Variance of the negative binomial N variable with parameters \( r \) and \( p \):

\[ \mathbb{P}(N = m) = \binom{m-1}{r-1} p^r (1-p)^{m-r} \]

\[ E(N) = \frac{r}{p} \leq \text{last time.} \]

Stop at \( r \) outs.

Compute: \( s \geq 1 \)
\[ E(N^5) = \sum_{m=r}^{\infty} m^5 \binom{m-1}{r-1} p^r (1-p)^{m-r} = \frac{r}{p} \sum_{m=r+1}^{\infty} (m-1)^{r-1} \binom{m-1}{r-1} p^r (1-p)^{m-r} \]

Set \( m = m+1 \)

\[ = \frac{r}{p} E((M-1)^{r-1}) \]

where \( M \) is negative binomial with parameters \( r+1, p \).

\( M \) includes the \( r+1 \) out of \( r \).
\[ E(N^2) = \frac{r}{\rho} E(M-1) = \frac{r}{\rho} \left( \frac{r+1}{\rho} - 1 \right) \]

\[ \text{var}(N) = E(N^2) - (E(N))^2 = \frac{r}{\rho} \left( \frac{r+1}{\rho} - 1 \right) - \left( \frac{r}{\rho} \right)^2 = \frac{r^2}{\rho^2} + \frac{r}{\rho^2} - \frac{r}{\rho} - \frac{r^2}{\rho^2} = \frac{r(1-\rho)}{\rho^2} \]

\[ E(N) = \frac{r}{\rho}, \quad \text{var}(N) = \frac{r(1-\rho)}{\rho^2} \]
The case when $r = 1$

$$P(N = m) = \binom{m-1}{r-1} p^r (1-p)^{m-r} = p^r (1-p)^{m-1}.$$
This case (negative binomial for \( r = 1 \)) is called the geometric variable.

Example: How to cheat at a casino

1. Have a strategy for beating the odds.
2. Don't get caught!

Assuming you have 1 and 2, how much should you bet?

\$17

Assume if we bet \$17, we win \( \lambda \) with probability \( p \) at one game, and lose \$17 with probability \( 1 - p \).
\[ p \leq \frac{1}{2}, \text{ we should not play, if we are playing for profit}. \]

Suppose we start with \$N. Suppose we bet \$aN.

Expected value after 1 game: \[ E = p(1+a)N + (1-p)(1-a)N = pN + aN \]

\[ = N + aN(2p-1). \]

The expected value after 1 game is maximized if I bet everything, assuming \( p \geq \frac{1}{2} \). (\( 2p-1 > 0 \).)
After playing the game \( k \) times: Expected value:

\[ N \left( 1 + a (2p-1) \right)^k \]

Still, the best strategy is to always bet everything.

But suppose I want to play indefinitely. What is the probability under the strategy I just outlined that after playing the game \( n \) times, I lose everything? (If I keep betting everything, all I need is to lose once, and I lose everything.)
Probability I lose everything in game \( n \) is geometrically distributed:

\[
p^{n-1}(1-p)
\]

(\text{caution: } p \text{ and } 1-p \text{ reversed from previous notation})

\[
\sum_{n=1}^{\infty} p^{n-1}(1-p) = 1,
\]

So if I keep playing indefinitely, I am sure to lose eventually.

Then this is the wrong strategy!
What is the right strategy for indefinite play?

Idea: if we take the logarithm of our net worth at each stage, then

\[ \ln((1+a)N) = \ln(1+a) + \ln N \]

\[ \ln((1-a)N) = \ln N + \ln(1-a) \]

Expected value of \( \ln(\text{Net worth}) \) after one game:

\[ \ln N + (p \ln(1+a) + (1-p) \ln(1-a)) \]

Best strategy for indefinitely lasting play
0 = \frac{d}{da} \left( p \ln(1+a) + (1-p) \ln(1-a) \right) = \\
= \frac{p}{1+a} - \frac{(1-p)}{1-a} \\
p(1-a) = (1+a)(1-p) \\
p - pa = 1 - p + a(1 - p) \\
2p - 1 = a \\
\text{We should set} \\
(2p - 1)N \text{ if our net worth} \\
\geq N \text{ and probability of}
At 8:1: If I bet an and win, I receive εaN, if I lose, I lose an.

Expected value of the logarithm after one game:

\[
\nu_1 = \ln N + p \ln (1 + \varepsilon z) + (1 - p) \ln (1 - a).
\]

0 = \frac{d}{da} \left( p \ln (1 + \varepsilon a) + (1 - p) \ln (1 - a) \right) =
\[
\frac{3\rho}{1+3a} - \frac{1-\rho}{1-a} = \frac{1-\rho}{1+3a}
\]

\[
\frac{3\rho}{1+3a} = \frac{1-\rho}{1-a}
\]

\[
3\rho - 3\rho a = 1 + 3a - 1 + 3a
\]

\[
\rho = \frac{3\rho + 1 - 1}{3}
\]

Don't play unless

\[
\frac{3\rho - 1}{\rho} > 0
\]

\[
3\rho > 1 - \rho
\]

This means

\[
\rho > \frac{1-\rho}{\rho}
\]

you are beating the odds.
Hypergeometric distribution (Caution: geometric binomial \text{not} hypergeometric)

"Like binomial, but without replacement"

Scenario: We have an urn containing $N$ balls of which $w$ are white and $N-w$ are black.

We choose a sample of $n$ balls (without replacement).

$X =$ the number of white balls selected.

Probability mass function:
\[ \Pr \{ X = i \} = \frac{\binom{m}{i} \binom{N-m}{m-i}}{\binom{N}{m}}. \]

Values: \( i = 0, \ldots, m. \)

\[ \left( 1 = \sum_{i=0}^{m} \Pr \{ X = i \} = \sum_{i=0}^{m} \frac{\binom{m}{i} \binom{N-m}{m-i}}{\binom{N}{m}} \right) \]

So \( \binom{N}{m} = \sum_{i=0}^{m} \binom{m}{i} \binom{N-m}{m-i} \).

**Comment:** The hypergeometric distribution has a symmetry by exchanging \( m \) and \( m. \)
\[
\frac{\binom{m}{i} \binom{N-m}{m-i}}{\binom{N}{m}} = \binom{m}{i} \frac{N-m}{m-i} \binom{N}{m}
\]

**Probabilistic argument:**

Think chosen = paint and

a symmetrical role.

Throw the ball back after each selection.

- If we choose with replacement, we lose the symmetry, and the distribution becomes binomial.
\[ P(Y = n) = \binom{m}{n} \left( \frac{m}{N} \right)^n \left( \frac{N-m}{N} \right)^{m-n} \]

If \( N \to \infty \), \( \frac{m}{N} = p \) is constant,

hypergeometric \( \to \) binomial.

---

**Expected value and variance of hypergeometric variable.**

Let \( X \) be a hypergeometric variable with parameters \( m, N, n \).
\[ E(X^k) = \sum_{i=0}^{m} i^k \binom{m}{i} \left( \frac{N-m}{m-i} \right) / \binom{N}{m} \quad \text{if } k \geq 1. \]

\[ \binom{m}{i} = m \binom{m-1}{i-1} \quad \binom{N}{m} = N \binom{N-1}{m-1} \]

\[ E(X^k) = \frac{m^m}{N} \sum_{i=1}^{m} i^{k-1} \binom{m-1}{i-1} \left( \frac{N-m}{m-i} \right) / \binom{N-1}{m-1} = \]

\[ = \frac{m^m}{N} \sum_{j=0}^{m-1} (j+1)^{k-1} \binom{m-1}{j} \left( \frac{N-m}{m-1-j} \right) / \binom{N-1}{m-1} \quad \text{if } j \leq k \]
\[ \frac{mm}{N} E \left( (Y+1)^{k-1} \right) \]

where \( Y \) is hypergeometric with parameters \( m \cdot 1, N-1, m-1 \).

Apply to \( k = 1 \):

\[ E(X) = \frac{m \cdot m}{N} \]

\[ E\left(X^2\right) = \frac{m \cdot m}{N} E \left( Y+1 \right) = \]

\[ E(Y) = \frac{(m-1)(m-1)}{N-1} \]
\[ \frac{m \cdot n}{N} \left( 1 + \frac{(m-1)(n-1)}{N-1} \right) \]

\[ \text{var}(X) = E(X^2) - (E(X))^2 = \frac{m \cdot m}{N} \left( 1 + \frac{(m-1)(n-1)}{N-1} - \frac{m \cdot m}{N} \right) \]

\[ \text{if } N \gg 0 \]
\[ \frac{m}{N} = p \text{ constant,} \]
\[ \text{these will nearly cancel out (\sim known)} \]
A little digression on jointly distributed discrete random variables $X, Y$:

$$E(X + Y) = E(X) + E(Y)$$

on the same sample space.

$$\text{var}(X + Y) = E((X + Y)^2) - (E(X) + E(Y))^2 =$$

$$= E(X^2 + 2XY + Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2 =$$

$$= E(X^2) + 2E(XY) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2 =$$

$$\text{var}(X) + \text{var}(Y) + 2\left(\frac{E(XY) - E(X)E(Y)}{}ight).$$
\[ \text{covariance} \]

\[ \text{cov} (X, Y) = E(KY) - E(X) \cdot E(Y) \]

\[ \text{Correlation:} \quad \text{corr} (X, Y) = \frac{\text{cov} (X, Y)}{\sqrt{\text{var} (X) \cdot \text{var} (Y)}}. \]

If \( \text{corr} (X, Y) = 0 \), then \( \text{var} (X + Y) = \text{var} (X) + \text{var} (Y) \).

\( X, Y \) are uncorrelated.

**Theorem:** If \( X, Y \) are independent then they are uncorrelated.
Recall: Two random variables $X$, $Y$ are **independent** if the event

$$\{X \leq x\} \text{ and } \{Y \leq y\}$$

are independent for any choices of $x, y$.

For **discrete variables**, this is the same as requiring that

$$\{X = x\} \text{ and } \{Y = y\}$$

be independent for all choices of $x, y$. 
If this happens,

\[ \text{Cov}(X, Y) = \sum_z p_{XY} = \sum_z x \cdot p_{X=x, Y=y} - \sum_x \sum_y x \cdot p_{X=x, Y=y} \]

\[ = \sum_z p_{XY} = \sum_x \sum_y x \cdot p_{X=x, Y=y} \]

\[ \text{independent} \]

\[ = \sum_z p_{XY} = \sum_z x \cdot p_{X=x, Y=y} \]

\[ \text{equal} \]

\[ = 0. \]
Exam next Wednesday: 5 questions, a & b

- independent event

- independent event & Bayes formula:

the insurance company

rating accident-prone drivers,

follow up: probability of accident

in 1st and 2nd year.

- Discrete random variables - everything,

binomial, Poisson, negative binomial,

hypergeometric + expectations and
HW: 1) Suppose a batch of 100 items contains 6 that are defective and 94 which are good. $X =$ number of defective items in a randomly drawn sample of 10 items from the batch, find

(a) $P(X = 0)$
(b) $P(X > 2)$. 
2. Suppose I am shooting free throws and the game is that I stop when I miss 3 (not necessarily consecutively). Then another player gets a chance to play. Assuming on average I take 10 shots in the game before I have to stop, what is my free throw percentage?

3. There are k types of coupons. Independently
of the types of previously collected coupons, each new coupon is of type \( i \) with probability \( p_i \), \( \sum_{i=1}^{n} p_i = 1 \). If \( m \) coupons are collected, find the expected number of distinct types of coupons that appear in the set. [Hint: Additivity of expected value.]