Chapters 1–2

Discrete random variables
Permutations
Binomial and related distributions
Expected value and variance

Prof. Tesler

Math 283
Fall 2018
Sample spaces and events

- Flip a coin 3 times. The possible outcomes are:
  
  \[\text{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}\]

- The sample space is the set of all possible outcomes:
  \[S = \{\text{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}\}\]

- An event is any subset of \(S\).
  The event that there are exactly two heads is
  \[A = \{\text{HHT, HTH, THH}\}\]

- The probability of heads is \(p\) and of tails is \(q = 1 - p\). The flips are independent, which gives these probabilities for each outcome:
  \[
  \begin{align*}
  P(\text{HHH}) &= p^3 \\
  P(\text{HHT}) &= P(\text{HTH}) = P(\text{THH}) = p^2q \\
  P(\text{TTT}) &= q^3 \\
  P(\text{HTT}) &= P(\text{THT}) = P(\text{TTH}) = pq^2 
  \end{align*}
  \]
  These are each between 0 and 1, and they add up to 1.

- The probability of an event is the sum of probabilities of its outcomes:
  \[
  P(A) = P(\text{HHT}) + P(\text{HTH}) + P(\text{THH}) = 3p^2q
  \]
A random variable $X$ is a function assigning a real number to each outcome.

Let $X$ denote the number of heads:

$X(\text{HHH}) = 3 \quad X(\text{HHT}) = X(\text{HTH}) = X(\text{THH}) = 2$

$X(\text{TTT}) = 0 \quad X(\text{HTT}) = X(\text{THT}) = X(\text{TTH}) = 1$

The range of $X$ is $\{0, 1, 2, 3\}$.

That range is a discrete set as opposed to a continuum, such as all real numbers $[0, 3]$. So $X$ is a discrete random variable.

The discrete probability density function (pdf) or probability mass function (pmf) is $p_X(k) = P(X = k)$, defined for all real numbers $k$:

$p_X(0) = q^3 \quad p_X(1) = 3pq^2 \quad p_X(2) = 3p^2q \quad p_X(3) = p^3$

$p_X(k) = 0$ otherwise, e.g. $p_X(2.5) = 0 \quad p_X(-1) = 0$

Use capital letters ($X$) for random variables and lowercase ($k$) to stand for numeric values.
Measure several properties at once using multiple random variables:

\[ X = \text{# heads} \]
\[ Y = \text{position of first head (1,2,3) or 4 if no heads} \]

\[ HHH: X = 3, \ Y = 1 \]
\[ HHT: X = 2, \ Y = 1 \]
\[ HTH: X = 2, \ Y = 1 \]
\[ HTT: X = 1, \ Y = 1 \]
\[ THH: X = 2, \ Y = 2 \]
\[ THT: X = 1, \ Y = 2 \]
\[ TTH: X = 1, \ Y = 3 \]
\[ TTT: X = 0, \ Y = 4 \]

Reorganize that in a two dimensional table:

| \( Y \) = 1 | \( X = 0 \) | \( X = 1 \) | \( X = 2 \) | \( X = 3 \) |
|-------------|-------------|-------------|-------------|
| \( Y = 1 \) | \( HTT \) | \( HHT, HTH \) | \( HHH \) |
| \( Y = 2 \) | \( THT \)  | \( THH \)    |            |
| \( Y = 3 \) |            | \( TTH \)  |            |
| \( Y = 4 \) |            |            | \( TTT \)  |
The (discrete) **joint probability density function** is 

\[ p_{X,Y}(x, y) = P(X = x, Y = y) : \]

<table>
<thead>
<tr>
<th>( p_{X,Y}(x, y) )</th>
<th>( x = 0 )</th>
<th>( x = 1 )</th>
<th>( x = 2 )</th>
<th>( x = 3 )</th>
<th>Total ( p_Y(y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y = 1 )</td>
<td>0</td>
<td>( pq^2 )</td>
<td>( 2p^2q )</td>
<td>( p^3 )</td>
<td>( p )</td>
</tr>
<tr>
<td>( y = 2 )</td>
<td>0</td>
<td>( pq^2 )</td>
<td>( p^2q )</td>
<td>0</td>
<td>( pq )</td>
</tr>
<tr>
<td>( y = 3 )</td>
<td>0</td>
<td>( pq^2 )</td>
<td>0</td>
<td>0</td>
<td>( pq^2 )</td>
</tr>
<tr>
<td>( y = 4 )</td>
<td>( q^3 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( q^3 )</td>
</tr>
<tr>
<td><strong>Total</strong> ( p_X(x) )</td>
<td>( q^3 )</td>
<td>3( pq^2 )</td>
<td>3( p^2q )</td>
<td>( p^3 )</td>
<td>1</td>
</tr>
</tbody>
</table>

It’s defined for all real numbers. It equals zero outside the table. 

**In table:** \( p_{X,Y}(3, 1) = p^3 \)  
**Not in table:** \( p_{X,Y}(1, -0.5) = 0 \)

**Row totals:** \( p_Y(y) = \sum_x p_{X,Y}(x, y) \)  
**Columns:** \( p_X(x) = \sum_y p_{X,Y}(x, y) \)

These are in the right and bottom margins of the table, so \( p_X(x) \), \( p_Y(y) \) are called **marginal densities** of the joint pdf \( p_{X,Y}(x, y) \).
Joint probability density — marginal density

<table>
<thead>
<tr>
<th>( p_{X,Y}(x, y) )</th>
<th>( x = 0 )</th>
<th>( x = 1 )</th>
<th>( x = 2 )</th>
<th>( x = 3 )</th>
<th>Total ( p_Y(y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y = 1 )</td>
<td>0</td>
<td>( pq^2 )</td>
<td>( 2p^2q )</td>
<td>( p^3 )</td>
<td>( p )</td>
</tr>
<tr>
<td>( y = 2 )</td>
<td>0</td>
<td>( pq^2 )</td>
<td>( p^2q )</td>
<td>0</td>
<td>( pq )</td>
</tr>
<tr>
<td>( y = 3 )</td>
<td>0</td>
<td>( pq^2 )</td>
<td>0</td>
<td>0</td>
<td>( pq^2 )</td>
</tr>
<tr>
<td>( y = 4 )</td>
<td>( q^3 )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( q^3 )</td>
</tr>
<tr>
<td>Total ( p_X(x) )</td>
<td>( q^3 )</td>
<td>( 3pq^2 )</td>
<td>( 3p^2q )</td>
<td>( p^3 )</td>
<td>1</td>
</tr>
</tbody>
</table>

Row totals

- **Row total for** \( y = 1 \):
  \[
pq^2 + 2p^2q + p^3 = p(q^2 + 2pq + p^2) = p(q + p)^2 = p \cdot 1^2 = p
  \]

- **Row total for** \( y = 2 \):
  \[
pq^2 + p^2q = pq(p + q) = pq \cdot 1 = pq
  \]

- Or, for \( y = 1, 2, 3 \), the probability that the first heads is flip \# \( y \) is
  \[
P(Y = y) = P(y - 1 \text{ tails followed by heads}) = q^{y-1}p
  \]
  and the probability of no heads is
  \[
P(Y = 4) = P(TTT) = q^3.
  \]
Conditional probability

Bob flips a coin 3 times and tells you that $X = 2$ (two heads), but no further information. What does that tell you about $Y$ (flip number of first head)?

The possible outcomes with $X = 2$ are $HHT$, $HTH$, $THH$, each with the same probability $p^2q$.

We’re restricted to three equally likely outcomes $HHT$, $HTH$, $THH$:  

- Probability $Y = 1$ is $2/3$ ($HHT$, $HTH$)
- Probability $Y = 2$ is $1/3$ ($THH$)
- Other values of $Y$ are not possible
You know that event $B$ holds. What’s the probability of event $A$?

The *conditional probability of $A$, given $B$, is*

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{P(A \cap B)}{P(B)}$$

The probability that $Y = 1$ given $X = 2$ is $P(Y = 1 \mid X = 2)$:

- The event $Y = 1$ is $A = \{HHH, HHT, HTH, HTT\}$.
- The event $X = 2$ is $B = \{HHT, HTH, THH\}$.

$$P(Y = 1 \mid X = 2) = \frac{P(X = 2 \text{ and } Y = 1)}{P(X = 2)} = \frac{P(\{HHT, HTH\})}{P(\{HHT, HTH, THH\})} = \frac{2p^2q}{3p^2q} = \frac{2}{3}$$
The conditional probability of \( A \), given \( B \), is

\[
P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{P(A \cap B)}{P(B)}
\]

The conditional probability that \( Y = y \) given that \( X = x \) is

\[
P(Y = y \mid X = x) = \frac{P(Y = y \text{ and } X = x)}{P(X = x)} = \frac{p_{X,Y}(x, y)}{p_X(x)}
\]

\[
P(Y = 1 \mid X = 2) = \frac{p_{X,Y}(2, 1)}{p_X(2)} = \frac{2p^2q}{3p^2q} = \frac{2}{3}
\]
Independent random variables

- In the previous example, knowing $X = 2$ affected the probabilities of the values of $Y$. So $X$ and $Y$ are dependent.

- Discrete random variables $U, V, W$ are independent if
  \[ P(U = u, V = v, W = w) = P(U = u)P(V = v)P(W = w) \]
  factorizes for all values of $u, v, w$, and dependent if there are any exceptions. This generalizes to any number of random variables.

- In terms of conditional probability, $X$ and $Y$ are independent if
  \[ P(Y = y | X = x) = P(Y = y) \]
  for all $x, y$ (with $P(X = x) \neq 0$).

Examples of independent random variables

- Let $U, V, W$ denote three flips of a coin, coded 0=tails, 1=heads.
- Let $X_1, \ldots, X_{10}$ denote the values of 10 separate rolls of a die.

Example of dependent random variables

- Drawing cards $U, V$ from a deck without replacement (so $V \neq U$).
Permutations of distinct objects

Permutations

Here are all the permutations of $A, B, C$:

$$ABC \quad ACB \quad BAC \quad BCA \quad CAB \quad CBA$$

- There are 3 items: $A, B, C$.
- There are 3 choices for which item to put first.
- There are 2 choices remaining to put second.
- There is 1 choice remaining to put third.
- Thus, the total number of permutations is $3 \cdot 2 \cdot 1 = 6$.

Factorials

- The number of permutations of $n$ distinct items is “$n$-factorial”:
  
  $$n! = n(n - 1)(n - 2) \cdots 1$$
  for integers $n = 1, 2, \ldots$
- $0! = 1$
Here are all the permutations of the letters of ALLELE:

<table>
<thead>
<tr>
<th>Permutation</th>
</tr>
</thead>
<tbody>
<tr>
<td>EEALLL</td>
</tr>
<tr>
<td>EELALL</td>
</tr>
<tr>
<td>EELLAL</td>
</tr>
<tr>
<td>EELLLA</td>
</tr>
<tr>
<td>EAELLL</td>
</tr>
<tr>
<td>EAELL</td>
</tr>
<tr>
<td>EALLEL</td>
</tr>
<tr>
<td>EALLLE</td>
</tr>
<tr>
<td>ELEALL</td>
</tr>
<tr>
<td>ELELAL</td>
</tr>
<tr>
<td>ELELLA</td>
</tr>
<tr>
<td>ELAEELL</td>
</tr>
<tr>
<td>ELAELL</td>
</tr>
<tr>
<td>EALLEA</td>
</tr>
<tr>
<td>ELLALE</td>
</tr>
<tr>
<td>ELLLEA</td>
</tr>
<tr>
<td>ELLLA</td>
</tr>
<tr>
<td>LEELLL</td>
</tr>
<tr>
<td>LAEELL</td>
</tr>
<tr>
<td>LAELLE</td>
</tr>
<tr>
<td>LALEE</td>
</tr>
<tr>
<td>LLAEEL</td>
</tr>
<tr>
<td>LLAEE</td>
</tr>
<tr>
<td>LLLAEED</td>
</tr>
</tbody>
</table>
There are $6! = 720$ ways to permute the subscripted letters $A_1, L_1, L_2, E_1, L_3, E_2$.

Here are all the ways to put subscripts on $EALLEL$:

- $E_1A_1L_1L_2E_2L_3$
- $E_1A_1L_1L_3E_2L_2$
- $E_2A_1L_1L_2E_1L_3$
- $E_2A_1L_1L_3E_1L_2$
- $E_1A_1L_2L_1E_2L_3$
- $E_1A_1L_2L_3E_2L_1$
- $E_2A_1L_2L_1E_1L_3$
- $E_2A_1L_2L_3E_1L_1$
- $E_1A_1L_3L_1E_2L_2$
- $E_1A_1L_3L_2E_2L_1$
- $E_2A_1L_3L_1E_1L_2$
- $E_2A_1L_3L_2E_1L_1$

Each rearrangement of $ALLELE$ has:
- $1! = 1$ way to subscript the A’s;
- $2! = 2$ ways to subscript the E’s; and
- $3! = 6$ ways to subscript the L’s,

giving $1! \cdot 2! \cdot 3! = 1 \cdot 2 \cdot 6 = 12$ ways to assign subscripts.

Since each permutation of $ALLELE$ is represented 12 different ways in permutations of $A_1L_1L_2E_1L_3E_2$, the number of permutations of $ALLELE$ is

$$\frac{6!}{1!2!3!} = \frac{720}{12} = 60.$$
Multinomial coefficients

For a word of length $n$ with $k_1$ of one letter, $k_2$ of a second letter, etc., the number of permutations is given by the multinomial coefficient:

$$\binom{n}{k_1, k_2, \ldots, k_r} = \frac{n!}{k_1! k_2! \cdots k_r!}$$

where $n, k_1, k_2, \ldots, k_r$ are integers $\geq 0$ and $n = k_1 + \cdots + k_r$. 
Mass Spectrometry (Mass Spec)
Peptide [242.3]D[I,L]SED[Q,K]D[I,L][Q,K]AEVN; Figure courtesy Nuno Bandeira
Peptide ABCDEF is ionized into fragments

\[ A / BCDEF, \ AB / CDEF, \text{ etc.} \]

giving a spectrum with intermingled peaks:

- **b-ions**: \( b_1 = \text{mass}(A), \ b_2 = \text{mass}(AB), \ldots, \ b_6 = \text{mass}(ABCDEF) \)
  successively separated by \( \text{mass}(B), \text{mass}(C), \ldots, \text{mass}(F) \)

- **y-ions**: \( y_1 = \text{mass}(F), \ y_2 = \text{mass}(EF), \ldots, \ y_6 = \text{mass}(ABCDEF) \)
  successively separated by \( \text{mass}(E), \text{mass}(D), \ldots, \text{mass}(A) \)

- Plus more peaks (multiple fragments, ± smaller chemicals, etc.).
## List of the 20 amino acids

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Code</th>
<th>Mass (Daltons)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alanine</td>
<td>A</td>
<td>71.037113787</td>
</tr>
<tr>
<td>Arginine</td>
<td>R</td>
<td>156.101111026</td>
</tr>
<tr>
<td>Aspartic acid</td>
<td>D</td>
<td>115.026943031</td>
</tr>
<tr>
<td>Asparagine</td>
<td>N</td>
<td>114.042927446</td>
</tr>
<tr>
<td>Cysteine</td>
<td>C</td>
<td>160.030648200</td>
</tr>
<tr>
<td>Glutamic acid</td>
<td>E</td>
<td>129.042593095</td>
</tr>
<tr>
<td>Glutamine</td>
<td>Q</td>
<td>128.058577510</td>
</tr>
<tr>
<td>Glycine</td>
<td>G</td>
<td>57.021463723</td>
</tr>
<tr>
<td>Histidine</td>
<td>H</td>
<td>137.058911861</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>I</td>
<td>113.084063979</td>
</tr>
<tr>
<td>Leucine</td>
<td>L</td>
<td>113.084063979</td>
</tr>
<tr>
<td>Lysine</td>
<td>K</td>
<td>128.094963016</td>
</tr>
<tr>
<td>Methionine</td>
<td>M</td>
<td>131.040484605</td>
</tr>
<tr>
<td>Asparagine</td>
<td>N</td>
<td>114.042927446</td>
</tr>
<tr>
<td>Proline</td>
<td>P</td>
<td>97.052763851</td>
</tr>
<tr>
<td>Serine</td>
<td>S</td>
<td>87.032028409</td>
</tr>
<tr>
<td>Threonine</td>
<td>T</td>
<td>101.047678473</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>W</td>
<td>186.079312952</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>Y</td>
<td>163.063328537</td>
</tr>
<tr>
<td>Valine</td>
<td>V</td>
<td>99.068413915</td>
</tr>
</tbody>
</table>

Note mass(I)=mass(L), mass(N)=mass(GG) and mass(GA)=mass(Q)≈mass(K).

A fragment of mass ≈ 242.3 could be
- mass(NE) = 243.09
- mass(LQ) = 241.14
- mass(KI) = 241.18
- mass(GGE) = 243.09
- mass(GAL) = 241.14

Or any permutations of those since they have the same mass: NE, EN, LQ, QL, KI, IK, GGE, GEG, EGG, GAL, GLA, ALG, etc.
Consider a biased 6-sided die:
- $q_i$ is the probability of rolling $i$, for $i = 1, 2, \ldots, 6$.
- Each $q_i$ is between 0 and 1, and $q_1 + \cdots + q_6 = 1$.
- 6 sides is an example; it could be any # sides.

The probability of a sequence of independent rolls is
\[
P(1131326) = q_1 q_1 q_3 q_1 q_3 q_2 q_6 = q_1^3 q_2 q_3^2 q_6 = \prod_{i=1}^{6} q_i^{\# i's}
\]

Roll the die $n$ times ($n = 0, 1, 2, 3, \ldots$).
Let $X_1$ be the number of 1’s, $X_2$ be the number of 2’s, etc.
\[
p_{X_1,X_2,\ldots,X_6}(k_1,k_2,\ldots,k_6) = P(X_1 = k_1, X_2 = k_2, \ldots, X_6 = k_6)
\]
\[
\begin{cases}
\binom{n}{k_1,k_2,\ldots,k_6} q_1^{k_1} q_2^{k_2} \cdots q_6^{k_6} & \text{if } k_1,\ldots,k_6 \text{ are integers } \geq 0 \text{ adding up to } n; \\
0 & \text{otherwise.}
\end{cases}
\]
Binomial coefficients

Suppose you flip a coin $n = 5$ times. How many sequences of flips are there with $k = 3$ heads? Ten:

\[ HHHTT \quad HHTHT \quad HHTTH \quad HTHHT \quad HTHTH \]

\[ HTTHH \quad THHHT \quad THHHT \quad THTHH \quad TTHHH \]

Definition (Binomial coefficient)

- “$n$ choose $k$” = \( \binom{n}{k} = \frac{n!}{k!(n-k)!} \)
  provided $n, k$ are integers and $0 \leq k \leq n$.
- \( \binom{n}{0} = 1 \)
- Some people use \( nC_k \) instead of \( \binom{n}{k} \).
- Binomial coefficient \( \binom{n}{k} \) = multinomial coefficient \( \binom{n}{k, n-k} \).

Top of slide: \( \binom{5}{3} = \frac{5!}{3!(5-3)!} = \frac{120}{6(2)} = 10. \)
Binomial distribution

- A biased coin has probability $p$ of heads, $q = 1 - p$ of tails.
- Flip the coin $n$ times ($n = 0, 1, 2, 3, \ldots$).
- $P(HHTHTTH) = ppqpqqp = p^4q^3 = p^\# \text{heads} q^\# \text{tails}$
- Let $X$ be the number of heads in the $n$ flips. The probability density function (pdf) of $X$ is

$$p_X(k) = P(X = k) = \begin{cases} \binom{n}{k}p^k q^{n-k} & \text{if } k = 0, 1, \ldots, n; \\ 0 & \text{otherwise}. \end{cases}$$

It's $\geq 0$ and the total is $\sum_{k=0}^{n} \binom{n}{k}p^k q^{n-k} = (p + q)^n = 1^n = 1$.

**Interpretation:** Repeat this experiment (flipping a coin $n$ times and counting the heads) a huge number of times. The fraction of experiments with $X = k$ will usually be approximately $p_X(k)$. 
Binomial distribution for $n = 10, \ p = \frac{3}{4}$

$$p_X(k) = \begin{cases} \binom{10}{k} \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right)^{10-k} & \text{if } k = 0, 1, \ldots, 10; \\ 0 & \text{otherwise.} \end{cases}$$

<table>
<thead>
<tr>
<th>$k$</th>
<th>pdf</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.000000095</td>
</tr>
<tr>
<td>1</td>
<td>0.00002861</td>
</tr>
<tr>
<td>2</td>
<td>0.00038624</td>
</tr>
<tr>
<td>3</td>
<td>0.00308990</td>
</tr>
<tr>
<td>4</td>
<td>0.01622200</td>
</tr>
<tr>
<td>5</td>
<td>0.05839920</td>
</tr>
<tr>
<td>6</td>
<td>0.14599800</td>
</tr>
<tr>
<td>7</td>
<td>0.25028229</td>
</tr>
<tr>
<td>8</td>
<td>0.28156757</td>
</tr>
<tr>
<td>9</td>
<td>0.18771172</td>
</tr>
<tr>
<td>10</td>
<td>0.05631351</td>
</tr>
<tr>
<td>other</td>
<td>0</td>
</tr>
</tbody>
</table>

Discrete probability density function

![Graph showing the discrete probability density function for the binomial distribution with $n=10$, $p=\frac{3}{4}$]
Where the distribution names come from

Binomial Theorem

For integers $n \geq 0$,

$$(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

$$(x + y)^3 = \binom{3}{0} x^0 y^3 + \binom{3}{1} x^1 y^2 + \binom{3}{2} x^2 y^1 + \binom{3}{3} x^3 y^0 = y^3 + 3xy^2 + 3x^2y + x^3$$

Multinomial Theorem

For integers $n \geq 0$,

$$(x + y + z)^n = \sum_{i=0}^{n} \sum_{j=0}^{n} \sum_{k=0}^{n} \binom{n}{i,j,k} x^i y^j z^k \quad \text{subject to} \quad i+j+k=n$$

$$(x + y + z)^2 = \binom{2}{2,0,0} x^2 y^0 z^0 + \binom{2}{0,2,0} x^0 y^2 z^0 + \binom{2}{0,0,2} x^0 y^0 z^2 + \binom{2}{1,1,0} x^1 y^1 z^0 + \binom{2}{1,0,1} x^1 y^0 z^1 + \binom{2}{0,1,1} x^0 y^1 z^1$$

$$= x^2 + y^2 + z^2 + 2xy + 2xz + 2yz$$

$$(x_1 + \cdots + x_m)^n \text{ works similarly with } m \text{ iterated sums.}$$
Genetics example

Consider a cross of two pea plants.

We will study the genes for plant height (alleles $T$=tall, $t$=short) and pea shape ($R$=round, $r$=wrinkled).

$T, R$ are dominant and $t, r$ are recessive.

The $T$ and $R$ loci are on different chromosomes so these recombine independently.

Consider a $TtRR \times TtRr$ cross of pea plants:

<table>
<thead>
<tr>
<th>Punnett Square</th>
<th>Genotype</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$TR (1/4)$</td>
<td>$TRR (1/8)$</td>
</tr>
<tr>
<td>$tR (1/4)$</td>
<td>$TRR (1/8)$</td>
<td>$TtRR (1/8)$</td>
</tr>
<tr>
<td>$tR (1/4)$</td>
<td>$TtRR (1/8)$</td>
<td>$ttRR (1/8)$</td>
</tr>
<tr>
<td>$Tr (1/4)$</td>
<td>$TTRr (1/8)$</td>
<td>$TtRr (1/8)$</td>
</tr>
<tr>
<td>$TR (1/2)$</td>
<td>$TTR (1/2)$</td>
<td></td>
</tr>
<tr>
<td>$tR (1/2)$</td>
<td>$ttR (1/2)$</td>
<td></td>
</tr>
</tbody>
</table>
If there are 27 offspring, what is the probability that 9 offspring have genotype TTRR, 2 have genotype TtRR, 3 have genotype TTRr, 5 have genotype TtRr, 7 have genotype ttRR, and 1 has genotype ttRr?

Use the multinomial distribution:

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Probability</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTRR</td>
<td>1/8</td>
<td>9</td>
</tr>
<tr>
<td>TtRR</td>
<td>1/4</td>
<td>2</td>
</tr>
<tr>
<td>TTRr</td>
<td>1/8</td>
<td>3</td>
</tr>
<tr>
<td>TtRr</td>
<td>1/4</td>
<td>5</td>
</tr>
<tr>
<td>ttRR</td>
<td>1/8</td>
<td>7</td>
</tr>
<tr>
<td>ttRr</td>
<td>1/8</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1</strong></td>
<td><strong>27</strong></td>
</tr>
</tbody>
</table>

\[
P = \frac{27!}{9!2!3!5!7!1!} \left( \frac{1}{8} \right)^9 \left( \frac{1}{4} \right)^2 \left( \frac{1}{8} \right)^3 \left( \frac{1}{4} \right)^5 \left( \frac{1}{8} \right)^7 \left( \frac{1}{8} \right)^1 \approx 2.19 \cdot 10^{-7}
\]
If there are 25 offspring, what is the probability that 9 offspring have genotype TTRR, 2 have genotype TtRR, 3 have genotype TTRr, 5 have genotype TtRr, 7 have genotype ttRR, and 1 has genotype ttRr?

\[ P = 0 \text{ because the numbers } 9, 2, 3, 5, 7, 1 \text{ do not add up to } 25. \]
# Genetics example

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Probability</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTRR</td>
<td>1/8</td>
<td>tall and round</td>
</tr>
<tr>
<td>TtRR</td>
<td>1/4</td>
<td>tall and round</td>
</tr>
<tr>
<td>TTRr</td>
<td>1/8</td>
<td>tall and round</td>
</tr>
<tr>
<td>TtRr</td>
<td>1/4</td>
<td>tall and round</td>
</tr>
<tr>
<td>ttRR</td>
<td>1/8</td>
<td>short and round</td>
</tr>
<tr>
<td>ttRr</td>
<td>1/8</td>
<td>short and round</td>
</tr>
</tbody>
</table>

For phenotypes,

\[
P(\text{tall and round}) = \frac{1}{8} + \frac{1}{4} + \frac{1}{8} + \frac{1}{4} = \frac{3}{4}
\]

\[
P(\text{short and round}) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}
\]

\[
P(\text{tall and wrinkled}) = P(\text{short and wrinkled}) = 0
\]

If there are 10 offspring, the number of tall offspring has a binomial distribution with \( n = 10, p = \frac{3}{4} \).

**Later:** We will see other bioinformatics applications that use the binomial distribution, including genome assembly and Haldane’s model of recombination.
Expected value of a random variable

(Technical name for long term average)

- Consider a biased coin with probability $p = 3/4$ for heads.

- Flip it 10 times and record the number of heads, $x_1$. Flip it another 10 times, get $x_2$ heads. Repeat to get $x_1, \ldots, x_{1000}$.

- **Estimate the average of $x_1, \ldots, x_{1000}$:** $10(3/4) = 7.5$

- **An estimate based on the pdf:**
  About $1000 p_X(k)$ of the $x_i$’s equal $k$ for each $k = 0, \ldots, 10$, so

  
  \[
  \text{average of } x_i \text{’s} \approx \frac{\sum_{i=1}^{1000} x_i}{1000} \approx \frac{\sum_{k=0}^{10} k \cdot 1000 p_X(k)}{1000} = \sum_{k=0}^{10} k \cdot p_X(k)
  \]
The expected value of a discrete random variable $X$ is

$$E(X) = \sum_x x \cdot p_X(x)$$

$E(X)$ is often called the mean value of $X$ and denoted $\mu$ (or $\mu_X$ if there are other random variables).

It turns out $E(X) = np$ for the binomial distribution.

On the previous slide, although $E(X) = np = 10(3/4) = 7.5$, this is not a possible value for $X$.

Expected value does not mean we anticipate observing that value.

It means the long term average of many independent measurements of $X$ will be approximately $E(X)$. 


Proof that $\mu = np$ for binomial distribution.

$$E(X) = \sum_k k \cdot p_X(k)$$

$$= \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k q^{n-k}$$

Calculus Trick:

$$(p + q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$$

Differentiate:

$$\frac{\partial}{\partial p} (p + q)^n = \sum_{k=0}^{n} k \binom{n}{k} p^{k-1} q^{n-k}$$

Times $p$:

$$p \frac{\partial}{\partial p} (p + q)^n = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} = E(X)$$

Evaluate left side:

$$p \frac{\partial}{\partial p} (p + q)^n = p \cdot n(p + q)^{n-1}$$

$$= p \cdot n \cdot 1^{n-1} = np \quad \text{since } p + q = 1.$$

So $E(X) = np$. □
Expected values of functions

Let $X = \text{roll of a biased 6-sided die}$ and $Z = (X - 3)^2$.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$p_X(x)$</th>
<th>$z = (x - 3)^2$</th>
<th>$p_Z(z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$q_1$</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$q_2$</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>$q_3$</td>
<td>0</td>
<td>$p_Z(0) = q_3$</td>
</tr>
<tr>
<td>4</td>
<td>$q_4$</td>
<td>1</td>
<td>$p_Z(1) = q_2 + q_4$</td>
</tr>
<tr>
<td>5</td>
<td>$q_5$</td>
<td>4</td>
<td>$p_Z(4) = q_1 + q_5$</td>
</tr>
<tr>
<td>6</td>
<td>$q_6$</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

pdf of $X$: Each $q_i \geq 0$ and $q_1 + \cdots + q_6 = 1$.

pdf of $Z$: Each probability is also $\geq 0$, and the total sum is also 1.

$E(Z)$, in terms of values of $Z$ and the pdf of $Z$, is

$$E(Z) = \sum_z z \cdot p_Z(z) = 0(q_3) + 1(q_2 + q_4) + 4(q_1 + q_5) + 9(q_6)$$

Regroup it in terms of $X$:

$$= 4q_1 + 1q_2 + 0q_3 + 1q_4 + 4q_5 + 9q_6 = \sum_{x=1}^{6} (x - 3)^2 q_x$$
Expected values of functions

- Define

\[ E(g(X)) = \sum_x g(x) \cdot p_X(x) \]

In general, if \( Z = g(X) \) then \( E(Z) = E(g(X)) \). The preceding slide demonstrates this for \( Z = (X - 3)^2 \).

- For functions of two variables, define

\[ E(g(X, Y)) = \sum_x \sum_y g(x, y)p_{X,Y}(x, y) \]

and for more variables, do more iterated sums.
Expected values — properties

- $E(aX + b) = aE(X) + b$ where $a, b$ are constants:

  \[
  E(aX + b) = \sum_x p_X(x) (ax + b) = a \sum_x xp_X(x) + b \sum_x p_X(x)
  \]

  \[
  = aE(X) + b \cdot 1 = aE(X) + b
  \]

- $E(a \ g(X)) = aE(g(X))$

  $E(a) = a$

  $E(g(X, Y) + h(X, Y)) = E(g(X, Y)) + E(h(X, Y))$

- **If $X$ and $Y$ are independent** then $E(XY) = E(X)E(Y)$:

  \[
  E(XY) = \sum_x \sum_y p_{X,Y}(x, y) \cdot xy
  \]

  \[
  = \sum_x \sum_y p_X(x)p_Y(y) \cdot xy \quad \text{if } X, Y \text{ independent!}
  \]

  \[
  = \left( \sum_x p_X(x)x \right) \left( \sum_y p_Y(y)y \right) = E(X)E(Y)
  \]
Expected value of a product — dependent variables

**Example (Dependent)**

- Let $U$ be the roll of a fair 6-sided die.
- Let $V$ be the value of the exact same roll of the die ($U = V$).
- $E(U) = E(V) = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = \frac{7}{2}$ and $E(U)E(V) = \frac{49}{4}$.
- $E(UV) = \frac{1 \cdot 1 + 2 \cdot 2 + 3 \cdot 3 + 4 \cdot 4 + 5 \cdot 5 + 6 \cdot 6}{6} = \frac{91}{6}$

**Example (Independent)**

- Now let $U, V$ be the values of two independent rolls of a fair 6-sided die.
- $E(UV) = \sum_{x=1}^{6} \sum_{y=1}^{6} \frac{x \cdot y}{36} = \frac{441}{36} = \frac{49}{4}$

and $E(U)E(V) = \left(\frac{7}{2}\right)\left(\frac{7}{2}\right) = \frac{49}{4}$
These distributions both have mean $= 0$, but the right one is more spread out.

The variance of $X$ measures the square of the spread from the mean:

$$\sigma^2 = \text{Var}(X) = E((X - \mu)^2)$$

The standard deviation of $X$ is $\sigma = \text{SD}(X) = \sqrt{\text{Var}(X)}$ and measures how wide the curve is.
Variance — properties

- **Var**($aX + b$) = $a^2$ **Var**($X$)  
  **SD**($aX + b$) = $|a|$ **SD**($X$)

Adding $b$ shifts the curve without changing the width, so $b$ disappears on the right side of the variance formula.

- Multiplying by $a$ dilates the width a factor of $a$, so variance goes up a factor $a^2$.
- For $Y = aX + b$, we have $\sigma_Y = |a| \sigma_X$ and $\mu_Y = a \mu_X + b$.
- **Example:** Convert measurements in °C to °F:
  
  $F = \left(\frac{9}{5}\right)C + 32$  
  $\mu_F = \left(\frac{9}{5}\right)\mu_C + 32$  
  $\sigma_F = \left(\frac{9}{5}\right)\sigma_C$
Useful alternative formula for variance

\[ \sigma^2 = \text{Var}(X) = E(X^2) - \mu^2 = E(X^2) - (E(X))^2 \]

Proof.

\[
\text{Var}(X) = E((X - \mu)^2) = E(X^2 - 2\mu X + \mu^2)
= E(X^2) - 2\mu E(X) + \mu^2
= E(X^2) - 2\mu \cdot \mu + \mu^2 = E(X^2) - \mu^2
\]

Proof of \( \text{Var}(aX + b) = a^2 \text{Var}(X) \).

\[
E((aX + b)^2) = E(a^2X^2 + 2ab X + b^2) = a^2E(X^2) + 2ab E(X) + b^2
\]

\[
(E(aX + b))^2 = (aE(X) + b)^2 = a^2(E(X))^2 + 2ab E(X) + b^2
\]

\[
\text{Var}(aX + b) = \text{difference} = a^2 \left( E(X^2) - (E(X))^2 \right)
= a^2 \text{Var}(X)
\]
We will show that if $X, Y$ are independent, then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Example (Dependent)

First consider this dependent example:
Let $X$ be any non-constant random variable and $Y = -X$.

$$\text{Var}(X + Y) = \text{Var}(0) = 0$$

$$\text{Var}(X) + \text{Var}(Y) = \text{Var}(X) + \text{Var}(-X)$$

$$= \text{Var}(X) + (-1)^2 \text{Var}(X) = 2 \text{Var}(X)$$

but usually $\text{Var}(X) \neq 0$ (the only exception would be if $X$ is a constant).
Variance of a sum — independent variables

**Theorem**

If $X, Y$ are independent, then $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

**Proof.**

\[
E((X + Y)^2) = E(X^2 + 2XY + Y^2) = E(X^2) + 2E(XY) + E(Y^2)
\]
\[
(E(X + Y))^2 = (E(X) + E(Y))^2 = (E(X))^2 + 2E(X)E(Y) + (E(Y))^2
\]

\[
\text{Var}(X + Y) = E((X + Y)^2) - (E(X + Y))^2
\]
\[
= (E(X^2) - (E(X))^2) + 2(E(XY) - E(X)E(Y)) + (E(Y^2) - (E(Y))^2)
\]
\[
= \text{Var}(X) + 2(E(XY) - E(X)E(Y)) + \text{Var}(Y)
\]

If $X, Y$ are independent, $E(XY) = E(X)E(Y)$, so the middle term is 0. □

**Generalization**

If $X, Y, Z, \ldots$ are pairwise independent:

\[
\text{Var}(X + Y + Z + \cdots) = \text{Var}(X) + \text{Var}(Y) + \text{Var}(Z) + \cdots
\]

\[
\text{Var}(aX + bY + cZ + \cdots) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + c^2 \text{Var}(Z) + \cdots
\]
Variance of a sum — dependent variables

**Covariance**

- For dependent variables, the cross-terms remain:

\[
\text{Var}(X + Y) = \text{Var}(X) + 2(E(XY) - E(X)E(Y)) + \text{Var}(Y)
\]

- Define \( \text{Cov}(X, Y) = E(XY) - E(X)E(Y) \). Then

\[
\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)
\]

**Two formulas for covariance:**

\[
\text{Cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y)) = E(XY) - E(X)E(Y)
\]

\[
E((X - \mu_X)(Y - \mu_Y)) = E(XY) - \mu_X E(Y) - E(X)\mu_Y + \mu_X \mu_Y \\
= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \\
= E(XY) - E(X)E(Y)
\]
Covariance properties

\[
\begin{align*}
\text{Var}(X) &= E((X - \mu_X)^2) = E(X^2) - (E(X))^2 \\
\text{Cov}(X, Y) &= E((X - \mu_X)(Y - \mu_Y)) = E(XY) - E(X)E(Y)
\end{align*}
\]

Additional properties

- \( \text{Cov}(X, X) = \text{Var}(X) \)
- \( \text{Cov}(X, Y) = \text{Cov}(Y, X) \)
- If \( X, Y \) are independent then \( \text{Cov}(X, Y) = 0 \).
- \( \text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y) \) (\( a, b, c, d \) are constants)
- \( \text{Cov}(X + Z, Y) = \text{Cov}(X, Y) + \text{Cov}(Z, Y) \) and \( \text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z) \)
- \( \text{Var}(X_1 + X_2 + \cdots + X_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j) \)
Mean and variance of the Binomial Distribution

- **A Bernoulli trial** is a single coin flip,
  \[ P(\text{heads}) = p, \quad P(\text{tails}) = 1 - p = q. \]

- Do \( n \) coin flips (\( n \) Bernoulli trials). Set
  \[ X_i = \begin{cases} 
  1 & \text{if flip } i \text{ is heads;} \\
  0 & \text{if flip } i \text{ is tails.} 
  \end{cases} \]

  The total number of heads in all flips is \( X = X_1 + X_2 + \cdots + X_n \).

- Flips **HTTHT**: \( X = 1 + 0 + 0 + 1 + 0 = 2 \).

- \( X_1, \ldots, X_n \) are independent and have the same pdfs, so they are i.i.d. (independent identically distributed) random variables.

\[
\begin{align*}
E(X_1) &= 0(1 - p) + 1p = p \\
E(X_1^2) &= 0^2(1 - p) + 1^2p = p \\
\text{Var}(X_1) &= E(X_1^2) - (E(X_1))^2 = p - p^2 = p(1 - p)
\end{align*}
\]

- \( E(X_i) = p \) and \( \text{Var}(X_i) = p(1 - p) \) for all \( i = 1, \ldots, n \) because they are identically distributed.
Mean and variance of the Binomial Distribution

- The total number of heads in all flips is $X = X_1 + X_2 + \cdots + X_n$.
- $E(X_i) = p$ and $\text{Var}(X_i) = p(1 - p)$ for all $i = 1, \ldots, n$.

Mean:

$$\mu_X = E(X) = E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) = p + \cdots + p = np$$

Variance:

$$\sigma_X^2 = \text{Var}(X) = \text{Var}(X_1 + \cdots + X_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n) \quad \text{by independence}$$

$$= p(1 - p) + \cdots + p(1 - p) = np(1 - p) = npq$$

Standard deviation:

$$\sigma_X = \sqrt{np(1 - p)} = \sqrt{npq}$$
For the binomial distribution,

**Mean:** \( \mu = np \)

**Variance:** \( \sigma^2 = np(1 - p) \)

**Standard deviation:** \( \sigma = \sqrt{np(1 - p)} \)

At \( n = 100 \) and \( p = \frac{3}{4} \):

\[ \mu = 100 \left( \frac{3}{4} \right) = 75 \]

\[ \sigma = \sqrt{100 \left( \frac{3}{4} \right) \left( \frac{1}{4} \right)} \approx 4.33 \]

Approximately 68% of the probability is for \( X \) between \( \mu \pm \sigma \). Approximately 95% of the probability is for \( X \) between \( \mu \pm 2\sigma \). More on that later when we do the normal distribution.
Consider a biased coin with probability $p$ of heads.
Flip it repeatedly (potentially $\infty$ times).
Let $X$ be the number of flips until the first head.

**Example:** $TTHTHTHTHHT$ has $X = 4$.

The pdf is

$$p_X(k) = \begin{cases} (1 - p)^{k-1}p & \text{for } k = 1, 2, 3, \ldots; \\ 0 & \text{otherwise} \end{cases}$$

- Mean: $\mu = 1/p$
- Variance: $\sigma^2 = (1 - p)/p^2$
- Standard deviation: $\sigma = \sqrt{1 - p}/p$
Consider a biased coin with probability $p$ of heads.

Flip it repeatedly (potentially $\infty$ times).

Let $X$ be the number of flips until the $r$th head ($r = 1, 2, 3, \ldots$ is a fixed parameter).

For $r = 3$, $TTTHTHHTTH$ has $X = 7$.

$X = k$ when

- **first $k - 1$ flips:** $r - 1$ heads and $k - r$ tails in any order;
- **$k$th flip:** heads

so the pdf is

$$p_X(k) = \binom{k - 1}{r - 1} p^{r-1} (1-p)^{k-r} \cdot p = \binom{k - 1}{r - 1} p^r (1-p)^{k-r}$$

provided $k = r, r + 1, r + 2, \ldots$;

$$p_X(k) = 0 \quad \text{otherwise.}$$
Consider the sequence of flips \( TTTHTHHTTH \).

Break it up at each heads:

\[
\begin{align*}
TTTH / TH / H / TTH \\
X_1 = 4 & \quad X_2 = 2 & \quad X_3 = 1 & \quad X_4 = 3
\end{align*}
\]

\( X_1 \) is the number of flips until the first heads;
\( X_2 \) is the number of additional flips until the 2nd heads;
\( X_3 \) is the number of additional flips until the 3rd heads; \ldots

The \( X_i \)'s are i.i.d. geometric random variables with parameter \( p \),
and \( X = X_1 + \cdots + X_r \).

Mean: \( E(X) = E(X_1) + \cdots + E(X_r) = 1/p + \cdots + 1/p = r/p \)

Variance: \( \sigma^2 = \frac{1-p}{p^2} + \cdots + \frac{1-p}{p^2} = \frac{r(1-p)}{p^2} \)

Standard deviation: \( \sigma = \sqrt{r(1-p)/p} \)
Geometric Distribution – example

- About 10% of the population is left-handed.
- Look at the handedness of babies in birth order in a hospital.
- **Number of births until first left-handed baby:**
  Geometric distribution with $p = .1$:

  $$p_X(x) = .9^{x-1} \cdot .1 \quad \text{for } x = 1, 2, 3, \ldots$$

- **Mean:** $1/p = 10$.
- **Standard deviation:** $\sigma = \frac{\sqrt{1-p}}{p} = \frac{\sqrt{.9}}{.1} \approx 9.487$, which is HUGE!
Negative Binomial Distribution – example

Number of births until 8th left-handed baby:
Negative binomial, \( r = 8, p = .1 \).

\[
p_{X}(x) = \binom{x-1}{8-1} (.1)^8 (.9)^{x-8} \quad \text{for } x = 8, 9, 10, \ldots
\]

\[\mu \pm \sigma\]

Mean: \( r/p = 8/.1 = 80 \).

Standard deviation: \( \sqrt{r(1-p)/p} = \sqrt{8(.9)/.1} \approx 26.833 \).

Probability the 50th baby is the 8th left-handed one:

\[
p_{X}(50) = \binom{50-1}{8-1} (.1)^8 (.9)^{50-8} = \binom{49}{7} (.1)^8 (.9)^{42} \approx 0.0103
\]
Where the distribution names come from

**Geometric series**

- For real $a, x$ with $|x| < 1$,
  \[
  \frac{a}{1 - x} = \sum_{i=0}^{\infty} a x^i
  \]
  \[
  = a + ax + ax^2 + \cdots
  \]
  The total probability for the geometric distribution is
  \[
  \sum_{k=1}^{\infty} (1 - p)^{k-1} p
  \]
  \[
  = \frac{p}{1 - (1 - p)}
  \]
  \[
  = \frac{p}{p} = 1
  \]

**Negative binomial series**

- For integer $r > 0$ and real $x$ with $|x| < 1$,
  \[
  \frac{1}{(1 - x)^r} = \sum_{k=r}^{\infty} \binom{k-1}{r-1} x^{k-r}
  \]
  The total probability for the negative binomial distribution is
  \[
  \sum_{k=r}^{\infty} \binom{k-1}{r-1} p^r (1 - p)^{k-r}
  \]
  \[
  = p^r \sum_{k=r}^{\infty} \binom{k-1}{r-1} (1 - p)^{k-r}
  \]
  \[
  = p^r \cdot \frac{1}{(1 - (1 - p))^r} = 1
  \]
Unfortunately, there are 4 versions of the definitions of these distributions. Our book uses versions 1 and 2 below, and you may see the others elsewhere. Authors should be careful to state which definition they’re using.

- **Version 1**: the definitions we already did (call the variable $X$).
- **Version 2 (geometric)**: Let $Y$ be the number of tails before the first heads, so $TTTHTTHHT$ has $Y = 3$.
  
  The probability mass function (pdf) is:
  
  $$p_Y(k) = \begin{cases} (1 - p)^k p & \text{for } k = 0, 1, 2, \ldots; \\ 0 & \text{otherwise} \end{cases}$$

  Since $Y = X - 1$, we have $E(Y) = \frac{1}{p} - 1$, $\text{Var}(Y) = \frac{1-p}{p^2}$.

- **Version 2 (negative binomial)**: Let $Y$ be the number of tails before the $r$th heads, so $Y = X - r$.
  
  The pdf is:
  
  $$p_Y(k) = \begin{cases} \binom{k+r-1}{r-1} p^r (1 - p)^k & \text{for } k = 0, 1, 2, \ldots; \\ 0 & \text{otherwise} \end{cases}$$

- **Versions 3 and 4**: switch the roles of heads and tails in the first two versions (so $p$ and $1 - p$ are switched).