cov}(X,Y) = \mathbf{E}(X-\mu_X)(Y-\mu_Y)$ $= \iint (x-\mu_X)(y-\mu_Y)f(x,y)\,dx\,dy$ $= \iint (xy)f(x,y)\,dx\,dy - \mu_x\mu_y$

$$f(x) = 2x^{-3} \text{ for } x \geq 1.$$

Its expectation equals

$$\mu = \mathbf{E}(X) = \int x\,f(x)dx = \int_{1}^{\infty} 2x^{-2}dx = -2x^{-1}\big|_{1}^{\infty} = 2.$$

Notes

Uniform Distributions

- a random variable with any thinkable distribution can be generated from a Uniform random variable
- Uniform distribution is used in any situation when a value is picked "at random" from a given interval; that is, without any preference to lower, higher, or medium values.
- Whenever the probability is proportional to the length of the interval, the random variable is uniformly distributed

Notes

Uniform Distributions

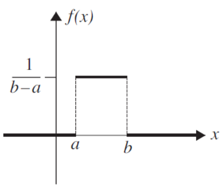
- Suppose that X is the value of the random point selected from an interval (a, b). Then X is called a uniform random variable over (a, b). Let F and f be distribution and probability density functions of X, respectively

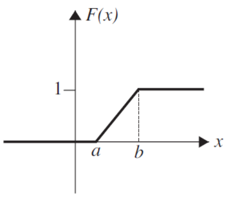
$$F(t) = \begin{cases} 0 & t \leq a \\ \dfrac{t-a}{b-a} & a \leq t \leq b \\ 1 & t \geq b \end{cases}$$

$$f(t) = F'(t) = \begin{cases} \dfrac{1}{b-a} & a < t < b \\ 0 & \text{otherwise} \end{cases}$$

Notes

Uniform Distributions





Notes

Uniform Distributions

- In random selections of a large number of points from (a, b), we expect that the average of the values of the points will be approximately $\frac{a+b}{2}$, the midpoint of (a, b).

$$E(X) = \int_a^b x \frac{1}{b-a} dx$$

$$= \frac{1}{b-a}\left[\frac{1}{2}x^2\right]_a^b$$

$$= \frac{1}{b-a}\left(\frac{1}{2}b^2 - \frac{1}{2}a^2\right)$$

$$= \frac{(b-a)(b+a)}{2(b-a)}$$

$$= \frac{a+b}{2}.$$

Notes

Uniform Distributions

- To find $Var(X)$, we have

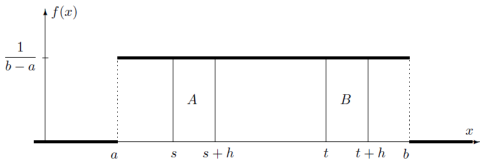
$$E(X^2) = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{3}\frac{b^3-a^3}{b-a} = \frac{1}{3}(a^2+ab+b^2)$$

Hence

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{1}{3}(a^2+ab+b^2) - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$$

Notes

Uniform Distributions



$$\begin{aligned}
(a,b) &= \text{range of values}\\
f(x) &= \frac{1}{b-a}, \quad a<x<b\\
\mathbf{E}(X) &= \frac{a+b}{2}\\
\text{Var}(X) &= \frac{(b-a)^2}{12}
\end{aligned}$$

Notes

Uniform Distributions

- Starting at 5:00 A.M., every half hour there is a flight from Karachi International airport to Islamabad International airport. Suppose that none of these planes is completely sold out and that they always have room for passengers. A person who wants to fly to Islamabad, arrives at the airport at a random time between 8:45 A.M. and 9:45 A.M. Find the probability that he waits
  1. at most 10 minutes
  2. at least 15 minutes

Notes

Uniform Distributions

- Starting at 5:00 A.M., every half hour there is a flight from Karachi International airport to Islamabad International airport. Suppose that none of these planes is completely sold out and that they always have room for passengers. A person who wants to fly to Islamabad, arrives at the airport at a random time between 8:45 A.M. and 9:45 A.M. Find the probability that he waits
- at most 10 minutes
- The passenger arrive at the airport X minutes past 8:45.
- Then X is a uniform random variable over the interval $(0, 60)$.
- Hence the probability density function of X is given by f

$$f(x = \begin{cases} \frac{1}{60} & 0<x<60 \\ 0 & \text{otherwise} \end{cases}$$

Notes

Uniform Distributions

- Starting at 5:00 A.M., every half hour there is a flight from Karachi International airport to Islamabad International airport. Suppose that none of these planes is completely sold out and that they always have room for passengers. A person who wants to fly to Islamabad, arrives at the airport at a random time between 8:45 A.M. and 9:45 A.M. Find the probability that he waits
- at most 10 minutes
- Now the passenger waits at most 10 minutes if she arrives between 8:50 and 9:00 or 9:20 and 9:30; that is, if $5 < X < 15 \text{ or } 35 < X < 45$

$$P(5 < X < 15) + P(35 < X < 45) = \int_{5}^{15} \frac{1}{60} + \int_{35}^{45} \frac{1}{60}$$

Notes

Uniform Distributions

- A person arrives at a bus station every day at 7:00 A.M. If a bus arrives at a random time between 7:00 A.M. and 7:30 A.M., what is the average time spent waiting
- It takes a professor a random time between 20 and 27 minutes to walk from his home to school every day. If he has a class at 9:00 A.M. and he leaves home at 8:37 A.M. , find the probability that he reaches his class on time.
- The time at which a bus arrives at a station is uniform over an interval (a, b) with mean 2:00 P.M. and standard deviation $\sqrt{12}$ minutes. Determine the values of a and b

Notes

Bernoulli and Binomial distribution

A random variable with two possible values, 0 and 1, is called a **Bernoulli variable**, its distribution is **Bernoulli distribution**, and any experiment with a *binary outcome* is called a **Bernoulli trial**.

Bernoulli distribution

$$\boxed{\begin{array}{lcl} p & = & \text{probability of success} \\ P(x) & = & \begin{cases} q = 1 - p & \text{if} \quad x = 0 \\ p & \text{if} \quad x = 1 \end{cases} \\ \mathbf{E}(X) & = & p \\ \mathrm{Var}(X) & = & pq \end{array}}$$

A variable described as the number of successes in a sequence of independent Bernoulli trials has **Binomial distribution**. Its parameters are n, the number of trials, and p, the probability of success.

Binomial distribution

$$\boxed{\begin{array}{lcl} n & = & \text{number of trials} \\ p & = & \text{probability of success} \\ P(x) & = & \binom{n}{x} p^x q^{n-x} \\ \mathbf{E}(X) & = & np \\ \mathrm{Var}(X) & = & npq \end{array}}$$

Notes

The Binomial Experiment:

An experiment that satisfies the following four conditions is called a binomial experiment.

a. There are n identical trials. In other words, the given experiment is repeated n times, where n is a positive integer. All of these repetitions are performed under identical conditions.
b. Each trial has two and only two outcomes. These outcomes are usually called a success and a failure, respectively. In case there are more than two outcomes for an experiment, we can combine outcomes into two events and then apply binomial probability distribution.
c. The probability of success is denoted by p and that of failure by q, and p + q= 1. The probabilities p and q remain constant for each trial.
d. The trials are independent. In other words, the outcome of one trial does not affect the outcome of another trial.

Notes

The Binomial Formula

- The random variable x that represents the number of successes in n trials for a binomial experiment is called a binomial random variable.
- The probability distribution of x in such experiments is called the binomial probability distribution.
- The binomial probability distribution is applied to find the probability of x successes in n trials for a binomial experiment.
- The number of successes x in such an experiment is a discrete random variable.

No. of successes

Combination of x successes from n trials

number of failures

$$P(X=x)={}^{n}C_{x}.p^{x}.(1-p)^{(n-x)}$$

random variable X

probability of success

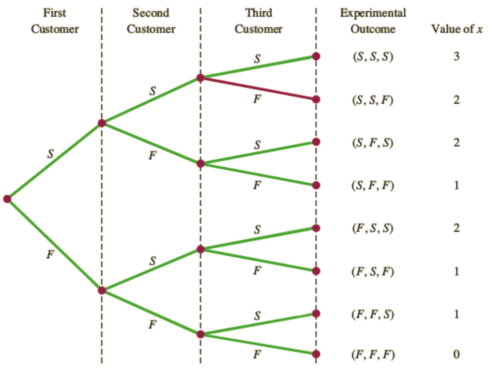
probability of failure

Notes

Example

- Question: Consider the purchase decisions of the next three customers who enter the Clothing Store. On the basis of past experience, the store manager estimates the probability that any one customer will make a purchase is .30. What is the probability that two of the next three customers will make a purchase?
- Let x be the number of Success in a sample of three.
- x can assume any of the values 0, 1, 2, and 3.
- it is a discrete random variable.

Notes



S = Purchase
F = No purchase
x = Number of customers making a purchase

Notes

Trial Outcomes				
1st Customer	2nd Customer	3rd Customer	Experimental Outcome	Probability of Experimental Outcome
Purchase	Purchase	No purchase	(S, S, F)	$pp(1-p)=p^2(1-p)$ $=(.30)^2(.70)=.063$
Purchase	No purchase	Purchase	(S, F, S)	$p(1-p)p=p^2(1-p)$ $=(.30)^2(.70)=.063$
No purchase	Purchase	Purchase	(F, S, S)	$(1-p)pp=p^2(1-p)$ $=(.30)^2(.70)=.063$

x	$f(x)$
0	$\frac{3!}{0!3!}(.30)^0(.70)^3=.343$
1	$\frac{3!}{1!2!}(.30)^1(.70)^2=.441$
2	$\frac{3!}{2!1!}(.30)^2(.70)^1=.189$
3	$\frac{3!}{3!0!}(.30)^3(.70)^0=\underline{.027}$
	1.000

Question: What is the probability of making exactly four sales to 10 customers entering the store.

we have a binomial experiment with n = 10, x = 4, and p =.30

Notes

Example

- Five percent of all DVD players manufactured by a large electronics company are defective. Three DVD players are randomly selected from the production line of this company. What is the probability that exactly one of these three DVD players is defective
- At the Express House Delivery Service, providing high-quality service to customers is the top priority of the management. The company guarantees a refund of all charges if a package it is delivering does not arrive at its destination by the specified time. It is known from past data that despite all efforts, 2% of the packages mailed through this company do not arrive at their destinations within the specified time. Suppose a corporation mails 10 packages through Express House Delivery Service on a certain day.

  Calculating the probability using the binomial formula.
  - Find the probability that exactly one of these 10 packages will not arrive at its destination within the specified time.
  - Find the probability that at most one of these 10 packages will not arrive at its destination within the specified time.
- An exciting computer game is released. Sixty percent of players complete all the levels. Thirty percent of them will then buy an advanced version of the game. Among 15 users, what is the expected number of people who will buy the advanced version? What is the probability that at least two people will buy it?

Notes

The Poisson Probability Distribution:

- The number of rare events occurring within a fixed period of time has Poisson distribution.
- The following examples also qualify for the application of the Poisson probability distribution.
- The number of accidents that occur on a given highway during a 1-week period
- The number of customers entering a grocery store during a 1-hour interval
- The number of television sets sold at a department store during a given week
- The following three conditions must be satisfied to apply the Poisson probability distribution.
  - x is a discrete random variable.
  - The occurrences are random.
  - The occurrences are independent.

Notes

Poisson distribution

$$\begin{aligned} \lambda &= \text{frequency, average number of events} \\ P(x) &= e^{-\lambda}\frac{\lambda^x}{x!},\ x = 0, 1, 2, \ldots \\ \mathbf{E}(X) &= \lambda \\ \mathrm{Var}(X) &= \lambda \end{aligned}$$

- PROPERTIES OF A POISSON EXPERIMENT
- The probability of an occurrence is the same for any two intervals of equal length.
- The occurrence or nonoccurrence in any interval is independent of the occurrence or nonoccurrence in any other interval.

Notes

- On average, a household receives 9.5 telemarketing phone calls per week. Using the Poisson probability distribution formula, find the probability that a randomly selected household receives exactly 6 telemarketing phone calls during a given week.
- A washing machine in a laundromat breaks down an average of three times per month. Using the Poisson probability distribution formula, find the probability that during the next month this machine will have
  - exactly two breakdowns
  - at most one breakdown
- The number of emails that I get in a weekday can be modeled by a Poisson distribution with an average of 0.2 emails per minute.
  - What is the probability that I get no emails in an interval of length 5 minutes?
  - What is the probability that I get more than 3 emails in an interval of length 10 minutes?

Notes

Problems

- A bus arrives every 20 minutes at a specified stop beginning at 6:40 A.M. and continuing until 8:40 A.M. A certain passenger does not know the schedule, but arrives randomly (uniformly distribut- ed) between 7:00 A.M. and 7:30 A.M. every morning. What is the probability that the passenger waits more than 5 minutes for a bus?
- Every day, a lecture may be cancelled due to inclement weather with probability 0.05. Class cancellations on different days are independent. Compute the probability that the tenth class this semester is the 3rd class cancelled?
- What are the expected value and variance of the number of full house hands in n poker hands? A poker hand consists of five randomly selected cards from an ordinary deck of 52 cards. It is a full house if three cards are of one denomination and two cards are of another denomination: for example, three queens and two 4's.

Notes

Problems

- A weather buoy has a service life (in years) that follows this pdf:

$$f(x) = \begin{cases} 0.35e^{-0.35x} & \text{if } x \geq 0 \\ 0 & otherwise \end{cases} \quad (1)$$

  - What is the probability that this buoy is still working after 4 years?
  - What is the probability that the buoy dies between 3 and 6 years from the time it is deployed in the sea?

Notes

Notes

Notes