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?

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

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

$$\mathrm{P}(X = x) = \frac{\binom{S}{x} \cdot \binom{N-S}{n-x}}{\binom{N}{n}} = \frac{{}_S C_x \cdot {}_{N-S} C_{n-x}}{{}_N C_n}$$

OR

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

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

EXCEL: $=\mathrm{HYPGEOMDIST}(x, n, S, N)$ gives $\mathrm{P}(X = x)$

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) $N = 10,\ S = 4;$

b) $N = 100,\ S = 40;$

c) $N = 1{,}000,\ S = 400;$

	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	$Var(X) = n \cdot p \cdot (1-p)$	$Var(X) = n \cdot \dfrac{S}{N} \cdot \left(1 - \dfrac{S}{N}\right) \cdot \dfrac{N-n}{N-1}$

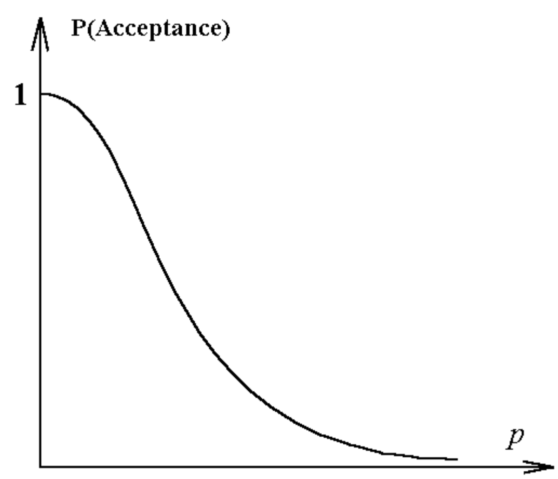
If the population size is large (**compared to the sample size**) Binomial Distribution can be used regardless of whether sampling is with or without replacement.

Acceptance Sampling

Suppose a retailer has to decide whether to accept or reject an incoming lot of supplies. In a retail store it is a lot of merchandise. If one examines the whole lot to determine p, the proportion of defectives, then it would be time consuming and costly. In order to devise a plan which is more practical one takes a sample from the lot. Note that usually the lot is very large compared to the size of the sample. So even if the sampling is without replacement, we can assume the number of defectives in sample to follow **Binomial** distribution.

Note that the proportion p of defectives is not known. However, any plan we draw should be such that we should accept the lot with high probability if p is close to zero and with low probability for p close to one.

Note that the probability of accepting the lot depends on the (unknown) proportion p of defectives. If we plot the probability of accepting for different values of p, we get a curve that is called operating characteristic (OC) curve of the sampling plan. A typical OC curve has the following form:



Example: Suppose a store chain has to make a decision to accept an incoming lot of 100,000 thingamabobs. The purchasing agent devises two plans.

Plan I: Sample 50 with replacement and accept the lot if there are at most two defectives in the sample.

Plan II: Sample 40 with replacement and accept the lot if there is at most one defective in the sample. If three or more defectives are found reject the lot. If there are two defectives in the sample, take a second sample of 20 with replacement and accept the lot if there is at most one defective in it, otherwise reject.

Plan II is called a multiple sampling plan. It is less expensive, as in most of the cases, the decision is made based on 40 items.

Note that for both plans number of defectives in each sample is a **Binomial** random variable. For Plan I, $n = 50$. For Plan II, $n = 40$ on the first stage and $n = 20$ on the second.

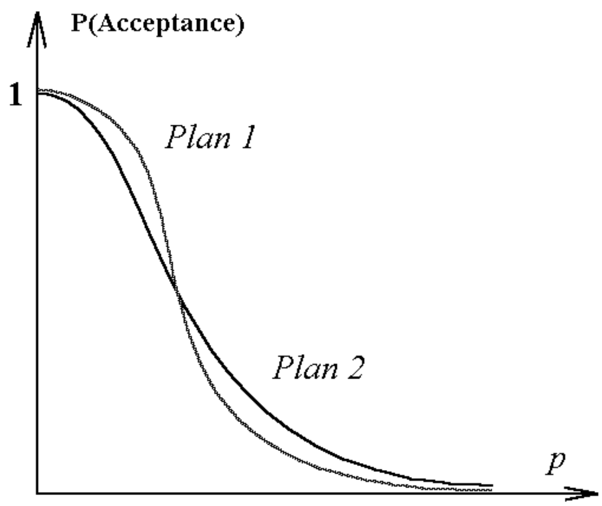
For Plan I:

$$\text{P(Acceptance)} = \text{P}(X \le 2 \mid n = 50, p).$$

For Plan II:

$$\text{P(Acceptance)} = \text{P}(X \le 1 \mid n = 40, p) + \text{P}(X = 2 \mid n = 40, p) \cdot \text{P}(X \le 1 \mid n = 20, p).$$

Ideally, we want a plan that would have high probability of acceptance for "good" lots (with p close to zero) and low probability of acceptance for "bad" lots (with p "large").



In this sense Plan 1 is superior to Plan 2 as Plan 1 has higher probability of accepting "good" lots and lower probability of accepting "bad" lots.

If we increase the sample size the OC curve becomes more "ideal". Unfortunately, the work load also increases. However, if one knows the maximum defective rate that one can tolerate, it is possible to come up with a reasonable sampling plan (i.e. set the sample size and the "critical" number of the defective items in the sample) so that the probability of accepting "good" lots is high, that the probability of accepting "bad" lots is low, and the work load is bearable.

Similar plans can also be used to control quality in a manufacturing process.

Multinomial Distribution:

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

Then

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

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

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?

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