STAT 400
UIUC

Examples for 2.4 (Extended)

Stepanov
Dalpiaz

5. Suppose that on Halloween 6 children come to a house to get treats. A bag contains 8 plain chocolate bars and 7 nut bars. Each child reaches into the bag and randomly selects 1 candy bar. Let X denote the number of nut bars selected.

a) Is the Binomial model appropriate for this problem?

No. Without replacement $\Rightarrow$ Trials are not independent.

b) Find the probability that exactly 2 nut bars were selected.

15 ↙ ↘
7 nut ↓ 2   8 plain ↓ 4

$$\frac{{}_7C_2 \cdot {}_8C_4}{{}_{15}C_6} = \frac{21 \cdot 70}{5{,}005} \approx \mathbf{0.2937}.$$

OR

$${}_6C_2 \cdot \left[\frac{7}{15} \cdot \frac{6}{14}\right] \cdot \left[\frac{8}{13} \cdot \frac{7}{12} \cdot \frac{6}{11} \cdot \frac{5}{10}\right] \approx \mathbf{0.2937}.$$

Hypergeometric Distribution:

N = population size,
S = number of "successes" in the population,
$N - S$ = number of "failures" in the population,
n = sample size.

X = number of "successes" in the sample when sampling is done without replacement.

Then

$$\mathrm{P}(X = x) = \frac{\binom{S}{x} \cdot \binom{N-S}{n-x}}{\binom{N}{n}} = \frac{{}_SC_x \cdot {}_{N-S}C_{n-x}}{{}_NC_n}$$

OR

$$\mathrm{P}(X = x) = \binom{n}{x} \cdot \left[\frac{S}{N} \cdot \frac{S-1}{N-1} \cdot \ldots \cdot \frac{S-x+1}{N-x+1}\right] \cdot \left[\frac{N-S}{N-x} \cdot \frac{N-S-1}{N-x-1} \cdot \ldots \cdot \frac{N-S-(n-x)+1}{N-n+1}\right]$$

$$\max(0,\, n + S - N) \le x \le \min(n,\, S).$$

c) Find the probability that at most 2 nut bars were selected.

$$P(X \le 2) = \frac{{}_7C_0 \cdot {}_8C_6}{{}_{15}C_6} + \frac{{}_7C_1 \cdot {}_8C_5}{{}_{15}C_6} + \frac{{}_7C_2 \cdot {}_8C_4}{{}_{15}C_6}$$

$$= \frac{1 \cdot 28}{5{,}005} + \frac{7 \cdot 56}{5{,}005} + \frac{21 \cdot 70}{5{,}005} \approx \mathbf{0.3776}.$$

d) Find the probability that at least 4 nut bars were selected.

$$P(X \ge 4) = \frac{{}_7C_4 \cdot {}_8C_2}{{}_{15}C_6} + \frac{{}_7C_5 \cdot {}_8C_1}{{}_{15}C_6} + \frac{{}_7C_6 \cdot {}_8C_0}{{}_{15}C_6}$$

$$= \frac{35 \cdot 28}{5{,}005} + \frac{21 \cdot 8}{5{,}005} + \frac{7 \cdot 1}{5{,}005} \approx \mathbf{0.23077}.$$

6. A jar has N marbles, S of them are orange and $N - S$ are blue. Suppose 3 marbles are selected. Find the probability that there are 2 orange marbles in the sample, if the selection is done …

	with replacement	without replacement
a) $N = 10$, $S = 4$;	${}_3C_2 \cdot (0.40)^2 \cdot (0.60)^1 = \mathbf{0.288}$.	$\frac{{}_4C_2 \cdot {}_6C_1}{{}_{10}C_3} = \mathbf{0.30}$.
b) $N = 100$, $S = 40$;	${}_3C_2 \cdot (0.40)^2 \cdot (0.60)^1 = \mathbf{0.288}$.	$\frac{{}_{40}C_2 \cdot {}_{60}C_1}{{}_{100}C_3} \approx \mathbf{0.289425}$.
c) $N = 1{,}000$, $S = 400$;	${}_3C_2 \cdot (0.40)^2 \cdot (0.60)^1 = \mathbf{0.288}$.	$\frac{{}_{400}C_2 \cdot {}_{600}C_1}{{}_{1000}C_3} \approx \mathbf{0.288144}$.

	Binomial	**Hypergeometric**
	with replacement	without replacement
Probability	$P(X = x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$	$P(X = x) = \dfrac{\binom{S}{x} \cdot \binom{N-S}{n-x}}{\binom{N}{n}}$
Expected Value	$E(X) = n \cdot p$	$E(X) = n \cdot \dfrac{S}{N}$
Variance	$Var(X) = n \cdot p \cdot (1-p)$	$Var(X) = n \cdot \dfrac{S}{N} \cdot \left(1 - \dfrac{S}{N}\right) \cdot \dfrac{N-n}{N-1}$

If the population size is large (**compared to the sample size**) Binomial Distribution can be used regardless of whether sampling is with or without replacement.

6½. In each of the following cases, is it appropriate to use Binomial model?

If yes, what are the values of its parameters n and p (if known)?

If no, explain why Binomial model is not appropriate.

a) A fair 6-sided die is rolled 7 times. X = # of 6's.

Yes. $n = 7$, $p = 1/6$.

b) A fair coin is tossed 3 times. X = # of H's.

Yes. $n = 3$, $p = 0.50$.

c) An exam consists of 10 questions, the first 4 are True-False, the last 6 are multiple choice questions with 4 possible answers each, only one of which is correct. A student guesses independently on each question. X = # of questions he answers correctly.

No. The probability of success is not the same for all trials.

d) Suppose 20% of the customers at a particular gas station select Premium gas
X = # of customers at this gas station on a particular day who selected Premium gas.

No. The number of trials is not fixed.

e) Suppose 20% of the customers at a particular gas station select Premium gas X = # of customers in the first 10 at a gas station on a particular day who selected Premium gas.

Yes. $n = 10$, $p = 0.20$.

f) A box contains 40 parts, 10 of which are defective. A person takes 7 parts out of the box with replacement. X = # of defective parts selected.

Yes. $n = 7$, $p = {}^{10}/_{40} = 0.25$.

g) A box contains 40 parts, 10 of which are defective. A person takes 7 parts out of the box without replacement. X = # of defective parts selected.

No. Trials are not independent.

h) A box contains 400,000 parts, 100,000 of which are defective. A person takes 7 parts out of the box without replacement. X = # of defective parts selected.

No. Trials are not independent. However, Binomial distribution can be used as an approximation.

i) Seven members of the same family are tested for a particular food allergy. X = # of family members who are allergic to this particular food.

Yes if we can assume independence, No if we cannot.

j) In Neverland, 10% of the labor force is unemployed. A random sample of 400 individuals is selected. X = # of individuals in the sample who are unemployed.

Yes. $n = 400$, $p = 0.10$.

k) Suppose that 5% of tax returns have arithmetic errors. 25 tax returns are selected at random. X = # of arithmetic errors in those 25 tax returns.

No. More than two possible outcomes for each trial.

l) Suppose that 5% of tax returns have arithmetic errors. 25 tax returns are selected at random. X = # of tax returns among those 25 with arithmetic errors.

Yes. $n = 25$, $p = 0.05$.

Multinomial Distribution:

- The number of trials, n, is fixed.
- Each trial has k possible outcomes, with probabilities $p_1, p_2, \ldots, p_k$, respectively. $(p_1 + p_2 + \ldots + p_k = 1)$
- The trials are independent.
- $X_1, X_2, \ldots, X_k$ represent the number of times outcome 1, outcome 2, … , outcome k occur, respectively. $(X_1 + X_2 + \ldots + X_k = n)$

Then

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{x_1!\,x_2!\ldots x_k!}\, p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k},$$

$$x_1 + x_2 + \ldots + x_k = n.$$

7. A particular brand of candy-coated chocolate comes in six different colors. Suppose 30% of all pieces are brown, 20% are blue, 15% are red, 15% are yellow, 10% are green, and 10% are orange. Thirty pieces are selected at random.

a) What is the probability that 10 are brown, 8 are blue, 7 are red, 3 are yellow, 2 are green, and none are orange?

$$\frac{30!}{10!\cdot 8!\cdot 7!\cdot 3!\cdot 2!\cdot 0!}\cdot(0.30)^{10}\cdot(0.20)^{8}\cdot(0.15)^{7}\cdot(0.15)^{3}\cdot(0.10)^{2}\cdot(0.10)^{0}$$

b) What is the probability that 10 are brown, 8 are blue, and 12 are of other colors?

$$\frac{30!}{10!\cdot 8!\cdot 12!}\cdot(0.30)^{10}\cdot(0.20)^{8}\cdot(0.50)^{12}$$