Chapter 5—Discrete Probability Distributions

Random variable (RV)--a variable that assumes numerical values associated with the random outcomes of an experiment, where only one numerical value is assigned to each sample point

- ***Discrete RVs***--- random variables that can assume a countable number of values
- ***Continuous RVs***---random variables that can assume values corresponding to any of the points contained in one or more intervals

Random Variables

Identify the following RVs as discrete or continuous

1. The diameter of a tree
2. The number of chapters in your statistics textbook
3. Number of commercials during your favorite TV show
4. The length of the first commercial shown during your favorite TV show
5. The number of registered voters who vote in a national election

Expectations for *RV*s

The ***expected value*** (*EV*) of a *RV* is the mean value of the variable *X* in the sample space, or population of possible outcomes.

EV can be interpreted as the mean value that would be obtained from an infinite number of observations of the random variable.

Probability Distributions for Discrete *RV*s

A complete description of a discrete *RV* requires specification of

- The possible values that the *RV* can assume
- The probability associated with each value

The probability distribution of a discrete *RV*, *X*, can be represented by a graph, table, or formula that specifies the probabilities associated with each possible value of *x*.

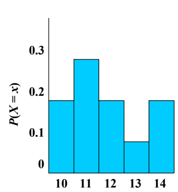
Requirements

1. $0 \le P(X = x) \le 1$ for any value of x
2. $\sum_{\text{all } x} P(X = x) = 1$

Example

Consider the following probability distribution function (pdf):

x	10	11	12	13	14
$P(X=x)$	.2	.3	.2	.1	.2



Cumulative Distribution Function (cdf)

x	10	11	12	13	14
$P(X=x)$	.2	.3	.2	.1	.2

x	10	11	12	13	14
$P(X \le x)$	.2	.5	.7	.8	1.0

Summary Calculations for Discrete *RVs*

The ***mean***, or ***expected value***, of a discrete RV is determined by its probability distribution

$$\mu = E(X) = \sum_{\text{all } x} x_i P(x_i)$$

The ***variance*** of a discrete RV is

$$\sigma^2 = V(X) = E\left[(X-\mu)^2\right] = \sum_{\text{all } x} (x_i - \mu)^2 P(x_i)$$

Calculator formula:

$$\sigma^2 = \sum_{\text{all } x} \left[x_i^2 \times P(x_i)\right] - \mu^2$$

Standard Deviation for Discrete *RVs*

The standard deviation of a discrete RV is equal to the square root of the variance:

$$\sigma = \sqrt{\sigma^2}$$

This quantifies how spread out the possible values of a discrete RV might be, weighted by how likely each value is to occur.

Expected Values for Discrete *RVs*

Example from Mathematical Statistics with Applications, Mendenhall et al, 1981, p. 99

Given the probability distribution of Y, ***find the mean, variance and SD for Y***

y	$P(Y=y)$
0	.125
1	.25
2	.375
3	.25

$$\begin{aligned}\mu_Y = E(Y) &= \sum_{i=1}^{4} y_i P(y_i) \\ &= (0)(.125)+(1)(.25)+(2)(.375)+(3)(.25) \\ &= 1.75\end{aligned}$$

Variance for Discrete *RVs*

Example from Mathematical Statistics with Applications, Mendenhall et al, 1981, p. 99 (continued)

$$\begin{aligned}\sigma_Y^2 = E\left[(Y-\mu)^2\right] &= \sum_{i=1}^{4} (y_i-\mu)^2 P(y_i) \\ &= (0-1.75)^2(.125)+(1-1.75)^2(.25)+ \\ &= (2-1.75)^2(.375)+(3-1.75)^2(.25) \\ &= 0.9375\end{aligned}$$

$$\sigma_Y = \sqrt{\sigma_Y^2} = \sqrt{0.9375} = 0.97$$

Rare Event Rule

If, under a given assumption, the probability of a particular observed event is extremely small, we conclude that the assumption is probably not correct.

- Range rule of thumb: $\mu \pm 2\sigma$
- Use probabilities

  Unusually high number of successes

  ➔ $P(x$ or more$) \leq 0.05$

  Unusually low number of successes

  ➔ $P(x$ or fewer$) \leq 0.05$

Example

Based on past results found in the *Information Please Almanac*, there is a 0.1818 probability that a baseball World Series contest will last four games, a 0.2121 probability that it will last five games, a 0.2323 probability that it will last six games, and a 0.3737 probability that it will last seven games. Is it unusual for a team to "sweep" by winning in four games?

Let Y = number of games played in a world series

y	$P(Y=y)$
4	.1818
5	.2121
6	.2323
7	.3737

Binomial Random Variables

Characteristics of a binomial experiment

1. The experiment consists of ***n* identical (fixed)** trials
2. The trials are **independent**
3. The experiment results in a ***dichotomous*** response; i.e., there are only two possible outcomes on each trial. One outcome is denoted by *S* (success) and the other by *F* (failure)
4. The probability of *S*, denoted as *p*, remains the same from trial to trial. The probability of *F*, denoted as *q*, is equal to $1 - p$.

The **binomial random variable**, *X*, is the number of *S*'s in *n* trials

Binomial Random Variables--Example

Example

Record the sequence of heads and tails in 3 tosses of an unfair coin where the *P*(H) = .6 and the *P*(T) = .4. We are interested in the distribution of the number of tails.

What is *n*, the number of trials?

Are the trials identical?

Are the trials independent?

What is *S*?

What is *X*?

Example (continued)

How many possible outcomes are there?
Number of outcomes = 2^3= 8
Possible outcomes are:

HHH	P(HHH) = 0.6 x 0.6 x 0.6 = 0.216
HHT	P(HHT) = 0.6 x 0.6 x 0.4 = 0.144
HTH	P(HTH) = 0.144
THH	P(THH) = 0.144
HTT	P(HTT) = 0.6 x 0.4 x 0.4 = 0.096
THT	P(THT) = 0.096
TTH	P(TTH) = 0.096
TTT	P(TTT) = 0.4 x 0.4 x 0.4 = 0.064

X	0	1	2	3
P(*X*)	0.216	0.432	0.288	0.064

The Binomial Probability Distribution

$$P(X=x)=\binom{n}{x}p^x q^{n-x} \quad (x=0,1,2,\ldots,n)$$

where

$p =$ probability of success on a single trial

$q = 1 - p$

$n =$ number of trials

$x =$ number of successes in n trials

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

The Binomial Probability Distribution

Mean, Variance and SD for a Binomial RV

Mean: $\mu = np$

Variance: $\sigma^2 = npq$

SD: $\sigma = \sqrt{npq}$

The Binomial Probability Distribution

For the tossing three coins example we can calculate various quantities:

$$P(X=2)=\binom{3}{2}(.4)^2(1-.4)^{3-2}=3(.4)^2(.6)^1=.288$$

$$\mu_X = np = 3(.4) = 1.2$$

$$V(X) = npq = 3(.4)(.6) = 0.72; \quad \sigma_X = \sqrt{0.72} = 0.85$$

Generic Example: $n = 6$; $p = 0.4$

Can calculate various quantities:

$$P(X=0)=\binom{6}{0}(.4)^0(1-.4)^{6-0}=1(1)(.6)^6=.0467$$

$$P(X=1)=\binom{6}{1}(.4)^1(1-.4)^{6-1}=6(.4)(.6)^5=.1866$$

$$P(X=2)=\binom{6}{2}(.4)^2(1-.4)^{6-2}=15(.4)^2(.6)^4=.3110$$

Can also calculation cumulative probabilities:

$$\begin{aligned}P(X\le 2)&=P(X=0)+P(X=1)+P(X=2)\\&=.3110+.1866+.0467\\&=.5443\end{aligned}$$

$$\begin{aligned}P(X>2)&=1-P(X\le 2)\\&=1-.5443\\&=.4557\end{aligned}$$

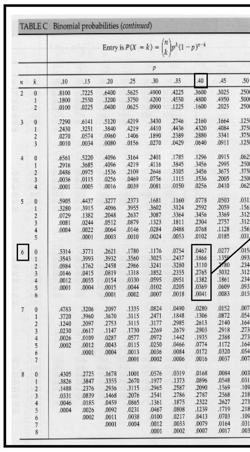


TABLE C Binomial probabilities (*continued*)

Entry is $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$

		p								
n	k	.10	.15	.20	.25	.30	.35	.40	.45	.50
2	0	.8100	.7225	.6400	.5625	.4900	.4225	.3600	.3025	.2500
	1	.1800	.2550	.3200	.3750	.4200	.4550	.4800	.4950	.5000
	2	.0100	.0225	.0400	.0625	.0900	.1225	.1600	.2025	.2500
3	0	.7290	.6141	.5120	.4219	.3430	.2746	.2160	.1664	.1250
	1	.2430	.3251	.3840	.4219	.4410	.4436	.4320	.4084	.3750
	2	.0270	.0574	.0960	.1406	.1890	.2389	.2880	.3341	.3750
	3	.0010	.0034	.0080	.0156	.0270	.0429	.0640	.0911	.1250
4	0	.6561	.5220	.4096	.3164	.2401	.1785	.1296	.0915	.0625
	1	.2916	.3685	.4096	.4219	.4116	.3845	.3456	.2995	.2500
	2	.0486	.0975	.1536	.2109	.2646	.3105	.3456	.3675	.3750
	3	.0036	.0115	.0256	.0469	.0756	.1115	.1536	.2005	.2500
	4	.0001	.0005	.0016	.0039	.0081	.0150	.0256	.0410	.0625
5	0	.5905	.4437	.3277	.2373	.1681	.1160	.0778	.0503	.0313
	1	.3280	.3915	.4096	.3955	.3602	.3124	.2592	.2059	.1563
	2	.0729	.1382	.2048	.2637	.3087	.3364	.3456	.3369	.3125
	3	.0081	.0244	.0512	.0879	.1323	.1811	.2304	.2757	.3125
	4	.0004	.0022	.0064	.0146	.0284	.0488	.0768	.1128	.1562
	5		.0001	.0003	.0010	.0024	.0053	.0102	.0185	.0312
6	0	.5314	.3771	.2621	.1780	.1176	.0754	.0467	.0277	.0156
	1	.3543	.3993	.3932	.3560	.3025	.2437	.1866	.1359	.0938
	2	.0984	.1762	.2458	.2966	.3241	.3280	.3110	.2780	.2344
	3	.0146	.0415	.0819	.1318	.1852	.2355	.2765	.3032	.3125
	4	.0012	.0055	.0154	.0330	.0595	.0951	.1382	.1861	.2344
	5	.0001	.0004	.0015	.0044	.0102	.0205	.0369	.0609	.0937
	6			.0001	.0002	.0007	.0018	.0041	.0083	.0156
7	0	.4783	.3206	.2097	.1335	.0824	.0490	.0280	.0152	.0078
	1	.3720	.3960	.3670	.3115	.2471	.1848	.1306	.0872	.0547
	2	.1240	.2097	.2753	.3115	.3177	.2985	.2613	.2140	.1641
	3	.0230	.0617	.1147	.1730	.2269	.2679	.2903	.2918	.2734
	4	.0026	.0109	.0287	.0577	.0972	.1442	.1935	.2388	.2734
	5	.0002	.0012	.0043	.0115	.0250	.0466	.0774	.1172	.1641
	6		.0001	.0004	.0013	.0036	.0084	.0172	.0320	.0547
	7				.0001	.0002	.0006	.0016	.0037	.0078
8	0	.4305	.2725	.1678	.1001	.0576	.0319	.0168	.0084	.0039
	1	.3826	.3847	.3355	.2670	.1977	.1373	.0896	.0548	.0313
	2	.1488	.2376	.2936	.3115	.2965	.2587	.2090	.1569	.1094
	3	.0331	.0839	.1468	.2076	.2541	.2786	.2787	.2568	.2188
	4	.0046	.0185	.0459	.0865	.1361	.1875	.2322	.2627	.2734
	5	.0004	.0026	.0092	.0231	.0467	.0808	.1239	.1719	.2188
	6		.0002	.0011	.0038	.0100	.0217	.0413	.0703	.1094
	7			.0001	.0004	.0012	.0033	.0079	.0164	.0312
	8					.0001	.0002	.0007	.0017	.0039

Table A-1, page 769 in Triola

.0467
.1866
.3110
.2765
.1382
.0369
.0041

TI 83/84 binompdf(n, p, x) and binomcdf( )

Press 2nd [DISTR]
Press ALPHA **A** for binompdf or
ALPHA **B** for binomcdf or
scroll through the list and press enter

Let n = 6 and p = 0.4.

To find $P(X = 3)$ use binompdf(6, .4, 3) → .27648

To find individual probabilities for more than one value of X at a time use binompdf(6, .4, {3, 4}) → {.27648 .13824}

To find $P(X \le 3)$ use binomcdf(6, .4, 3) →.8208

To find $P(X < 3)$ use binomcdf(6, .4, **2**) → .54432

To find $P(1 \le X \le 3)$ use binomcdf(6, .4, 3) – binomcdf(6, .4, **0**) → .774144

Example

Consider the discrete probability distribution:

X	10	12	18	20
$P(X)$	.2	.3	.1	.4

Calculate μ, σ^2, and σ

What is $P(x<15)$?

Calculate $\mu \pm 2\sigma$

What is the probability that X is in the interval $\mu \pm 2\sigma$?

Example

Problem 4.47 from McClave and Sincich, 9th edition, pg. 200

A Federal Trade Commission (FTC) study of the pricing accuracy of electronic checkout scanners at stores found that one of every 30 items is priced incorrectly. Suppose the FTC randomly selects five items at a retail store and checks the accuracy of the scanner price of each. Let X represent the number of the five items that is priced incorrectly.

a) Show that X is a binomial RV.

b) Use the information in the FTC study to estimate p for the binomial experiment.

c) What is the probability that exactly one of the five items is priced incorrectly by the scanner?

d) What is the probability that at least one of the five items is priced incorrectly by the scanner?

e) What is the probability that X is in the interval $\mu \pm 2\sigma$?