Section 1

MANE 3332.03

Subsection 1

Lecture 7, February 13

Agenda

- Start Chapter 3 lecture
- **New: CDF Practice Problems (assigned 2/13/2025, due 2/18/2025 11:59pm)**

Handouts

- Chapter 3 Slides
- Chapter 3 Slides marked

Random Variable

- A **random variable** is a function that assigns a number real number to each outcome in the sample space of a random experiment.
- A **discrete** random variable is a random variable with a finite or (countably infinite) range.
    - Examples include number of scratches on a surface, proportion of defective parts among 1000 tested, number of transmitted bits received in error
- A **continuous** random variable is a random variable with an interval of real numbers for its range.
    - Examples include electrical current, length, pressure, temperature, time voltage, weight

Definitions

There are three terms commonly used in describing the mathematical relationship between events and probabilities for discrete random variables

Probability distribution

of a random variable is a description of the probabilities associated with the possible values of X

Probability mass function

for a random variable X with possible values $x_1, x_2, \ldots, x_n$ is

$$f(x_i) = P(X = x_i)$$

Cumulative distribution function

of a random variable X is

$$F(x) = P(X \le x) = \sum f(x_i)$$

Probability Distributions

Can be described in three different ways:

1. Graphically using a histogram,
2. in a tabular manner, see problem 3.1.13 on page p-15 or,
3. using a mathematical function (PMF), see problem 3.1.11 on page p-15.

Probability Mass Functions

A PMF for a discrete random variable X with possible values of $x_1, x_2, \ldots, x_n$ is function with the following properties:

- $f(x_i) \geq 0$
- $\sum_{i=1}^{n} f(x_i) = 1$
- $f(x_i) = P(X = x_i)$

Cumulative Distribution Function

There are three special properties that a function must satisfy to be a cumulative distribution function (CDF):

1. $F(x) = P\left(X \leq x\right) = \sum_{x_i \leq x} f(x_i)$
2. $0 \leq F(x) \leq 1$
3. If $x \leq y$, then $F(x) \leq F(y)$

Using a CDF

- Knowledge of the CDF can simplify calculating probabilities
- Example consider a sample of 20 items and we count the number of defects, X
  - Find $P(X > 8)$

$$\begin{aligned} P(X > 8) = & \sum_{i=9}^{20} P(X = i) \\ = & F(20) - F(8) \end{aligned}$$

This can also be written another way

$$\begin{aligned} P(X > 8) = & 1 - P(X \leq 8) \\ = & 1 - F(8) \end{aligned}$$

- Care must be taken when using CDF regarding less than or less than or

CDF Practice Problems

Mean and Variance of a Discrete Random Variable

- The mean or expected value of a random variable (denoted $E(X)$) is

$$\mu = E(X) = \sum_{i=1}^{n} x_i f(x_i)$$

- The variance of X is

$$\sigma^2 = V(X) = E(X-\mu)^2 = \sum_{i=1}^{n} (x_i - \mu)^2 f(x_i) = \sum_{i=1}^{n} x_i^2 f(x_i) - \mu^2$$

- The standard deviation of X is

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

- Fortunately, we won't often use these formulas. Distributions will have

Bernoulli Distribution

The Bernoulli distribution is one of the simplest statistical distributions.

- The Bernoulli distribution is a random variable that can take only two values
- Usually the events are labelled 0 and 1
- The distribution is defined by a single parameter p $(0 \leq p \leq 1)$, takes the values 0 and 1 with $P(X = 0) = 1 - p$ and $P(X = 1) = p$
- The mean is

$$\mu = E(X) = p$$

- The standard deviation is

$$\sigma = \sqrt{p(1-p)}$$

Summary of Common Probability Distributions (Discrete)

A-4 APPENDIX A Statistical Tables and Charts

TABLE I Summary of Common Probability Distributions

Name	Probability Distribution	Mean	Variance	Section in Book
Discrete				
Uniform	$\frac{1}{n}, a \le b$	$\frac{(b+a)}{2}$	$\frac{(b-a+1)^2-1}{12}$	3-5
Binomial	$\binom{n}{x} p^x(1-p)^{n-x}$ $x = 0, 1, \ldots, n, 0 \le p \le 1$	np	$np(1-p)$	3-6
Geometric	$(1-p)^{x-1}p$ $x = 1, 2, \ldots, 0 \le p \le 1$	$1/p$	$(1-p)/p^2$	3-7
Negative binomial	$\binom{x-1}{r-1}(1-p)^{x-r}p^r$ $x = r, r+1, r+2, \ldots, 0 \le p \le 1$	r/p	$r(1-p)/p^2$	3-7
Hypergeometric	$\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$ $x = \max(0, n-N+K), 1, \ldots$ $\min(K, n), K \le N, n \le N$	np where $p = \frac{K}{N}$	$np(1-p)\left(\frac{N-n}{N-1}\right)$	3-8

Discrete Uniform Distribution

- A random variable X is a discrete uniform rv if each of the n values in its range, $x_1, x_2, \ldots, x_n$ has equal probability
- The PMF of a discrete uniform is defined to be

$$f(x_i) = \frac{1}{n}$$

- If the discrete uniform random variable is defined on the consecutive integers $a, a+1, \ldots, b$ for $a \leq b$. The mean is

$$\mu = E(X) = \frac{b+a}{2}$$

and the standard deviation is

$$\sqrt{(b-a+1)^2 - 1}$$

Problem 3.80

3-80. The lengths of plate glass parts are measured to the nearest tenth of a millimeter. The lengths are uniformly distributed with values at every tenth of a millimeter starting at 590.0 and continuing through 590.9. Determine the mean and variance of the lengths.

Figure 1: Problem 3.80

Binomial Distribution

- A very common and important distribution. See examples on pages 80
- A **binomial** experiment is an experiment consisting of n repeated trials such that
  1. the trials are independent
  2. each trial results in a Bernoulli outcome
  3. the probability of success on each trial, denoted as p, remains constant
- To be a binomial distribution, the sampling must be done **with replacement**. In some situations, the binomial distribution can be used when the sampling is done without replacement

Binomial Distribution

- The binomial PMF is

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

where $\binom{n}{x} = \frac{n!}{x!(n-x)!}$

- The mean of a binomial random variable is

$$\mu = E(X) = np$$

- The standard deviation of X is

$$\sigma = \sqrt{np(1-p)}$$

Example Problem

Source: Montgomery, Runger, Hubele (2004). Engineering Statistics.

(a) Sketch the probability mass function of X.
(b) Sketch the cumulative distribution.
(c) What value of X is most likely?
(d) What value(s) of X is (are) least likely?

3-79. The random variable X has a binomial distribution with $n = 20$ and $p = 0.5$. Determine the following probabilities.

(a) $P(X = 15)$
(b) $P(X \leq 12)$
(c) $P(X \geq 19)$
(d) $P(13 \leq X < 15)$
(e) Sketch the cumulative distribution function.

Excel Formula for Binomial Example

Problem 3-79

	A	B	C
1	x	f(x)	F(x)
2	0	=BINOMDIST(A2,20,0.5,FALSE)	=BINOMDIST(A2,20,0.5,TRUE)
3	1	=BINOMDIST(A3,20,0.5,FALSE)	=BINOMDIST(A3,20,0.5,TRUE)
4	2	=BINOMDIST(A4,20,0.5,FALSE)	=BINOMDIST(A4,20,0.5,TRUE)
5	3	=BINOMDIST(A5,20,0.5,FALSE)	=BINOMDIST(A5,20,0.5,TRUE)
6	4	=BINOMDIST(A6,20,0.5,FALSE)	=BINOMDIST(A6,20,0.5,TRUE)
7	5	=BINOMDIST(A7,20,0.5,FALSE)	=BINOMDIST(A7,20,0.5,TRUE)
8	6	=BINOMDIST(A8,20,0.5,FALSE)	=BINOMDIST(A8,20,0.5,TRUE)
9	7	=BINOMDIST(A9,20,0.5,FALSE)	=BINOMDIST(A9,20,0.5,TRUE)
10	8	=BINOMDIST(A10,20,0.5,FALSE)	=BINOMDIST(A10,20,0.5,TRUE)
11	9	=BINOMDIST(A11,20,0.5,FALSE)	=BINOMDIST(A11,20,0.5,TRUE)
12	10	=BINOMDIST(A12,20,0.5,FALSE)	=BINOMDIST(A12,20,0.5,TRUE)
13	11	=BINOMDIST(A13,20,0.5,FALSE)	=BINOMDIST(A13,20,0.5,TRUE)
14	12	=BINOMDIST(A14,20,0.5,FALSE)	=BINOMDIST(A14,20,0.5,TRUE)
15	13	=BINOMDIST(A15,20,0.5,FALSE)	=BINOMDIST(A15,20,0.5,TRUE)
16	14	=BINOMDIST(A16,20,0.5,FALSE)	=BINOMDIST(A16,20,0.5,TRUE)
17	15	=BINOMDIST(A17,20,0.5,FALSE)	=BINOMDIST(A17,20,0.5,TRUE)
18	16	=BINOMDIST(A18,20,0.5,FALSE)	=BINOMDIST(A18,20,0.5,TRUE)
19	17	=BINOMDIST(A19,20,0.5,FALSE)	=BINOMDIST(A19,20,0.5,TRUE)
20	18	=BINOMDIST(A20,20,0.5,FALSE)	=BINOMDIST(A20,20,0.5,TRUE)
21	19	=BINOMDIST(A21,20,0.5,FALSE)	=BINOMDIST(A21,20,0.5,TRUE)
22	20	=BINOMDIST(A22,20,0.5,FALSE)	=BINOMDIST(A22,20,0.5,TRUE)

Cumulative Binomial Probability Tables

TABLE II Cumulative Binomial Probabilities $P(X \leq x)$ (*continued*)

		P										
n	*x*	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95	0.99
14	0	0.2288	0.0440	0.0068	0.0008	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.5846	0.1979	0.0475	0.0081	0.0009	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.8416	0.4481	0.1608	0.0398	0.0065	0.0006	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.9559	0.6982	0.3552	0.1243	0.0287	0.0039	0.0002	0.0000	0.0000	0.0000	0.0000
	4	0.9908	0.8702	0.5842	0.2793	0.0898	0.0175	0.0017	0.0000	0.0000	0.0000	0.0000
	5	0.9985	0.9561	0.7805	0.4859	0.2120	0.0583	0.0083	0.0004	0.0000	0.0000	0.0000
	6	0.9998	0.9884	0.9067	0.6925	0.3953	0.1501	0.0315	0.0024	0.0000	0.0000	0.0000
	7	1.0000	0.9976	0.9685	0.8499	0.6047	0.3075	0.0933	0.0116	0.0002	0.0000	0.0000
	8	1.0000	0.9996	0.9917	0.9417	0.7880	0.5141	0.2195	0.0439	0.0015	0.0000	0.0000
	9	1.0000	1.0000	0.9983	0.9825	0.9102	0.7207	0.4158	0.1298	0.0092	0.0004	0.0000
	10	1.0000	1.0000	0.9998	0.9961	0.9713	0.8757	0.6448	0.3018	0.0441	0.0042	0.0000
	11	1.0000	1.0000	1.0000	0.9994	0.9935	0.9602	0.8392	0.5519	0.1584	0.0301	0.0003
	12	1.0000	1.0000	1.0000	0.9999	0.9991	0.9919	0.9525	0.8021	0.4154	0.1530	0.0084
	13	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9932	0.9560	0.7712	0.5123	0.1313
15	0	0.2059	0.0352	0.0047	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.5490	0.1671	0.0353	0.0052	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.8159	0.3980	0.1268	0.0271	0.0037	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.9444	0.6482	0.2969	0.0905	0.0176	0.0019	0.0001	0.0000	0.0000	0.0000	0.0000
	4	0.9873	0.8358	0.5155	0.2173	0.0592	0.0093	0.0007	0.0000	0.0000	0.0000	0.0000
	5	0.9978	0.9389	0.7216	0.4032	0.1509	0.0338	0.0037	0.0001	0.0000	0.0000	0.0000
	6	0.9997	0.9819	0.8689	0.6098	0.3036	0.0950	0.0152	0.0008	0.0000	0.0000	0.0000
	7	1.0000	0.9958	0.9500	0.7869	0.5000	0.2131	0.0500	0.0042	0.0000	0.0000	0.0000
	8	1.0000	0.9992	0.9848	0.9050	0.6964	0.3902	0.1311	0.0181	0.0003	0.0000	0.0000
	9	1.0000	0.9999	0.9963	0.9662	0.8491	0.5968	0.2784	0.0611	0.0022	0.0001	0.0000
	10	1.0000	1.0000	0.9993	0.9907	0.9408	0.7827	0.4845	0.1642	0.0127	0.0006	0.0000
	11	1.0000	1.0000	0.9999	0.9981	0.9824	0.9095	0.7031	0.3518	0.0556	0.0055	0.0000
	12	1.0000	1.0000	1.0000	0.9997	0.9963	0.9729	0.8732	0.6020	01841	0.0362	0.0004
	13	1.0000	1.0000	1.0000	1.0000	0.9995	0.9948	0.9647	0.8329	0.4510	0.1710	0.0096
	14	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9953	0.9648	0.7941	0.5367	0.1399
20	0	0.1216	0.0115	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	1	0.3917	0.0692	0.0076	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	2	0.6769	0.2061	0.0355	0.0036	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	3	0.8670	0.4114	0.1071	0.0160	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	4	0.9568	0.6296	0.2375	0.0510	0.0059	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	5	0.9887	0.8042	0.4164	0.1256	0.0207	0.0016	0.0000	0.0000	0.0000	0.0000	0.0000
	6	0.9976	0.9133	0.6080	0.2500	0.0577	0.0065	0.0003	0.0000	0.0000	0.0000	0.0000
	7	0.9996	0.9679	0.7723	0.4159	0.1316	0.0210	0.0013	0.0000	0.0000	0.0000	0.0000
	8	0.9999	0.9900	0.8867	0.5956	0.2517	0.0565	0.0051	0.0001	0.0000	0.0000	0.0000
	9	1.0000	0.9974	0.9520	0.7553	0.4119	0.1275	0.0171	0.0006	0.0000	0.0000	0.0000
	10	1.0000	0.9994	0.9829	0.8725	0.5881	0.2447	0.0480	0.0026	0.0000	0.0000	0.0000
	11	1.0000	0.9999	0.9949	0.9435	0.7483	0.4044	0.1133	0.0100	0.0001	0.0000	0.0000
	12	1.0000	1.0000	0.9987	0.9790	0.8684	0.5841	0.2277	0.0321	0.0004	0.0000	0.0000
	13	1.0000	1.0000	0.9997	0.9935	0.9423	0.7500	0.3920	0.0867	0.0024	0.0000	0.0000
	14	1.0000	1.0000	1.0000	0.9984	0.9793	0.8744	0.5836	0.1958	0.0113	0.0003	0.0000
	15	1.0000	1.0000	1.0000	0.9997	0.9941	0.9490	0.7625	0.3704	0.0432	0.0026	0.0000
	16	1.0000	1.0000	1.0000	1.0000	0.9987	0.9840	0.8929	0.5886	0.1330	0.0159	0.0000

Binomial Practice Problems

Hypergeometric Distribution

The hypergeometric distribution is one of the commonly occurring distributions in quality.

- A random variable is hypergeometric when a set of N objects contains
  - K objects classified as successes and
  - $N - K$ objects classified as failures
  - a sample of size n is selected **without replacement** from the N objects, where $K \leq N$ and $n \leq N$

Hypergeometric Distribution

- The hypergeometric PMF is

$$f(x) = \frac{\left(\begin{array}{c} K \\ x \end{array}\right)\left(\begin{array}{c} N-K \\ n-x \end{array}\right)}{\left(\begin{array}{c} N \\ n \end{array}\right)}$$

- The mean of X is

$$E(X) = \mu = np$$

- The variance of X is

$$\sigma^2 = V(X) = np(1-p)\left[\frac{N-n}{N-1}\right]$$

Hypergeometric Example Problem

Hypergeometric Example

3-92. Printed circuit cards are placed in a functional test after being populated with semiconductor chips. A lot contains 140 cards, and 20 are selected without replacement for functional testing.

(a) If 20 cards are defective, what is the probability that at least 1 defective card is in the sample?

(b) If 5 cards are defective, what is the probability that at least 1 defective card appears in the sample?

Source: Montgomery & Runger (2003). Applied Statistics & Probability for Engineers

facts: $N = 140$, $n = 20$

part a) $K = 20$, $P[X \geq 1] = 1 - P(X = 0)$

$$f(0) = \frac{\binom{20}{0}\binom{140-20}{20-0}}{\binom{140}{20}} = 0.0356$$

$$P[X \geq 1] = 1 - .0356 = .9644$$

part b) $K = 5$, find $P[X \geq 1] = 1 - P(X = 0)$

$$f(0) = \frac{\binom{5}{0}\binom{140-5}{20-0}}{\binom{140}{20}} = 0.4571$$

$$P[X \geq 1] = 1 - .4571 = .5429$$

Excel for Hypergeometric Example

Hypergeometric Example

x	f(x)	F(x)
0	0.4571	0.4571
1	0.3940	0.8511
2	0.1280	0.9790
3	0.0195	0.9986
4	0.0014	1.0000
5	0.0000	1.0000

Excel Code

	A	B	C
1	x	f(x)	F(x)
2	0	=HYPGEOMDIST(A2,20,5,140)	=B2
3	1	=HYPGEOMDIST(A3,20,5,140)	=C2+B3
4	2	=HYPGEOMDIST(A4,20,5,140)	=C3+B4
5	3	=HYPGEOMDIST(A5,20,5,140)	=C4+B5
6	4	=HYPGEOMDIST(A6,20,5,140)	=C5+B6
7	5	=HYPGEOMDIST(A7,20,5,140)	=C6+B7

Binomial Approximation to the Hypergeometric Distribution

- The mean and variance of the hypergeometric and binomial distribution are very similar. The variance only differs by the finite population correction factor,

$$\frac{N-n}{N-1}$$

- **Sampling with replacement** is equivalent to sampling from an infinite set (without replacement) because the proportion remains constant
- If n is small relative to N, then the finite correction is negligible and the binomial distribution can be used as an approximation to the hypergeometric.
- A rule of thumb is to use this approximation when $N/n > 20$.

Geometric Distribution

- Montgomery and Runger (2003) define a geometric random variable to be the number of trials until the first success of a series of independent Bernoulli trials, with constant probability p of success
- The PMF of a geometric distribution is

$$f(x) = (1-p)^{x-1}p, \; x = 1, 2, \ldots$$

- The mean of a geometric random variable is

$$\mu = E(X) = \frac{1}{p}$$

- The variance of a geometric random variable is

$$\sigma^2 = V(X) = \frac{1-p}{p^2}$$

Geometric Distribution Example

Geometric Distribution Example

3-72. Suppose the random variable X has a geometric distribution with a mean of 2.5. Determine the following probabilities:
(a) $P(X = 1)$ (b) $P(X = 4)$
(c) $P(X = 5)$ (d) $P(X \le 3)$
(e) $P(X > 3)$

Source: Montgomery & Runger (2003). Applied Statistics & Probability for Engineers

Note $\mu = \frac{1}{p} = 2.5 \Rightarrow p = \frac{1}{2.5} = 0.4$

Part a) $P(X=1) = (1-p)^{1-1} p = (1-0.4)^0 \, 0.4 = 0.4$

Part d) $P(X \le 3) = P(X=1) + P(X=2) + P(X=3)$

$$P(X=2) = (1-.4)^{2-1} .4 = 0.24$$
$$P(X=3) = (1-.4)^{3-1} .4 = 0.144$$
$$P(X \le 3) = .4 + .24 + .144 = .784$$

Part e) $P(X>3) = 1 - (P(X=1) + P(X=2))$

$$= 1 - (.4 + .24)$$
$$= .36$$

Negative Binomial Distribution

- Montgomery and Runger (2003) define a negative binomial random variable to be the number of trials until r successes are observed of a series of independent Bernoulli trials, with constant probability p of success
- The geometric distribution is a special case of the negative binomial distribution with $r = 1$
- The PMF of a negative binomial distribution is

$$f(x) = \left(\begin{array}{c} x-1 \\ r-1 \end{array} \right) (1-p)^{x-r} p^r, \; x = r, r+1, \ldots$$

- The mean of a negative binomial random variable is

$$\mu = E(X) = \frac{r}{p}$$

Negative Binomial Example

Negative Binomial Distribution

3-83 An electronic scale in an automated filling operation stops the manufacturing line after three underweight packages are detected. Suppose that the probability of an underweight package is 0.001 and each fill is independent.

(a) What is the mean number of fills before the line is stopped?

(b) What is the standard deviation of the number of fills before the line is stopped?

Source: Montgomery & Runger (2003). Applied Statistics & Probability for Engineers.

part a) $r = 3, \; p = 0.001$

$$\mu = \frac{r}{p} = \frac{3}{.001} = 3{,}000$$

part b) $$\sigma = \sqrt{\frac{r(1-p)}{p^2}} = \sqrt{\frac{3(1-.001)}{.001^2}} = 1{,}731.18$$

Poisson Process

- The number of events over an interval (such as time) is a discrete random variable that is often modelled by the Poisson distribution
- The length of the interval between events is often modeled by the (continuous) exponential distribution
- These two distributions are related

Poisson Process

- The number of events over an interval (such as time) is a discrete random variable that is often modelled by the Poisson distribution
- The length of the interval between events is often modelled by the (continuous) exponential distribution
- These two distributions are related

Poisson Process

Assume that the events occur at random throughout the interval. If the interval can be partitioned into subintervals of small enough length such that

1. The probability of more than one count in a subinterval is zero
2. The probability of one count in a subinterval is the same for all subintervals and proportional to the length of the subinterval, and
3. The count in each subinterval is independent of other subintervals, the random experiment is called a *Poisson process*

Poisson Distribution

If the mean number of counts in the interval is $\lambda > 0$, the random variable X that equals the number of counts in the interval has a **Poisson distribution** with parameter λ

- The Poisson PMF is

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!},\ x = 0, 1, 2, \ldots$$

- The mean of a Poisson random variable is

$$E(X) = \mu = \lambda$$

- The variance of a Poisson random variable is

$$V(X) = \sigma^2 = \lambda$$

Poisson Practice Problems

Poisson Example

Poisson Example

3-100. When network cards are communicating, bits can occasionally be corrupted in transmission. Engineers have determined that the number of bits in error follows a Poisson distribution with mean of 3.2 bits/kb (per kilobyte).

(a) What is the probability of 5 bits being in error during the transmission of 1 kb?

(b) What is the probability of 8 bits being in error during the transmission of 2 kb?

(c) What is the probability of no error bits in 3 kb?

Source: Montgomery, Runger, Hubele (2004). Engineering Statistics

part a) find $P(X=5)$ $\lambda = 3.2$

$$f(5) = \frac{e^{-3.2} 3.2^5}{5!} = 0.114$$

part b) find $P(X=8)$ note λ units changed from 1 KB to 2 KB

$$\lambda = 2(3.2) = 6.4$$

$$f(8) = \frac{e^{-6.4} 6.4^8}{8!} = 0.116$$

part c) find $P(X=0)$ Note: λ units changed again

$$\lambda = 3(3.2) = 9.6$$

$$f(0) = \frac{e^{-9.6} 9.6^0}{0!} = e^{-9.6} = 0.0001$$