ISyE 3770
Chapter 3: Discrete Random Variables

Instructor: Dan Li
Slides by Prof. Nagi Gebraeel

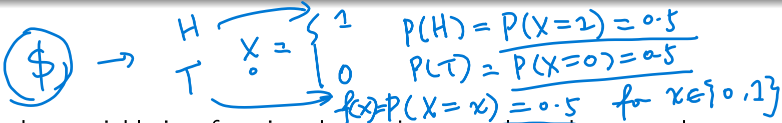
School of Industrial and Systems Engineering
Georgia Tech

Spring 2021

1. Discrete Random Variables
    - pmf
    - cdf
    - Mean & Variance

2. Discrete Distributions
    - Discrete Uniform Distribution
    - Binomial Distribution
    - Geometric Distribution
    - Negative Binomial and Hypergeometric Distributions
    - Poisson Distribution

Recall: Random Variables



- A random variable is a function that assigns a real number to each outcome in the sample space of a random experiment.
- A random variable is denoted by an uppercase letter such as X. After the experiment is conducted, the measured value of the random variable is denoted by a lowercase letter such as $x = 70$ milliamperes. X and x are shown in italics, e.g., $P(X = x)$.
- We often focus on describing the distribution of a particular random variable.

X = outcome of experiment

$x = \{1, 2, \ldots, 6\}$, $x = 1, 2 \ldots$

Probability Distribution

- The probability distribution of the random variable X is a description of the probabilities with the possible numerical values of X.
- A probability distribution of a discrete random variable can be:
  1. A list of the possible values along with their probabilities.
  2. A formula that is used to calculate the probability in response to an input of the random variable's value.

① $P(X=x_1)=0.1$ $P(X=x_2)=0.1$ .... $P(X=x_n)$

$S=\{x_1,\ldots,x_n\}$ $x \in S$

② $P(X=x)=f(x)$ $\forall x \in S$

Example: Probability Distribution

- There is a chance that a bit transmitted through a digital transmission channel is received in error.
- Let X equal the number of bits received in error of the next 4 transmitted. S = {0, 1, 2, 3, 4}
- The associated probability distribution of X is shown as a graph and as a table.

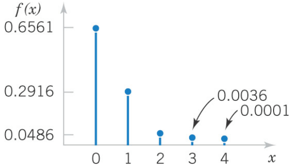


Figure 3-1 Probability distribution for bits in error.

P(X=0) =	0.6561
P(X=1) =	0.2916
P(X=2) =	0.0486
P(X=3) =	0.0036
P(X=4) =	0.0001
	1.0000

Probability Mass Function

- In this section, we focus on the analysis of some popular discrete random variables that frequently arise in applications.
- For a discrete random variable X with possible values $x_1, x_2, \ldots, x_n$, a probability mass function is a function such that has the following properties:
  1. $f(x_i) \geq 0$
  2. $\sum_{i=1}^{n} f(x_i) = 1$
  3. $f(x_i) = P(X = x_i)$

$S = \{x_1, \ldots, x_n\}$.

$\sum f(x_i) = 1 = f(x_1) + \cdots + f(x_n)$

def. of pmf.

Cumulative Distribution

$f(x) = P(X = x)$

cdf $F(x) = P(X \le x)$

- The cumulative distribution function is the probability that a random variable X with a given probability distribution will be found at a value less than or equal to x.
- For a discrete rv X, we define the *cumulative distribution function* (cdf) for a state space $\mathcal{S} = \{x_1, x_2, \ldots, x_n\}$ as follows:
  1. $F_X(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$
  2. $0 \le F_X(x) \le 1$
  3. If $x \le y$, then $F_X(x) \le F_X(y)$

$F(x) = P(X \le x)$

$S\{0, 1, 2, 3\}$

$F(2) = f(0) + f(1) + f(2)$

$F_X(y) = P(X \le y) = P(X \le x) + P(x \le X \le y) \ge P(X \le x) = F_X(x)$, where $P(x \le X \le y) \ge 0$

Example

- A day's production of 850 parts contains 50 defective parts. Two parts are selected at random without replacement.

  Let the random variable X equal the number of defective parts in the sample. Find the cumulative distribution function of X.

$$S = \{0, 1, 2\}$$

$$f(0) = P(X=0) = \frac{800}{850} \times \frac{799}{849} = 0.886$$

$$f(2) = P(X=1) = 1 - f(0) - f(2) = 0.111$$

$$f(2) = P(X=2) = \frac{50}{850} \times \frac{49}{849} = 0.003$$

$$F(0) = P(X \le 0) = f(0) = 0.886$$

$$F(1) = P(X \le 1) = f(0) + f(1) = 0.997$$

$$F(2) = 1$$

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.886 & 0 \le x < 1 \\ 0.997 & 1 \le x < 2 \\ 1 & x \ge 2 \end{cases}$$

Example contd.

Summary Numbers of Probability Distributions

$X = \begin{cases} 0 \\ 1 \end{cases}$ $p(X=0) = 0.1$ $p(X=1) = 0.9$

$\mu \neq 0.5$

$\mu = 0.1 \times 0 + 0.9 \times 1 = 0.9.$

- The **mean** is a measure of the center of a probability distribution. $\mu = E(X)$.
- The **variance** is a measure of the dispersion or variability of a probability distribution. $\sigma^2 = V(X)$.
- The **standard deviation** is another measure of the dispersion. It is the square root of the variance.

$\sigma = \sqrt{V(X)}$.

Mean

- For a discrete rv X with a state space $\mathcal{S} = \{x_1, x_2, \ldots, x_n\}$ and pmf P_X, we define its *expectation* or *mean* to be a weighted average of the values in the state space:

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i P(X = x_i).$$

- The expectation is a measure for the average value in the state space.

Variance

- For a discrete X with a state space $\mathcal{S} = \{x_1, x_2, \ldots, x_n\}$ and pmf P_X, we define its *variance* to be:

$$\mathbb{V}(X) = \sum_{i=1}^{n} (x_i - \mathbb{E}(X))^2 P(X = x_i)$$

$$= \sum_{i=1}^{n} x_i^2 P(X = x_i) - \left(\sum_{i=1}^{n} x_i P(X = x_i) \right)^2$$

- A computationally efficient formula for calculating variance

$$\mathbb{V}(X) = \sum_{i=1}^{n} x_i^2 P(X = x_i) - \mathbb{E}(X)^2$$

Example 3-9

- There is a chance that a bit transmitted through a digital transmission channel is received in error. X is the number of bits received in error of the next 4 transmitted. The probabilities are $P(X = 0) = 0.6561, P(X = 1) = 0.2916, P(X = 2) = 0.0486, P(X = 3) = 0.0036, P(X = 4) = 0.0001,$

$\mu = E(X) = 0$

Example 3-9

x					
0					
1					
2					
3					
4					

Functions of Discrete Random Variables

- If X is a discrete random variable with probability mass function $f(x)$ and $h(X)$ is a function of X
- $\mathbb{E}[h(X)] = \sum_X h(x)f(x)$
- What if $h(x) = (X - \mu)^2$?
- **Example:** X is the number of bits in error in the next four bits transmitted. What is the expected value of the square of the number of bits in error?

Example

Discrete Uniform Distribution

- A random variable X has a discrete uniform distribution if each of the n values in its range, say $x_1, x_2, \ldots, x_n$ has equal probability, $f(x_i) = \frac{1}{n}$
- **Example:** The first digit of a part's serial number is equally likely to be the digits 0 through 9. If one part is selected from a large batch, and X is the 1^{st} digit of the serial number, then X has a discrete uniform distribution of the following form.

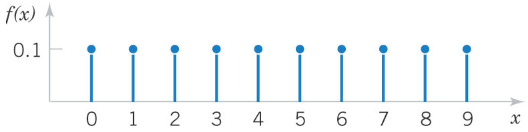


Figure 3-7 Probability mass function, $f(x) = 1/10$ for $x = 0, 1, 2, \ldots, 9$

Discrete Uniform Distribution

- Let X be a discrete uniform random variable from a to b for $a < b$. There are $b - a + 1$ values in the inclusive interval. Therefore:

$$f(x) = \frac{1}{b - a + 1}$$

- The mean and variance of this distribution are
  - $\mu = \mathbb{E}[X] = (b + a)/2$
  - $\sigma^2 = V(X) = \frac{(b-a+1)^2-1}{12}$

Example of Discrete Uniform Distribution

- Let the random variable X denote the number of the 48 voice lines that are in use at a particular time. Assume that X is a discrete uniform random variable with a range of 0 to 48. Find $\mathbb{E}[X]$ & $\sigma(X)$.

Binomial Distribution

- A Binomial distributions results from experiments that involve the following:
  - Fixed number of trials n.
  - Each trial is termed a success or failure. X is the number of successes.
  - The probability of success in each trial is constant p.
  - The outcomes of successive trials are independent.
- Examples of binomial random variables are
  - Flip a coin 10 times. $X =$ number heads obtained.
  - A multiple-choice test contains 10 questions, each with 4 choices, and you guess. $X =$ number of correct answers.

Binomial Distribution slide 36

- The random variable X that equals the number of trials that result in a success is a binomial random variable with parameters $0 < p < 1$ and $n = 0, 1, \ldots$
- The probability mass function is:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

- If X is a binomial random variable with parameters p and n,

$$\mu = \mathbb{E}[X] = np$$

$$\sigma^2 = V(X) = np(1-p)$$

Example of Binomial Distribution

- Each sample of water has a 10% chance of containing a particular organic pollutant. Assume that the samples are independent with regard to the presence of the pollutant. Find the probability that, in the next 18 samples, exactly 2 contain the pollutant.
- Determine the probability that at least 4 samples contain the pollutant.
- Now determine the probability that $3 \leq X \leq 7$

Example of Binomial Distribution

Example of Binomial Distribution

Example of Binomial Distribution

- For the number of transmitted bit received in error, $n = 4$ and $p = 0.1$. Find the mean and variance of the binomial random variable.

Geometric Distribution

- Similar to the binomial distribution–a series of Bernoulli trials with fixed parameter p.
- Whereas the Binomial distribution has a (1) Fixed number of trials, and (2) Random number of successes;
- The Geometric distribution has reversed roles:
  - Random number of trials.
  - Fixed number of successes, in this case 1.
- The pmf of a geometric random variable is

$$f(x) = p(1-p)^{x-1}$$

where $x = 1, 2, \ldots \infty$, the number of failures until the 1st success and $0 < p < 1$, the probability of success.

Example of Geometric Distribution

- The probability that a wafer contains a large particle of contamination is 0.01. Assume that the wafers are independent. What is the probability that exactly 125 wafers need to be analyzed before a particle is detected?

Geometric Distribution

- If X is a geometric random variable with parameter p,

$$\mu = \mathbb{E}[X] = \frac{1}{p} \qquad \text{and} \qquad \sigma^2 = V(X) = \frac{1-p}{p}$$

- Consider the transmission of bits in the previous example. Here, $p = 0.1$. Find the mean and standard deviation.

A Note on Geometric Random Variables

- For a geometric random variable, the trials are independent. Thus the count of the number of trials until the next success can be started at any trial without changing the probability.
- This is called the Lack of Memory Property

Negative Binomial Distribution

- In a series of independent trials with constant probability of success, let the random variable X denote the number of trials until r successes occur. Then X is a negative binomial random variable with parameters $0 < p < 1$ and $r = 1, 2, 3, \ldots$
- The probability mass function is

$$f(x) = C_{r-1}^{x-1} p^r (1-p)^{x-r}$$

for $x = r, (r+1), (r+2), \ldots$

Negative Binomial Distribution

- Let X_1 denote the number of trials to the 1st success.
- Let X_2 denote the number of trials to the 2nd success, since the 1st success.
- Let X_3 denote the number of trials to the 3rd success, since the 2nd success.

- Let the X_i be geometric random variables (independent), so without memory.
- Then $X = X_1 + X_2 + X_3$
- Therefore, X is a negative binomial random variable, a sum of three geometric rv's.

Negative Binomial Distribution

- If X is a negative binomial random variable with parameters p and r,

$$\mu = \mathbb{E}[X] = \frac{r}{p} \qquad \text{and} \qquad \sigma^2 = V(X) = \frac{r(1-p)}{p}$$

Example: Negative Binomial Distribution

- A Web site contains 3 identical computer servers. Only one is used to operate the site, and the other 2 are spares that can be activated in case the primary system fails. The probability of a failure in the primary computer (or any activated spare) from a request for service is 0.0005. Assume that each request represents an independent trial. What is the probability that all 3 servers fail within 5 requests?

Hypergeometric Distribution

- Applies to sampling without replacement, that is, trials are not independent & a tree diagram used.
- A set of N objects contains:
  - K objects classified as success
  - $N - K$ objects classified as failures
- A sample of size n objects is selected without replacement from the N objects, where:
  - $K \leq N$ and $n \leq N$ objects classified as success

Hypergeometric Distribution

- Let the random variable X denote the number of successes in the sample. Then X is a hypergeometric random variable.

$$f(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

- where $x = \max(0, n + K - N)$ to $\min(K, n)$

Example of Hypergeometric Random Variable

- A batch of parts contains 100 parts from supplier A and 200 parts from Supplier B. If 4 parts are selected randomly, without replacement,
  - what is the probability that they are all from Supplier A?
  - What is the probability that two or more parts are from Supplier A?
  - What is the probability that at least one part is from Supplier A?

Example of Hypergeometric Random Variable

Example of Hypergeometric Random Variable

Hypergeometric Mean and Variance

- If X is a hypergeometric random variable with parameters N, K, and n, then

$$\mu = \mathbb{E}[X] = np \qquad \text{and} \qquad \sigma^2 = V(X) = np(1-p)\left(\frac{N-n}{N-1}\right)$$

where $p = K/N$

Poisson Distribution

- In general, the Poisson random variable X is the number of events (counts) per interval.
- As the number of trials n in a binomial experiment increases to infinity while the binomial mean np remains constant, the binomial distribution becomes the Poisson distribution.
- The random variable X that equals the number of events in a Poisson process is a Poisson random variable with parameter $\lambda > 0$, and the probability mass function is:

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad \text{for} \qquad x = 1, 2, 3, \ldots$$

Poisson Distribution

- It is important to use consistent units in the calculation of Poisson:
  - Probabilities
  - Means
  - Variances
- If X is a Poisson random variable with parameter λ, then:

$$\mu = \mathbb{E}[X] = \lambda \quad \text{and} \quad \sigma^2 = V(X) = \lambda$$

- The mean and variance of the Poisson model are the same. If the mean and variance of a data set are not about the same, then the Poisson model would not be a good representation of that set.

Poisson Distribution

- Contamination is a problem in the manufacture of optical storage disks (CDs). The number of particles of contamination that occur on a CD has a Poisson distribution. The average number of particles per square cm of media is 0.1. The area of a disk under study is 100 cm2. Let X denote the number of particles of a disk.
    - Find $P(X = 12)$
    - Determine the probability that 12 or fewer particles occur on the disk. That will require 13 terms in the sum of probabilities.