University of California, Los Angeles
Department of Statistics

Statistics 100A **Instructor: Nicolas Christou**

Some special discrete probability distributions

- **Bernoulli random variable:**
  It is a variable that has 2 possible outcomes: "success", or "failure". Success occurs with probability p and failure with probability $1 - p$.

- **Binomial probability distribution:**

  Suppose that n independent Bernoulli trials each one having probability of success p are to be performed. Let X be the number of successes among the n trials. We say that X follows the binomial probability distribution with parameters n, p.

  Probability mass function of X:

  $$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, 3, \cdots, n$$

  or

  $$P(X = x) = nCx\ p^x (1-p)^{n-x}, \quad x = 0, 1, 2, 3, \cdots, n$$

  where

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

Expected value of X:	$E(X) = np$
Variance of X:	$\sigma^2 = np(1-p)$
Standard deviation of X:	$\sigma = \sqrt{np(1-p}$

- **Geometric probability distribution:**
  Suppose that repeated independent Bernoulli trials each one having probability of success p are to be performed. Let X be the number of trials needed until the first success occurs. We say that X follows the geometric probability distribution with parameter p.
  Probability mass function of X:

$$P(X = x) = (1 - p)^{x-1} p, \quad x = 1, 2, 3, \cdots$$

Expected value of X:	$E(X) = \frac{1}{p}$
Variance of X:	$\sigma^2 = \frac{1-p}{p^2}$
Standard deviation of X:	$\sigma = \sqrt{\frac{1-p}{p^2}}$

- More on geometric probability distribution $\cdots$
  Repeated Bernoulli trials are performed until the first success occurs. Find the probability that
  - the first success occurs after the k_{th} trial
  - the first success occurs on or after the k_{th} trial
  - the first success occurs before the k_{th} trial
  - the first success occurs on or before the k_{th} trial

- **Negative binomial probability distribution:**
  Suppose that repeated Bernoulli trials are performed until r successes occur. The number of trials required X, follows the so called negative binomial probability distribution.
  Probability mass function of X is:

$$P(X=x)=\binom{x-1}{r-1}p^{r-1}(1-p)^{x-r}p, \quad \text{or}$$

$$P(X=x)=\binom{x-1}{r-1}p^{r}(1-p)^{x-r}$$

$$x=r, r+1, r+2, \cdots$$

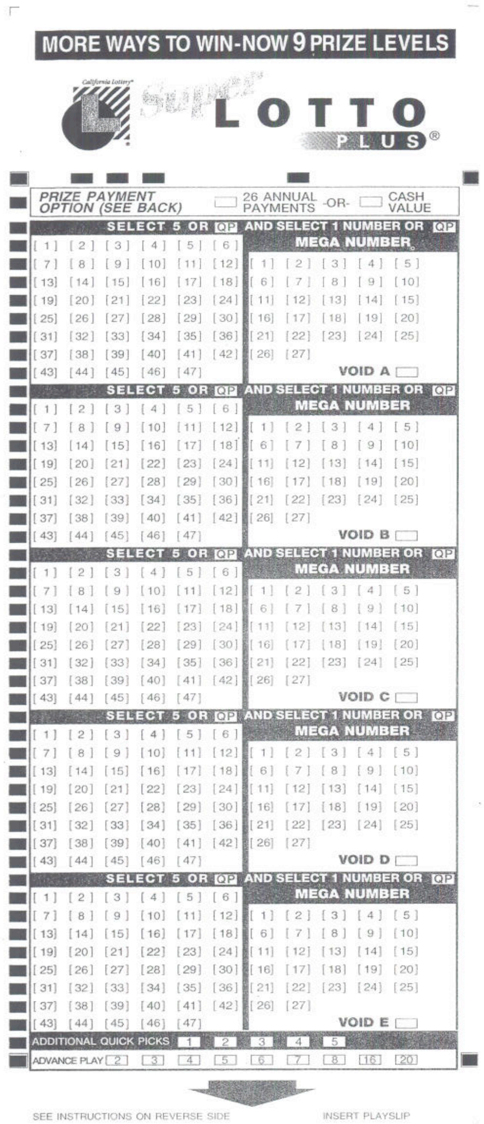
Expected value of X:	$E(X)=\frac{r}{p}$
Variance of X:	$\sigma^2=r\frac{1-p}{p^2}$
Standard deviation of X:	$\sigma=\sqrt{r\frac{1-p}{p^2}}$

- **Hypergeometric probability distribution:**
  Select without replacement n from N available items (of which r are labeled as "hot items", and $N - r$ are labeled as "cold items"). Let X be the number of hot items among the n.
  Probability mass function of X:

$$P(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

Expected value of X:	$E(X) = \frac{nr}{N}$
Variance of X:	$\sigma^2 = \frac{nr(N-r)(N-n)}{N^2(N-1)}$
Standard deviation of X:	$\sigma = \sqrt{\frac{nr(N-r)(N-n)}{N^2(N-1)}}$

California Super Lotto Plus:



Play SuperLOTTO Plus®

- Mark 5 numbers from 1 to 47 and 1 MEGA number from 1 to 27 for each play. Do not use red ink
- Or mark Quick Pick® to randomly select just the 5 numbers, or only the MEGA number, or both the 5 and the MEGA numbers.
- You may select 1 to 5, 10 or 20 additional Quick Pick® plays on one ticket.
- Play the same numbers for consecutive draws by marking Advance Play®. You may select 2-8,16, 20 or 26 Advance plays.
- Choose how you want to be paid if you win the Jackpot prize. Select either "26 Annual Payments" or "Cash Value".
- 26 Annual Payments: Jackpot paid in 26 annual payments that increase each year.
- Cash Value: Single cash payment, estimated at 40% to 50% of the Announced Jackpot. If you do not choose a Prize Payment Option, the Jackpot will be paid in 26 Annual Payments.
- Your ticket will show your numbers, draw date(s), and payment selection.
- If you make a mistake mark the VOID box. DO NOT ERASE.
- Check your ticket selections for accuracy. If inaccurate, cancel the ticket on the same day of purchase and before pool closure, from the retailer that sold the ticket.
- Your ticket not the playslip is your valid receipt. Sign the back of your ticket to indicate ownership.

HOW TO FILL OUT YOUR PLAY SELECTION

CORRECT INCORRECT

[X] [✓] [•]

If you make a mistake, mark the VOID box. DO NOT ERASE

PRIZE CATEGORIES AND ODDS

MATCHING NUMBERS	PRIZES	(ODDS 1:)
All 5 of 5 and MEGA	Pari-mutuel*	41,416,353
All 5 of 5	Pari-mutuel*	1,592,937
Any 4 of 5 and MEGA	Pari-mutuel*	197,221
Any 4 of 5	Pari-mutuel*	7,585
Any 3 of 5 and MEGA	Pari-mutuel*	4,810
Any 3 of 5	Pari-mutuel*	185
Any 2 of 5 and MEGA	Pari-mutuel*	361
Any 1of 5 and MEGA	Pari-mutuel*	74
None of 5, only MEGA	Pari-mutuel*	49
Overall Odds of Winning		23

* Prize amounts will vary depending on sales levels and number of winners

On any play, only the highest prize won is awarded

For winning Lottery numbers please log on to www.calottery.com

Play Responsibly. Gambling Problem Help Line: 1-888-277-3115

For any other Lottery information call 1-800-LOTTERY

Telecommunications Device for the deaf (TDD) 1-800-345-4275

TO CLAIM YOUR PRIZE

- All winning tickets must be claimed within 180 days after the draw in which the prize was won.
- There may be different claiming periods for replays, entries, coupons and promotions. Contact the Lottery for details.
- Winning tickets under $600 per play may be claimed at any participating cashing retailer.
- For winning tickets $600 or more, per play, submit a claim form and your original ticket to the Lottery for payment.
- Federal, State law and Lottery rules and regulations apply. Must be at least 18 years of age to purchase, play and claim.

LO 9/04 B1

- **Poisson probability distribution:**
  The Poisson probability mass function with parameter $\lambda > 0$ (where λ is the average number of events occur per time, area, volume, etc.) is:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \cdots$$

Expected value of X:	$E(X) = \lambda$
Variance of X:	$\sigma^2 = \lambda$
Standard deviation of X:	$\sigma = \sqrt{\lambda}$

- **Multinomial probability distribution:**
  A sequence of n independent experiments is performed and each experiment can result in one of r possible outcomes with probabilities $p_1, p_2, \ldots, p_r$ with $\Sigma_{i=1}^r p_i = 1$. Let X_i be the number of the n experiments that result in outcome i, $i = 1, 2, \ldots, r$. Then,

$$P(X_1 = n_1, X_2 = n_2, \ldots, X_r = n_r) = \frac{n!}{n_1! n_2! \cdots n_r!} p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$$

Example:
Suppose 20 patients arrive at a hospital on any given day. Assume that 10% of all the patients of this hospital are emergency cases.

a. Find the probability that exactly 5 of the 20 patients are emergency cases.
b. Find the probability that none of the 20 patients are emergency cases.
c. Find the probability that all 20 patients are emergency cases.
d. Find the probability that at least 4 of the 20 patients are emergency cases.
e. Find the probability that more than 4 of the 20 patients are emergency cases.
f. Find the probability that at most 3 of the 20 patients are emergency cases.
g. Find the probability that less than 3 of the 20 patients are emergency cases.
h. On average how many patients (from the 20) are expected to be emergency case?

Example:
Patients arrive at a hospital. Assume that 10% of all the patients of this hospital are emergency cases.

a. Find the probability that at any given day the 20_{th} patient will be the first emergency case.
b. On average how many patients must arrive at the hospital to find the first emergency case?
c. Find the probability that the first emergency case will occur after the arrival of the 20_{th} patient.
d. Find the probability that the first emergency case will occur on or before the arrival of the 15_{th} patient.

Example:
The probability for a child to catch a certain disease is 20%. Find the probability that the 12_{th} child exposed to the disease will be the 3_{rd} to catch it.

Example:
A basket contains 20 fruits of which 10 are oranges, 8 are apples, and 2 are tangerines. You randomly select 5 and give them to your friend. What is the probability that among the 5, your friend will get 2 tangerines?

Example:
In a bag there are 10 green and 5 white chips. You randomly select chips (one at a time) with replacement until a green is obtained. Your friend does the same. Find the variance of the sum of the number of trials needed until the first green is obtained by you and your friend.

Example:
The number of industrial accidents at a particular plant is found to average 3 per month. Find the probability that 8 accidents will occur at any given month.

Review problems

Problem 1
Suppose $X \sim b(n, p)$.

a. The calculation of binomial probabilities can be computed by means of the following *recursion formula*. Verify this formula.

$$P(X = x + 1) = \frac{p(n - x)}{(x + 1)(1 - p)} P(X = x)$$

b. Let $X \sim b(8, 0.25)$. Use the above result to calculate $P(X = 1)$, and $P(X = 2)$. You are given that $P(X = 0) = 0.1001$.

Problem 2
The amount of flour used per week by a bakery is a random variable X having an exponential distribution with mean equal to 4 tons. The cost of the flour per week is given by $Y = 3X^2+1$.

a. Find the median of X.

b. Find the 20_{th} percentile of the distribution of X.

c. What is the variance of X?

d. Find $P(X > 6/X > 2)$.

e. What is the expected cost?

Problem 3
Answer the folowing questions:

a. If the probabilities of having a male or female offspring are both 0.50, find the probability that a family's fifth child is their second son.

b. Suppose the probability that a car will have a flat tire while driving on the 405 freeway is 0.0004. What is the probability that of 10000 cars driving on the 405 freeway fewer than 3 will have a flat tire. Use the Poisson approximation to binomial for faster calculations.

c. A doctor knows from experience that 15% of the patients who are given a certain medicine will have udesirable side effects. What is the probability that the tenth patient will be the first to show these side effects.

d. Suppose X follows the geometric probability distribution with $p = 0.2$.
Find $P(X \geq 10)$.

e. Let $X \sim b(n, 0.4)$. Find n so that $P(X \geq 1) = 0.99$.

Problem 4
For a certain section of a pine forest, the number of diseased trees per acre, X, follows the Poisson distribution with $\lambda = 10$. The diseased trees are sprayed with an insecticide at a cost of \$3.00 per tree, plus a fixed overhead cost for equipment rental of \$50.00.

a. Find the probability that a randomly selected acre from this forest will contain at least 12 diseased trees.

b. Letting C denote the total cost for a randomly selected acre, find the expected value and standard deviation of C.

Problem 5
A particular sale involves 4 items randomly selected from a large lot that is known to contain 10% defectives. Let X denote the number of defectives among the 4 sold. The purchaser of the items will return the defectives for repair, and the repair cost is given by $C = 3X^2 + X + 2$. Find the expected repair cost.

Problem 6
The telephone lines serving an airline reservation office all are busy 60% of the time.

a. If you are calling this office, what is the probability that you complete your call on the first try? the second try? the third try?

b. If you and your friend must both complete calls to this office, what is the probability that it takes a total of 4 tries for both of you to get through?

Problem 7
In the daily production of a certain kind of rope, the number of defects per foot X is assumed to have a Poisson distribution with mean $\lambda = 2$. The profit per foot when the rope is sold is given by Y, where $Y = 50 - 2X - X^2$. Find the expected profit per foot.

Problem 8
The number of eggs laid on a tree leaf by an insect of a certain type is a Poisson random variable with parameter λ. However, such a random variable can only be observed if it is positive, since if it is 0 then we cannot know that such an insect was on the leaf. If we let Y denote the observed number of eggs, then

$$P(Y = x) = P(X = x | X > 0)$$

where X is Poisson with parameter λ. Find $E(Y)$.

Discrete random variables - some examples

Example 1

In a gambling game a person who draws a jack or a queen is paid \$15 and \$5 for drawing a king or an ace from an ordinary deck of fifty-two playing cards. A person who draws any other card pays \$4. If a person plays the game, what is the expected gain?

Example 2

A purchaser of electrical components buys them in lots of size 10. It is his policy to inspect 3 components randomly from a lot and to accept the lot only if all 3 components are nondefective. If 30 percent of the lots have 4 defective components and 70 percent have only 1, what proportion of lots does the purchaser reject?

Example 3

An urn contains N white and M black balls. Balls are randomly selected with replacement, one at a time, until a black one is obtained.

a. What is the probability that exactly n draws are needed?

a. What is the probability that at least k draws are needed?

Example 4

An electronic fuse is produced by five production lines in a manufacturing operation. The fuses are costly, are quite reliable, and are shipped to suppliers in 100-unit lots. Because testing is destructive, most buyers of the fuses test only a small number of fuses before deciding to accept or reject lots of incoming fuses. All five production lines usually produce only 2% defective fuses. Unfortunately, production line 1 suffered mechanical difficulty and produced 5% defectives during the previous month. This situation became known to the manufacturer after the fuses had been shipped. A customer received a lot produced last month and tested three fuses. One failed.

a. What is the probability that the lot produced on line 1?

b. What is the probability that the lot came from one of the four other lines?

Example 5

To determine whether or not they have a certain disease, 100 people are to have their blood tested. However, rather than testing each individual separately, it has been decided first to group the people in groups of 10. The blood samples of the 10 people in each group will be pooled and analyzed together. If the test is negative, one test will suffice for the 10 people. If the test is positive each of the 10 people will also be individually tested. Suppose the probability that a person has the disease is 0.10 for all people independently from each other. Compute the expected number of tests necessary for each group.

Example 6

A standard deck of 52 cards has been reduced to 51 because an unknown card was lost. As a reminder a standard deck of 52 cards consists of 13 clubs ($\clubsuit$), 13 spades ($\spadesuit$), 13 hearts ($\heartsuit$), and 13 diamonds ($\diamondsuit$).

a. A card is selected at random from the 51 cards, and we observe that this card is a diamond. Given this information, find the probability that the lost card is a club.

b. Four cards are selected at random without replacement from the 51 cards, and we observe that all are hearts. Given this information, find the probability that the lost card is a club.

Example 7
Suppose that $P(X = 0) = 1 - P(X = 1)$. If $E(X) = 3\text{Var}(X)$, find:

a. $P(X = 0)$.

b. $Var(3X)$.

Example 8
Let $X \sim b(n, p)$. Show that $Var(X) = np(1-p)$. *Hint:* First find $E[(X(X-1)]$ and then use $\sigma^2 = EX^2 - \mu^2$.

Example 9
Let X be a geometric random variable with probability of success p. The probability mass function of X is:

$$P(X = k) = (1 - p)^{k-1}p, \quad k = 1, 2, \cdots$$

Show that the probabilities sum up to 1.

Example 10
In a game of darts, the probability that a particular player aims and hits treble twenty with one dart is 0.40. How many throws are necessary so that the probability of hitting the treble twenty at least once exceeds 90%?

Example 11
Show that the mean and variance of the hypergeometric distribution are:

$$\mu = \frac{nr}{N}$$

$$\sigma^2 = \frac{nr(N - r)(N - n)}{N^2(N - 1)}$$

Example 12
Using steps similar to those employed to find the mean of the binomial distribution show that the mean of the negative binomial distribution is $\mu = \frac{r}{p}$. Also show that the variance of the negative binomial distribution is $\sigma^2 = \frac{r}{p}(\frac{1}{p} - 1)$ by first evaluating $E[X(X + 1)]$.

Example 13
Show that if we let $p = \frac{r}{N}$, the mean and the variance of the hypergeometric distribution can be written as $\mu = np$ and $\sigma^2 = np(1 - p)\frac{N-n}{N-1}$. What does this remind you?

Example 14
Suppose 5 cards are selected at random and without replacement from an ordinary deck of playing cars.

a. Construct the probability distribution of X, the number of clubs among the five cards.

b. Use (a) to find $P(X \leq 1)$.

Example 15
The manager of an industrial plant is planning to buy a new machine of either type A, or type B. The number of daily repairs X_A required to maintain a machine of type A is a random variable with mean and variance both equal to $0.10t$, where t denotes the number of hours of daily operation. The number of daily repairs X_B for a machine of type B is a random variable with mean and variance both equal $0.12t$. The daily cost of operating A is $C_A(t) = 10t + 30X_A^2$, and for B is $C_B(t) = 8t + 30X_B^2$. Assume that the repairs take negligible time and that each night the machines are tuned so that they operate essentially like new machines at the start of the next day. Which machine minimizes the expected daily cost if a workday consists of

a. 10 hours

b. 20 hours

Discrete random variables - some examples (solutions)

Example 1
Let X be the payoff:

Card	$X(\$)$	$P(X)$
Jack	15	$\frac{4}{52}$
Queen	15	$\frac{4}{52}$
King	5	$\frac{4}{52}$
Ace	5	$\frac{4}{52}$
Else	-4	$\frac{36}{52}$

Therefore the expected gain is:
$E(X) = 15\frac{4}{52} + 15\frac{4}{52} + 5\frac{4}{52} + 5\frac{4}{52} - 4\frac{36}{52} = \$\frac{16}{52}$.

Example 2
This is hypergeometric problem. The probability of accepting the lot is the probability that all 3 are found nondefective. Therefore the probability that the lot is rejected is 1-P(accept the lot).

$$P(\text{accept}) = 0.30\frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}} + 0.70\frac{\binom{1}{0}\binom{9}{3}}{\binom{10}{3}} = 0.54.$$

So, P(reject)= 1−P(accept)=1-0.54=0.46.

Example 3
This is geometric problem with $p = \frac{M}{M+N}$.

a. $$P(X = n) = (\frac{N}{N+M})^{n-1}\frac{M}{N+M} = \frac{MN^{n-1}}{(M+N)^n}.$$

b.

$$P(X \geq k) = \sum_{x=k}^{\infty}(\frac{N}{N+M})^{x-1}\frac{M}{N+M} =$$
$$\frac{M}{M+N}\sum_{x=k}^{\infty}(\frac{N}{M+N})^{x-1} =$$
$$\frac{M}{M+N}[(\frac{N}{M+N})^{k-1} + (\frac{N}{M+N})^{k} + \cdots] = (\frac{N}{M+N})^{k-1}.$$

Example 4
This is a binomial problem together with some probability (law of total probability and Bayes' theorem).
Let L_i be the event that lot was produced on line i.
Let X be the number of defective fuses among the 3 selected.

a.

$$P(L_1/X=1) = \frac{P(L_1 \cap X=1)}{P(X=1)} =$$
$$\frac{P(X=1/L_1)P(L_1)}{P(X=1/L_1)P(L_1) + P(X=1/L_2)P(L_2) + P(X=1/L_3)P(L_3) + P(X=1/L_4)P(L_4) + P(X=1/L_5)P(L_5)} =$$
$$\frac{\binom{3}{1}0.05^1 10.95^2\frac{1}{5}}{\binom{3}{1}0.05^1 0.95^2\frac{1}{5} + \binom{3}{1}0.02^1 0.98^2\frac{1}{5} + \binom{3}{1}0.02^1 0.98^2\frac{1}{5} + \binom{3}{1}0.02^1 0.98^2\frac{1}{5} + \binom{3}{1}0.02^1 0.98^2\frac{1}{5}} \Rightarrow$$
$$P(L_1/X=1) = 0.37.$$

b. The probability that the lot came from one of the other 4 lines is:

$$P(L_2 \cup L_3 \cup L_4 \cup L_5/X=1) = 1 - P(L_1/X=1) = 1 - 0.37 = 0.63.$$

Example 5
Let X be the number of tests needed for each group of 10 people. Then, if nobody has the disease 1 test is enough. But if the test is positive then there will be 11 test $(1+10)$. The probability distribution of X is:

X	$P(X)$
1	$\binom{10}{0}0.10^0 0.90^{10} = 0.90^{10}$
11	$1-\binom{10}{0}0.10^0 0.90^{10} = 1-0.90^{10}$

Therefore the expected number of tests is:
$E(X) = 1(0.90)^{10} + 11(1-0.90^{10} = 7.51.$

Example 6
This is a hypergeometric problem. Let's define the events:
D: Selected card is $\diamondsuit$ (part a)
$H4$: All 4 selected cards are $\heartsuit$ (part b)
L_c: Lost card is $\clubsuit$
L_s: Lost card is $\spadesuit$
L_h: Lost card is $\heartsuit$
L_d: Lost card is $\diamondsuit$

a. We need to find $P(L_c/D)$. This equal to:

$$\begin{aligned} P(L_c/D) = \frac{P(L_c \cap D)}{P(D)} &= \\ \frac{P(L_c \cap D)}{P(L_c \cap D) + P(L_s \cap D) + P(L_h \cap D) + P(L_d \cap D)} &= \\ \frac{P(D/L_c)P(L_c)}{P(D/L_c)P(L_c) + P(D/L_s)P(L_s) + P(D/L_h)P(L_h) + P(D/L_d)P(L_d)} &= \\ \frac{\frac{13}{51}\frac{1}{4}}{\frac{13}{51}\frac{1}{4} + \frac{13}{51}\frac{1}{4} + \frac{13}{51}\frac{1}{4} + \frac{12}{51}\frac{1}{4}} = \frac{13}{51} = 0.2549. \end{aligned}$$

b. We need to find $P(L_c/H4)$. This equal to:

$$\begin{aligned} P(L_c/H4) = \frac{P(L_c \cap H4)}{P(H4)} &= \\ \frac{P(L_c \cap H4)}{P(L_c \cap H4) + P(L_s \cap H4) + P(L_h \cap H4) + P(L_d \cap H4)} &= \\ \frac{P(H4/L_c)P(L_c)}{P(H4/L_c)P(L_c) + P(H4/L_s)P(L_s) + P(H4/L_h)P(L_h) + P(H4/L_d)P(L_d)} &= \\ \frac{\frac{\binom{13}{4}\binom{38}{0}}{\binom{51}{4}}\frac{1}{4}}{\frac{\binom{13}{4}\binom{38}{0}}{\binom{51}{4}}\frac{1}{4} + \frac{\binom{13}{4}\binom{38}{0}}{\binom{51}{4}}\frac{1}{4} + \frac{\binom{12}{4}\binom{39}{0}}{\binom{51}{4}}\frac{1}{4} + \frac{\binom{13}{4}\binom{38}{0}}{\binom{51}{4}}\frac{1}{4}} &= \\ \frac{\binom{13}{4}}{\binom{13}{4} + \binom{13}{4} + \binom{12}{4} + \binom{13}{4}} = 0.2708. \end{aligned}$$

Example 7
Here, X takes on two values 0 or 1. Let $P(X=1) = p$. The probability distribution of X is:

X	$P(X)$
0	1-p
1	p

a. $E(X) = 0(1-p) + 1(p) \Rightarrow E(X) = p.$
$Var(X) = 0^2(1-p) + 1^2(p) - P^2 \Rightarrow Var(X) = p - p^2.$
We know that $E(X) = 3Var(X) \Rightarrow p = 3(p-p^2) \Rightarrow p = \frac{2}{3}.$
Therefore $P(X=0) = 1 - P(X=1) = 1 - \frac{2}{3} \Rightarrow P(X=0) = \frac{1}{3}.$

b. $Var(3X) = 9Var(X) = 9(0^2\frac{1}{3} + 1^2\frac{2}{3}) - (\frac{2}{3})^2 \Rightarrow Var(3X) = 2.$

Example 8
Let $X \sim b(n,p)$. Show that $Var(X) = np(1-p)$. We start with $E[(X(X-1)]$ and then use $\sigma^2 = EX^2 - \mu^2$.

$$
\begin{aligned}
EX(X-1) = \sum_{x=0}^{n} x(x-1)\binom{n}{x} p^x (1-p)^{n-x} \quad &= \\
\text{Since for } x=0,\ x=1 \text{ the expression is zero, we can begin from } x=2: \\
\sum_{x=2}^{n} x(x-1)\frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} \quad &= \\
\text{We write } \ x! = x(x-1)(x-2)! \text{ to cancell } x(x-1) \text{ from the numerator} \\
\sum_{x=2}^{n} x(x-1)\frac{n!}{(n-x)!(x(x-1)(x-2)!} p^x (1-p)^{n-x} = \sum_{x=2}^{n} \frac{n!}{(n-x)!(x-2)!} p^x (1-p)^{n-x} \quad &= \\
\text{We factor outside from the summation } n(n-1)p^2. \\
n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(n-x)!(x-2)!} p^x - 2(1-p)^{n-x} \quad &= \\
\text{Now we let } y = x-2: \\
n(n-1)p^2 \sum_{y=0}^{n-2} \frac{(n-2)!}{(n-y-2)!(y)!} p^y (1-p)^{n-y-2} \quad &\Rightarrow \\
EX(X-1) = n(n-1)p^2 \sum_{y=0}^{n-2} \binom{n-2}{y} p^y (1-p)^{n-y-2} \quad (1)
\end{aligned}
$$

We can see now that $Y \sim b(n-2,p)$, that is, Y is a binomial random variable with $n-2$ number of trials and probability of success p. Therefore

$$
\sum_{y=0}^{n-2} \binom{n-2}{y} p^y (1-p)^{n-y-2} = 1.
$$

And expression (1) can be written as $EXX(X-1) = n(n-1)p^2$. To find the variance we use $\sigma^2 = EX^2 - \mu^2$:

$$
\begin{aligned}
EX(X-1) = n(n-1)p^2 \Rightarrow EX^2 - EX = n(n-1)p^2 \quad &\Rightarrow \\
EX^2 = n(n-1)p^2 + EX \quad &\Rightarrow \\
\text{We know that } EX = \mu = np: \\
EX^2 = n(n-1)p^2 + np. \\
\text{And the variance is:} \\
\sigma^2 = EX^2 - \mu^2 = n(n-1)p^2 + np - (np)^2 \quad &\Rightarrow \\
\sigma^2 = n(n-1)p^2 + np(1-np) = np[(n-1)p + 1 - np] = np(np - p + 1 - np) \quad &\Rightarrow \\
\sigma^2 = np(1-p).
\end{aligned}
$$

Example 9
Let X be a geometric random variable with probability of success p. The probability mass function of X is:

$$
P(X = x) = (1-p)^{x-1}p, \quad k = 1, 2, \cdots
$$

The above summation must equal to 1 since $P(x)$ is a probability mass function. This is an easy proof since we have an infinite sme of geometric series with common ratio $1-p$.

$$
\sum_{x=1}^{\infty} (1-p)^{x-1}p = p + (1-p)p^2 + (1-p)^2p^3 + \cdots = p\frac{1}{1-(1-p)} = 1
$$

Example 10
We want $P(X \geq 1) > 0.90$. We can write this as: $1 - P(X = 0) > 0.90 \Rightarrow 1 - \binom{n}{0} 0.40^0 (1-0.40)^n > 0.90 \Rightarrow 1 - 0.60^n > 0.90 \Rightarrow 0.60^n < 0.10 \Rightarrow n > \frac{log(0.10)}{log(0.60)} \Rightarrow n > 4.51$ or $n > 5$.

Example 11
We know that the probability mass function of the hypergeometric probability distribution is:

$$P(X=x)=\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

Therefore the expected value of X is:

$$\mu=E(X)=\sum_{x=0}^{n} x\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}=\sum_{x=1}^{n}\frac{r!}{(r-x)!(x-1)!}\frac{\binom{N-r}{n-x}}{\binom{N}{n}}=$$

$$\frac{r}{\binom{N}{n}}\sum_{x=1}^{n}\frac{(r-1)!}{(r-x)!(x-1)!}\binom{N-r}{n-x}=\frac{r}{\binom{N}{n}}\sum_{x=1}^{n}\binom{r-1}{x-1}\binom{N-r}{n-x}\Rightarrow$$

Let $y=x-1$

$$E(X)=\frac{r}{\binom{N}{n}}\sum_{y=0}^{n-1}\binom{r-1}{y}\binom{N-r}{n-y-1}\qquad(1)$$

At this point we have to use a result from combinatorial analysis from the beginning of the course (hw1 ex6). Here is the result:

$$\binom{n+m}{r}=\binom{n}{0}\binom{m}{r}+\binom{n}{1}\binom{m}{r-1}+\cdots\binom{n}{r}\binom{m}{0}.$$

Or we can write it as:

$$\binom{n+m}{r}=\sum_{k=0}^{r}\binom{n}{k}\binom{m}{r-k}.$$

Therefore the summationof expression (1) above is equal to $\binom{N-1}{n-1}$. Now we can write expression (1) as:

$$E(X)=\frac{r}{\binom{N}{n}}\sum_{y=0}^{n-1}\binom{r-1}{y}\binom{N-r}{n-y-1}=\frac{r}{\binom{N}{n}}\binom{N-1}{n-1}=$$

$$\frac{r(N-n)!n!}{N!}\frac{(N-1)!}{(N-n)!(n-1)!}=\frac{rn(n-1)!(N-1)!}{N(N-1)!(n-1)!}\Rightarrow$$

$$E(X)=r\frac{n}{N}.$$

To find the variance we start with $EX(X-1)$:

$$EX(X-1)=\sum_{x=0}^{n}x(x-1)\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}=\sum_{x=2}^{n}x(x-1)\frac{r!}{(r-x)!x!}\frac{\binom{N-r}{n-x}}{\binom{N}{n}}=$$

$$\frac{r(r-1)}{\binom{N}{n}}\sum_{x=2}^{n}\frac{(r-2)!}{(r-x)!(x-2)!}\binom{N-r}{n-x}=\frac{r(r-1)}{\binom{N}{n}}\sum_{x=2}^{n}\binom{r-2}{x-2}\binom{N-r}{n-x}\Rightarrow$$

Let $y=x-2$

$$EX(X-1)=\frac{r(r-1)}{\binom{N}{n}}\sum_{y=0}^{n-2}\binom{r-2}{y}\binom{N-r}{n-y-2}\qquad(2)$$

Using the same result as before, the summation of expression (2) is equal to $\binom{N-2}{n-2}$. Now we can write expression (2) as:

$$EX(X-1)=\frac{r(r-1)}{\binom{N}{n}}\sum_{y=0}^{n-2}\binom{r-2}{y}\binom{N-r}{n-y-2}=\frac{r(r-1)}{\binom{N}{n}}\binom{N-2}{n-2}=$$

$$\frac{r(r-1)(N-n)!n!}{N!}\frac{(N-2)!}{(N-n)!(n-2)!}=\frac{r(r-1)n(n-1)(n-2)!(N-2)!}{N(N-1)(N-2)!(n-2)!}\Rightarrow$$

$$EX(X-1)=\frac{r(r-1)n(n-1)}{N(N-1)}\Rightarrow EX^2-EX=\frac{r(r-1)n(n-1)}{N(N-1)}\Rightarrow$$

$$EX^2=\frac{r(r-1)n(n-1)}{N(N-1)}+r\frac{n}{N}.$$

Therefore the variance is:

$$\sigma^2 = EX^2 - \mu^2 = \frac{r(r-1)n(n-1)}{N(N-1)} + r\frac{n}{N} - (r\frac{n}{N})^2 = \frac{r(r-1)n(n-1)N + rnN(N-1) - r^2n^2(N-1)}{N^2(N-1)} =$$

$$nr\frac{(r-1)(n-1)N + N(N-1) - rn(N-1)}{N^2(N-1)} = nr\frac{Nrn - Nr - Nn + N + N^2 - Nrn - N + rn}{N^2(N-1)} =$$

$$nr\frac{-Nr - Nn + N^2 + rn}{N^2(N-1)} = nr\frac{N(N-r) - n(N-r)}{N^2(N-1)} \Rightarrow \sigma^2 = \frac{nr(N-r)(N-n)}{N^2(N-1)}$$

Example 12
We know that the probability mass function of a negative binomial random variable is:

$$P(X=x) = \binom{x-1}{r-1}p^r(1-p)^{x-r}, \qquad x = r, r+1, r+2, \cdots$$

where X is the number of trials required until r successes occur. Therefore the expected value of X is:

$$\mu = E(X) = \sum_{x=r}^{\infty} x\binom{x-1}{r-1}p^r(1-p)^{x-r} =$$

$$\sum_{x=r}^{\infty} x\frac{(x-1)!}{(x-r)!(r-1)!}p^r(1-p)^{x-r} = \sum_{x=r}^{\infty}\frac{x!}{(x-r)!(r-1)!}p^r(1-p)^{x-r} =$$

Factor outside of the summation $\frac{r}{p}$:

$$\frac{r}{p}\sum_{x=r}^{\infty}\frac{x!}{(x-r)!r!}p^{r+1}(1-p)^{x-r} = \frac{r}{p}\sum_{x=r}^{\infty}\binom{x}{r}p^{r+1}(1-p)^{x-r} =$$

Now let $y = x+1$:

$$\frac{r}{p}\sum_{y=r+1}^{\infty}\binom{y-1}{r}p^{r+1}(1-p)^{y-r-1}.$$

We observe now that Y is a negative binomial random variable that represents the number of trials required until $r+1$ successes occur. Therefore the summation is equal to 1 (it is the sum of all the probabilities). And we have that $\mu = E(X) = \frac{r}{p}$.

To find the variance we start with $EX(X+1)$ whic is equal to:

$$= EX(X+1) = \sum_{x=r}^{\infty} x(x+1)\binom{x-1}{r-1}p^r(1-p)^{x-r} =$$

$$\sum_{x=r}^{\infty} x(x+1)\frac{(x-1)!}{(x-r)!(r-1)!}p^r(1-p)^{x-r} = \sum_{x=r}^{\infty}\frac{(x+1)!}{(x-r)!(r-1)!}p^r(1-p)^{x-r} =$$

Factor outside of the summation $\frac{r(r+1)}{p^2}$:

$$\frac{r(r+1)}{p^2}\sum_{x=r}^{\infty}\frac{(x+1)!}{(x-r)!(r+1)!}p^{r+2}(1-p)^{x-r} = \frac{r(r+1)}{p^2}\sum_{x=r}^{\infty}\binom{x+1}{r+1}p^{r+2}(1-p)^{x-r} =$$

Now let $y = x+2$:

$$\frac{r(r+1)}{p^2}\sum_{y=r+2}^{\infty}\binom{y-1}{r+1}p^{r+2}(1-p)^{y-r-2}$$

We observe now that Y is a negative binomial random variable that represents the number of trials required until $r+2$ successes occur. Therefore the summation is equal to 1 (it is the sum of all the probabilities). And we have that $EX(X+1) = \frac{r(r+1)}{p^2}$.
We can find EX^2:

$$EX(X+1) = \frac{r(r+1)}{p^2} \Rightarrow EX^2 + EX = \frac{r(r+1)}{p^2} \qquad \Rightarrow$$

$$EX^2 = \frac{r(r+1)}{p^2} - EX \quad \Rightarrow EX^2 = \tfrac{r(r+1)}{p^2} - \tfrac{r}{p}$$

And now the variance:

$$\sigma^2 = EX^2 - \mu^2 = \frac{r(r+1)}{p^2} - \frac{r}{p} - (\frac{r}{p})^2 = \frac{r(r+1) - rp - r^2}{p^2} \Rightarrow \sigma^2 = \frac{r(1-p)}{p^2}.$$

Example 13
The mean and variance of the hypergeometric distribution are:

$$\mu = \frac{nr}{N}$$

$$\sigma^2 = \frac{nr(N-r)(N-n)}{N^2(N-1)}$$

If we let $p = \frac{r}{N}$, then the previous two expressions will become:

$$\mu = \frac{nr}{N} \Rightarrow \mu = np.$$

$$\sigma^2 = \frac{nr(N-r)(N-n)}{N^2(N-1)} = n\frac{r}{N}\frac{N-r}{N}\frac{N-n}{N-1} \quad =$$
$$n\frac{r}{N}(1-\frac{r}{N})\frac{N-n}{N-1} \Rightarrow \sigma^2 = np(1-p)\frac{N-n}{N-1}$$

This reminds us the mean and variance of the binomial distribution, specially when n is much smaller than N then $\frac{N-n}{N-1} \approx 1$.

Example 14
This is a hypergeometric probability distribution problem:

a.

X	$P(X)$
0	$\frac{\binom{13}{0}\binom{39}{5}}{\binom{52}{5}} = 0.2215$
1	$\frac{\binom{13}{1}\binom{39}{4}}{\binom{52}{5}} = 0.4114$
2	$\frac{\binom{13}{2}\binom{39}{3}}{\binom{52}{5}} = 0.2743$
3	$\frac{\binom{13}{3}\binom{39}{2}}{\binom{52}{5}} = 0.0815$
4	$\frac{\binom{13}{4}\binom{39}{1}}{\binom{52}{5}} = 0.0107$
5	$\frac{\binom{13}{5}\binom{39}{0}}{\binom{52}{5}} = 0.0005$

b. Using the probability distribution we constructed in part (a) we have:
$P(X \leq 1) = P(X=0) + P(X=1) = 0.2215 + 0.4114 = 0.6329$.

Example 15
We are given: $\mu_A = 0.10t, \sigma_A^2 = 0.10t$ and $\mu_B = 0.12t, \sigma_B^2 = 0.12t$. The cost functions are $C_A = 10t + 30X_A^2$, and $C_B = 8t + 30X_B^2$.

a. For $t = 10$:
$E(C_A) = E(10t+30X_A^2) = E(100+30X_A^2) = 100+30EX_A^2 = 100+30(\sigma_A^2+\mu_A^2) = 100+30[0.1(10)+(0.1(10))^2] = 160$.
$E(C_B) = E(8t+30X_B^2) = E(80+30X_B^2) = 80+30EX_B^2 = 80+30(\sigma_B^2+\mu_B^2) = 80+30[0.12(10)+(0.12(10))^2] = 159.2$.
Therefore machine B has less expected daily cost.

a. For $t = 20$:
$E(C_A) = E(10t+30X_A^2) = E(200+30X_A^2) = 200+30EX_A^2 = 200+30(\sigma_A^2+\mu_A^2) = 200+30[0.1(20)+(0.1(20))^2] = 380$.
$E(C_B) = E(8t+30X_B^2) = E(160+30X_B^2) = 160+30EX_B^2 = 160+30(\sigma_B^2+\mu_B^2) = 160+30[0.12(20)+(0.12(20))^2] = 404.8$.
Therefore machine A has less expected daily cost.