Chapters 1–2

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

Prof. Tesler

Math 283
Fall 2019

Sample spaces and events

- Flip a coin 3 times. The possible *outcomes* are

$$HHH \quad HHT \quad HTH \quad HTT \quad THH \quad THT \quad TTH \quad TTT$$

- The *sample space* is the set of all possible outcomes:

$$S = \{HHH,\ HHT,\ HTH,\ HTT,\ THH,\ THT,\ TTH,\ TTT\}$$

- An *event* is any subset of S.
  The event that there are exactly two heads is

$$A = \{HHT, HTH, THH\}$$

- The probability of heads is p and of tails is $q = 1 - p$. The flips are independent, which gives these *probabilities for each outcome*:

$$P(HHH) = p^3 \quad P(HHT) = P(HTH) = P(THH) = p^2q$$
$$P(TTT) = q^3 \quad P(HTT) = P(THT) = P(TTH) = pq^2$$

- These are each between 0 and 1, and they add up to 1:

$$p^3 + 3p^2q + 3pq^2 + q^3 = (p+q)^3 = 1^3 = 1$$

Sample spaces and events

- Flip a coin 3 times. The possible *outcomes* are

$$HHH \quad HHT \quad HTH \quad HTT \quad THH \quad THT \quad TTH \quad TTT$$

- The *sample space* is the set of all possible outcomes:

$$S = \{HHH,\ HHT,\ HTH,\ HTT,\ THH,\ THT,\ TTH,\ TTT\}$$

- An *event* is any subset of S.
  The event that there are exactly two heads is

$$A = \{HHT, HTH, THH\}$$

- The probability of heads is p and of tails is $q = 1 - p$. The flips are independent, which gives these *probabilities for each outcome*:

$$P(HHH) = p^3 \quad P(HHT) = P(HTH) = P(THH) = p^2 q$$
$$P(TTT) = q^3 \quad P(HTT) = P(THT) = P(TTH) = pq^2$$

- The *probability of an event* is the sum of probabilities of its outcomes:

$$P(A) = P(HHT) + P(HTH) + P(THH) = 3p^2 q$$

Random variables

- A *random variable* X is a function assigning a real number to each outcome.
- Let X be the number of heads:

$$X(HHH) = 3 \quad X(HHT) = X(HTH) = X(THH) = 2$$
$$X(TTT) = 0 \quad X(HTT) = X(THT) = X(TTH) = 1$$

- 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:

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

$$p_X(k) = 0 \text{ otherwise:} \qquad p_X(2.5) = 0 \qquad p_X(-1) = 0$$

- Use capital letters (X) for random variables and lowercase (k) to stand for numeric values.

Joint probability density

- 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: $X = 3$, $Y = 1$	THH: $X = 2$, $Y = 2$
HHT: $X = 2$, $Y = 1$	THT: $X = 1$, $Y = 2$
HTH: $X = 2$, $Y = 1$	TTH: $X = 1$, $Y = 3$
HTT: $X = 1$, $Y = 1$	TTT: $X = 0$, $Y = 4$

- Reorganize as a two dimensional table:

	$X = 0$	$X = 1$	$X = 2$	$X = 3$
$Y = 1$		HTT	HHT, HTH	HHH
$Y = 2$		THT	THH	
$Y = 3$		TTH		
$Y = 4$	TTT			

Joint probability density

- The (discrete) *joint probability density function* is $p_{X,Y}(x,y) = P(X=x, Y=y)$:

$p_{X,Y}(x,y)$	$x=0$	$x=1$	$x=2$	$x=3$	Total $p_Y(y)$
$y=1$	0	pq^2	$2p^2q$	p^3	p
$y=2$	0	pq^2	p^2q	0	pq
$y=3$	0	pq^2	0	0	pq^2
$y=4$	q^3	0	0	0	q^3
Total $p_X(x)$	q^3	$3pq^2$	$3p^2q$	p^3	1

- 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:** $p_{X,Y}(1,-.5) = 0$

- **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

$p_{X,Y}(x,y)$	$x=0$	$x=1$	$x=2$	$x=3$	Total $p_Y(y)$
$y=1$	0	pq^2	$2p^2q$	p^3	p
$y=2$	0	pq^2	p^2q	0	pq
$y=3$	0	pq^2	0	0	pq^2
$y=4$	q^3	0	0	0	q^3
Total $p_X(x)$	q^3	$3pq^2$	$3p^2q$	p^3	1

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

- These are called *conditional probabilities*.

Conditional probability formula

- You know that event B holds. What's the probability of event A?

Conditional Probability Formula

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 $B = \{HHT, HTH, THH\}$.

$$\begin{aligned} P(Y = 1 \mid X = 2) &= \frac{P(X = 2 \text{ and } Y = 1)}{P(X = 2)} \\ &= \frac{P(\{HHT, HTH\})}{P(\{HHT, HTH, THH\})} = \frac{2p^2q}{3p^2q} = \frac{2}{3} \end{aligned}$$

Conditional probability formula

Bayes' Theorem

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

$$P(Y = y \,|\, X = x) = \frac{P(Y = y \text{ and } X = x)}{P(X = x)} = \frac{p_{X,Y}(x, y)}{p_X(x)}$$

$$P(Y = 1 \,|\, X = 2) = \frac{p_{X,Y}(2, 1)}{p_X(2)} = \frac{2p^2q}{3p^2q} = \frac{2}{3}$$

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 $n = 1, 2, \ldots$
- $0! = 1$

Permutations with repetitions

Here are all the permutations of the letters of ALLELE:

EEALLL	*EELALL*	*EELLAL*	*EELLLA*	*EAELLL*	*EALELL*
EALLEL	*EALLLE*	*ELEALL*	*ELELAL*	*ELELLA*	*ELAELL*
ELALEL	*ELALLE*	*ELLEAL*	*ELLELA*	*ELLAEL*	*ELLALE*
ELLLEA	*ELLLAE*	*AEELLL*	*AELELL*	*AELLEL*	*AELLLE*
ALEELL	*ALELEL*	*ALELLE*	*ALLEEL*	*ALLELE*	*ALLLEE*
LEEALL	*LEELAL*	*LEELLA*	*LEAELL*	*LEALEL*	*LEALLE*
LELEAL	*LELELA*	*LELAEL*	*LELALE*	*LELLEA*	*LELLAE*
LAEELL	*LAELEL*	*LAELLE*	*LALEEL*	*LALELE*	*LALLEE*
LLEEAL	*LLEELA*	*LLEAEL*	*LLEALE*	*LLELEA*	*LLELAE*
LLAEEL	*LLAELE*	*LLALEE*	*LLLEEA*	*LLLEAE*	*LLLAEE*

Permutations with repetitions

- 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:

$$\begin{array}{cccc}
E_1A_1L_1L_2E_2L_3 & E_1A_1L_1L_3E_2L_2 & E_2A_1L_1L_2E_1L_3 & E_2A_1L_1L_3E_1L_2 \\
E_1A_1L_2L_1E_2L_3 & E_1A_1L_2L_3E_2L_1 & E_2A_1L_2L_1E_1L_3 & E_2A_1L_2L_3E_1L_1 \\
E_1A_1L_3L_1E_2L_2 & E_1A_1L_3L_2E_2L_1 & E_2A_1L_3L_1E_1L_2 & E_2A_1L_3L_2E_1L_1
\end{array}$$

- 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 $\geqslant 0$ and $n = k_1 + \cdots + k_r$.

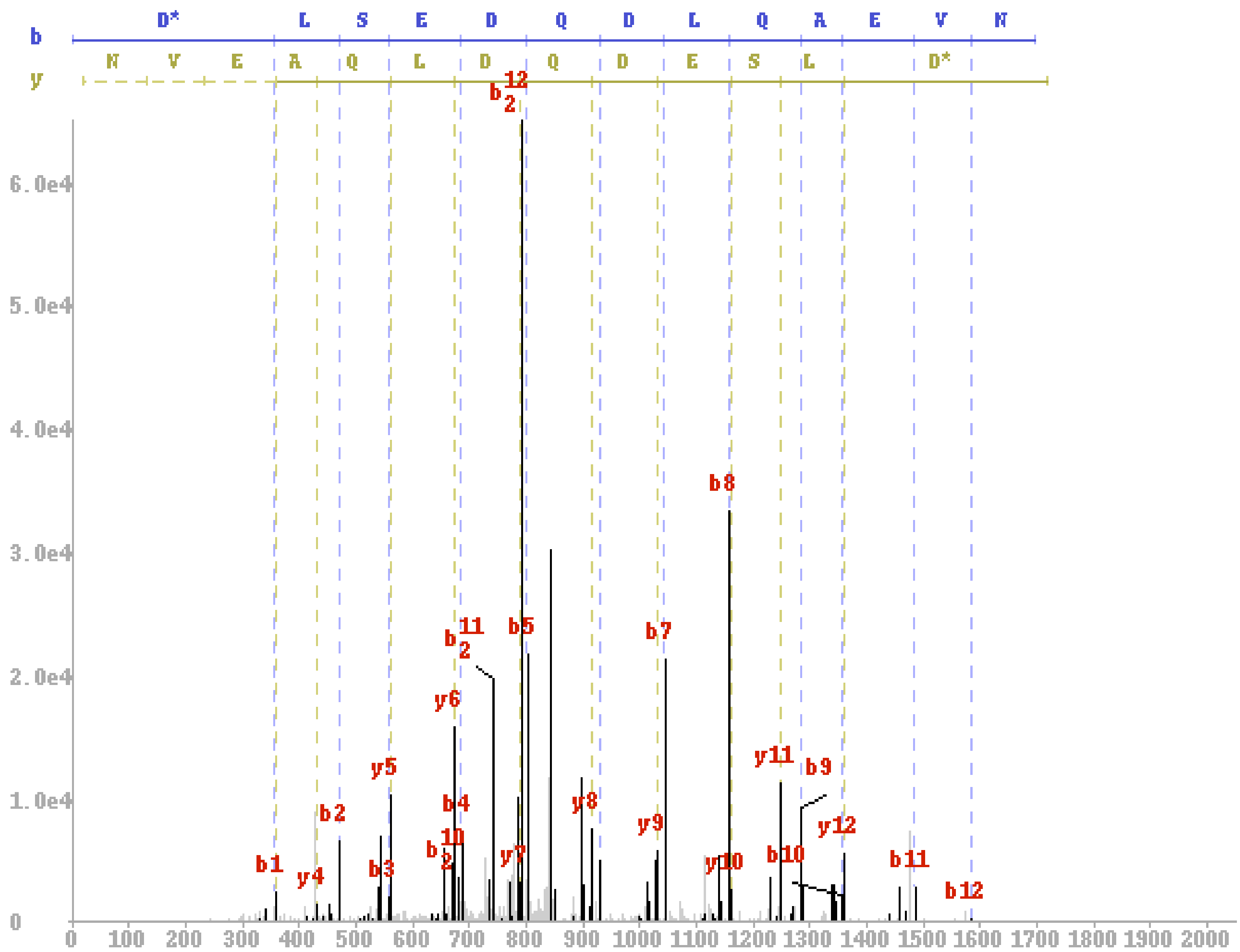
Previous slide example: ALLELE

$n = 6$ letters, with 1 A, 2 E's, 3 L's:

$$\binom{6}{1, 2, 3} = \frac{6!}{1!\,2!\,3!} = \frac{720}{12} = 60$$

Mass Spectrometry (Mass Spec)

Peptide [242.3]D[I,L]SED[Q,K]D[I,L][Q,K]AEVN; Figure courtesy Nuno Bandeira



Mass Spectrometry

Peptide ABCDEF is ionized into fragments

A / BCDEF, AB / CDEF, etc.

giving a spectrum with intermingled peaks:

- **b-ions:** $b_1 =$ mass(A), $b_2 =$ mass(AB), . . . , $b_6 =$ mass(ABCDEF) successively separated by mass(B), mass(C), . . . , mass(F)
- **y-ions:** $y_1 =$ mass(F), $y_2 =$ mass(EF), . . . , $y_6 =$ mass(ABCDEF) successively separated by mass(E), mass(D), . . . , mass(A)
- Plus more peaks (multiple fragments, $\pm$ smaller chemicals, etc.).

Mass Spectrometry — Amino Acid Composition

List of the 20 amino acids

Amino Acid	Code	Mass (Daltons)	Amino Acid	Code	Mass (Daltons)
Alanine	A	71.037113787	Leucine	L	113.084063979
Arginine	R	156.101111026	Lysine	K	128.094963016
Aspartic acid	D	115.026943031	Methionine	M	131.040484605
Asparagine	N	114.042927446	Phenylalanine	F	147.068413915
Cysteine	C	160.030648200	Proline	P	97.052763851
Glutamic acid	E	129.042593095	Serine	S	87.032028409
Glutamine	Q	128.058577510	Threonine	T	101.047678473
Glycine	G	57.021463723	Tryptophan	W	186.079312952
Histidine	H	137.058911861	Tyrosine	Y	163.063328537
Isoleucine	I	113.084063979	Valine	V	99.068413915

- Note mass(I)=mass(L), mass(N)=mass(GG) and mass(GA)=mass(Q)$\approx$mass(K).
- A fragment of mass $\approx$ 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.

Multinomial distribution

- 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

$$P(1131326) = q_1\, q_1\, q_3\, q_1\, q_3\, q_2\, q_6 = {q_1}^3\, q_2\, {q_3}^2\, q_6 = \prod_{i=1}^{6} {q_i}^{\#\ i\text{'s}}$$

- 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.

$$p_{X_1,X_2,\ldots,X_6}(k_1, k_2, \ldots, k_6) = P(X_1 = k_1, X_2 = k_2, \ldots, X_6 = k_6)$$

$$= \begin{cases} \binom{n}{k_1,k_2,\ldots,k_6} {q_1}^{k_1} {q_2}^{k_2} \ldots {q_6}^{k_6} & \text{if } k_1, \ldots, k_6 \text{ are integers} \geqslant 0 \text{ adding up to } n; \\ 0 & \text{otherwise.} \end{cases}$$

Binomial coefficients

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

$$\begin{array}{ccccc} HHHTT & HHTHT & HHTTH & HTHHT & HTHTH \\ HTTHH & THHHT & THHTH & THTHH & TTHHH \end{array}$$

Definition (Binomial coefficient)

- "n choose k" $= \binom{n}{k} = \frac{n!}{k!(n-k)!}$
  provided n, k are integers and $0 \leqslant k \leqslant n$.
- $\binom{n}{0} = 1$
- 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 ($n = 0, 1, 2, 3, \ldots$).
- $P(HHTHTTH) = ppqpqqp = p^4 q^3 = p^{\text{\# heads}} q^{\text{\# tails}}$
- Let X be the number of heads in the n flips.
  The probability density function (pdf) of X is

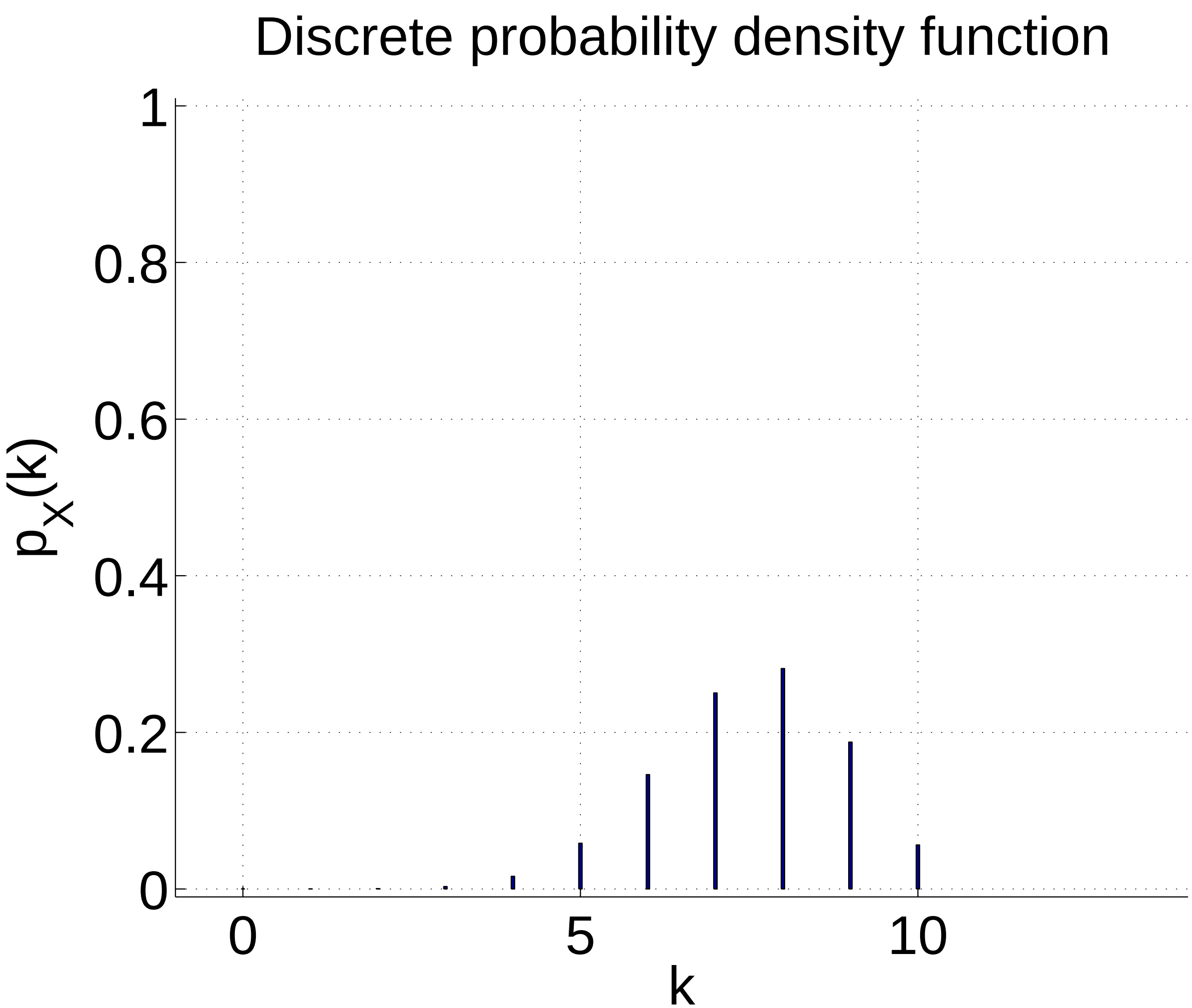
$$p_X(k) = P(X = k) = \begin{cases} \binom{n}{k} p^k q^{n-k} & \text{if } k = 0, 1, \ldots, n; \\ 0 & \text{otherwise.} \end{cases}$$

  It's $\geqslant 0$ 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 $n = 10$, $p = 3/4$

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

k	**pdf**
0	0.00000095
1	0.00002861
2	0.00038624
3	0.00308990
4	0.01622200
5	0.05839920
6	0.14599800
7	0.25028229
8	0.28156757
9	0.18771172
10	0.05631351
other	0



Where the distribution names come from

Binomial Theorem

For integers $n \geqslant 0$,

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

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

Multinomial Theorem

For integers $n \geqslant 0$,

$$(x+y+z)^n = \underbrace{\sum_{i=0}^{n} \sum_{j=0}^{n} \sum_{k=0}^{n}}_{i+j+k=n} \binom{n}{i,j,k} x^i y^j z^k$$

$$\begin{aligned}
(x+y+z)^2 &= \tbinom{2}{2,0,0}x^2y^0z^0 + \tbinom{2}{0,2,0}x^0y^2z^0 + \tbinom{2}{0,0,2}x^0y^0z^2 \\
&\quad + \tbinom{2}{1,1,0}x^1y^1z^0 + \tbinom{2}{1,0,1}x^1y^0z^1 + \tbinom{2}{0,1,1}x^0y^1z^1 \\
&= x^2 + y^2 + z^2 + 2xy + 2xz + 2yz
\end{aligned}$$

$(x_1 + \cdots + x_m)^n$ works similarly with m 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:

Punnett Square

	TR $(1/2)$	tR $(1/2)$
TR $(1/4)$	$TTRR$ $(1/8)$	$TtRR$ $(1/8)$
Tr $(1/4)$	$TTRr$ $(1/8)$	$TtRr$ $(1/8)$
tR $(1/4)$	$TtRR$ $(1/8)$	$ttRR$ $(1/8)$
tr $(1/4)$	$TtRr$ $(1/8)$	$ttRr$ $(1/8)$

Genotype	**Prob.**
$TTRR$	$1/8$
$TtRR$	$2/8 = 1/4$
$TTRr$	$1/8$
$TtRr$	$2/8 = 1/4$
$ttRR$	$1/8$
$ttRr$	$1/8$

Genetics example

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:

Genotype	**Probability**	**Frequency**
TTRR	$1/8$	9
TtRR	$1/4$	2
TTRr	$1/8$	3
TtRr	$1/4$	5
ttRR	$1/8$	7
ttRr	$1/8$	1
Total	1	27

$$P = \frac{27!}{9!\,2!\,3!\,5!\,7!\,1!}\left(\frac{1}{8}\right)^9\left(\frac{1}{4}\right)^2\left(\frac{1}{8}\right)^3\left(\frac{1}{4}\right)^5\left(\frac{1}{8}\right)^7\left(\frac{1}{8}\right)^1 \approx 2.19 \cdot 10^{-7}$$

Genetics example

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?

$P = 0$ because the numbers $9, 2, 3, 5, 7, 1$ do not add up to 25.

Genetics example

Genotype	Probability	Phenotype
TTRR	$1/8$	tall and round
TtRR	$1/4$	tall and round
TTRr	$1/8$	tall and round
TtRr	$1/4$	tall and round
ttRR	$1/8$	short and round
ttRr	$1/8$	short and round

For phenotypes,

$$P(\text{tall and round}) = 1/8 + 1/4 + 1/8 + 1/4 = 3/4$$

$$P(\text{short and round}) = 1/8 + 1/8 = 1/4$$

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

If there are 10 offspring, the number of tall offspring has a binomial distribution with $n = 10$, $p = 3/4$.

Later: We'll cover other Bioinformatics applications using 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 $x_1, \cdots, x_{1000}$.
- **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

$$\text{average of } x_i\text{'s} = \frac{\sum_{i=1}^{1000} x_i}{1000} \approx \frac{\sum_{k=0}^{10} k \cdot 1000\, p_X(k)}{1000} = \sum_{k=0}^{10} k \cdot p_X(k)$$

Expected value of a random variable

(Technical name for long term average)

- 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 μ (or μ_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 $E(X)$.

Mean of the Binomial Distribution

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

$$\begin{aligned} E(X) &= \textstyle\sum_k k \cdot p_X(k) \\ &= \textstyle\sum_{k=0}^{n} k \cdot \binom{n}{k} p^k q^{n-k} \end{aligned}$$

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

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

So $E(X) = np$. □

Expected values of functions

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

x	$p_X(x)$	$z = (x-3)^2$	$p_Z(z)$
1	q_1	4	
2	q_2	1	
3	q_3	0	$p_Z(0) = q_3$
4	q_4	1	$p_Z(1) = q_2 + q_4$
5	q_5	4	$p_Z(4) = q_1 + q_5$
6	q_6	9	$p_Z(9) = q_6$

  pdf of X: Each $q_i \geqslant 0$ and $q_1 + \cdots + q_6 = 1$.
  pdf of Z: Each probability is also $\geqslant 0$, 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:

$$\begin{aligned} E(aX+b) &= \sum_x p_X(x)(ax+b) = a\sum_x x p_X(x) + b\sum_x p_X(x) \\ &= aE(X) + b\cdot 1 = aE(X)+b \end{aligned}$$

- $E(a\,g(X)) = aE(g(X))$
  $E(a) = a$
  $E(g(X,Y)+h(X,Y)) = E(g(X,Y)) + E(h(X,Y))$

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

$$\begin{aligned} E(XY) &= \sum_x \sum_y p_{X,Y}(x,y)\cdot xy \\ &= \sum_x \sum_y p_X(x)p_Y(y)\cdot xy \quad \text{if } X, Y \text{ independent!} \\ &= \left(\sum_x p_X(x)x\right)\left(\sum_y p_Y(y)y\right) = E(X)E(Y) \end{aligned}$$

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 $E(U)E(V) = \frac{49}{4}$.
- $$E(UV) = \frac{1\cdot1+2\cdot2+3\cdot3+4\cdot4+5\cdot5+6\cdot6}{6} = \frac{91}{6}$$

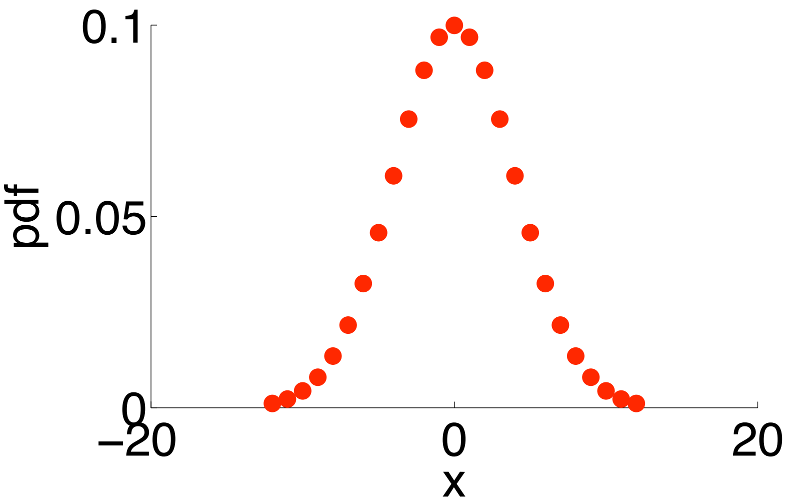
Example (Independent)

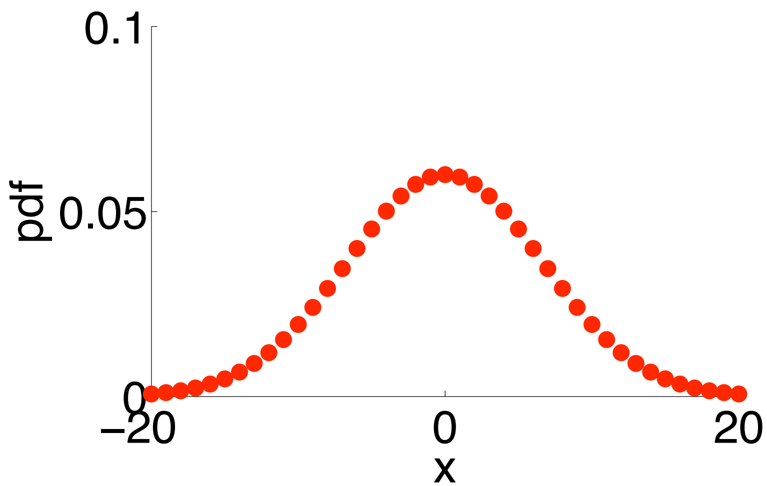
- Now let U, V be the values of two independent rolls of a fair 6-sided die.
- $$E(UV) = \sum_{x=1}^{6}\sum_{y=1}^{6}\frac{x\cdot y}{36} = \frac{441}{36} = \frac{49}{4}$$

  and $E(U)E(V) = (7/2)(7/2) = 49/4$

Variance

- These distributions both have mean=0, but the right one is more spread out.



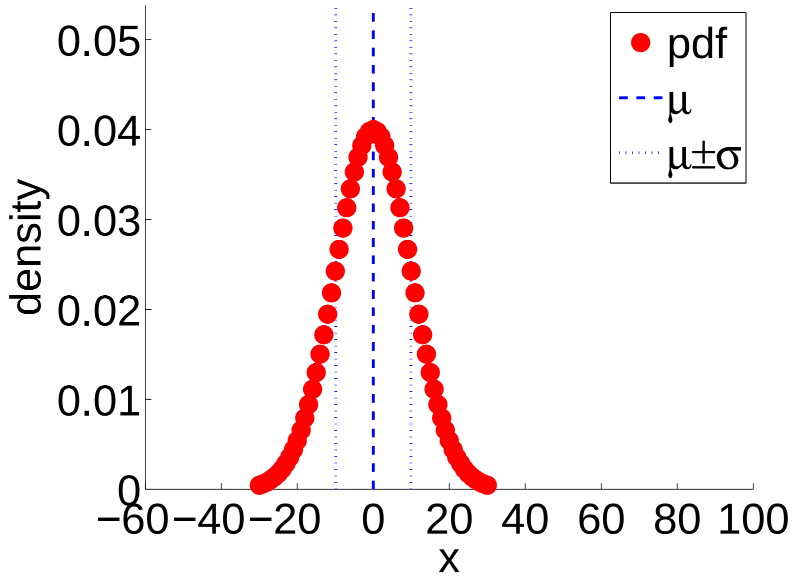


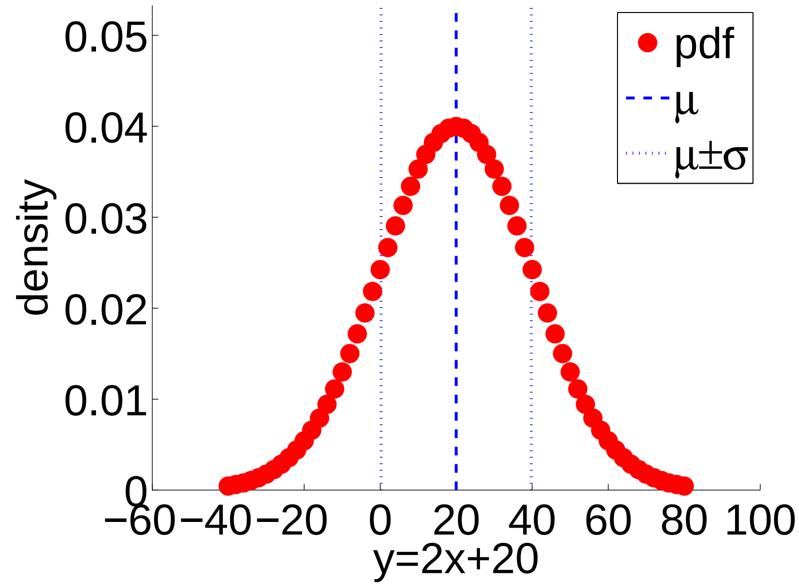
- Variance measures the square of the spread from the mean:

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

- Standard deviation measures how wide the curve is:

$$\sigma = \mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}$$





- $\text{Var}(aX+b) = a^2\,\text{Var}(X) \qquad \text{SD}(aX+b) = |a|\,\text{SD}(X)$
- 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 $^\circ C$ to $^\circ F$:
  $F = (9/5)C + 32 \qquad \mu_F = (9/5)\mu_C + 32 \qquad \sigma_F = (9/5)\sigma_C$

Variance — properties

Useful alternative formula for variance

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

Proof.

$$\begin{aligned}\mathrm{Var}(X) = E((X-\mu)^2) &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - 2\mu \cdot \mu + \mu^2 = E(X^2) - \mu^2 \qquad \square\end{aligned}$$

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

$$\begin{aligned}E((aX+b)^2) &= E(a^2X^2 + 2ab\,X + b^2) &&= a^2E(X^2) + 2ab\,E(X) + b^2 \\ (E(aX+b))^2 &= (aE(X)+b)^2 &&= a^2(E(X))^2 + 2ab\,E(X) + b^2 \\ \mathrm{Var}(aX+b) &= \text{difference} &&= a^2\left(E(X^2) - (E(X))^2\right) \\ & &&= a^2\,\mathrm{Var}(X) \qquad \square\end{aligned}$$

Variance of a sum — dependent variables

- We will show that if X, Y are independent, then

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

Example (Dependent)

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

$$\begin{aligned} \operatorname{Var}(X+Y) &= \operatorname{Var}(0) = 0 \\ \operatorname{Var}(X) + \operatorname{Var}(Y) &= \operatorname{Var}(X) + \operatorname{Var}(-X) \\ &= \operatorname{Var}(X) + (-1)^2 \operatorname{Var}(X) = 2 \operatorname{Var}(X) \end{aligned}$$

but usually $\operatorname{Var}(X) \neq 0$ (the only exception would be if X is a constant).

Variance of a sum — independent variables

Theorem

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

Proof.

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

$$\begin{aligned}
\mathrm{Var}(X+Y) &= E((X+Y)^2) - (E(X+Y))^2 \\
&= \left(E(X^2) - (E(X))^2\right) \\
&\qquad + 2\,(E(XY) - E(X)E(Y)) \\
&\qquad + \left(E(Y^2) - (E(Y))^2\right) \\
&= \mathrm{Var}(X) + 2(E(XY) - E(X)E(Y)) + \mathrm{Var}(Y)
\end{aligned}$$

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:

$$\mathrm{Var}(X+Y+Z+\cdots) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Var}(Z) + \cdots$$

$$\mathrm{Var}(aX+bY+cZ+\cdots) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + c^2\,\mathrm{Var}(Z) + \cdots$$

Variance of a sum — dependent variables

Covariance

- For dependent variables, the cross-terms remain:

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + 2(E(XY) - E(X)E(Y)) + \operatorname{Var}(Y)$$

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

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y)$$

Two formulas for covariance:

$$\operatorname{Cov}(X,Y) \;=\; E((X-\mu_X)(Y-\mu_Y)) \;=\; E(XY) - E(X)E(Y)$$

$$\begin{aligned}
E((X-\mu_X)(Y-\mu_Y)) &= E(XY) - \quad \mu_X E(Y) - E(X)\mu_Y \quad + \mu_X\mu_Y \\
&= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \\
&= E(XY) - E(X)E(Y)
\end{aligned}$$

Covariance properties

$$\begin{aligned}
\mathrm{Var}(X) &= E((X-\mu_X)^2) &&= E(X^2)-(E(X))^2 \\
\mathrm{Cov}(X,Y) &= E((X-\mu_X)(Y-\mu_Y)) &&= E(XY)-E(X)E(Y)
\end{aligned}$$

Additional properties

- $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$
- $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$
- If X, Y are independent then $\mathrm{Cov}(X,Y) = 0$.
  Beware, this is not reversible: $\mathrm{Cov}(X,Y)$ ***could be 0 for dependent variables.***
- $\mathrm{Cov}(aX+b, cY+d) = ac\,\mathrm{Cov}(X,Y)$ (a, b, c, d are constants)
- $\mathrm{Cov}(X+Z,Y) = \mathrm{Cov}(X,Y)+\mathrm{Cov}(Z,Y)$ and
  $\mathrm{Cov}(X,Y+Z) = \mathrm{Cov}(X,Y)+\mathrm{Cov}(X,Z)$
- $\mathrm{Var}(X_1+X_2+\cdots+X_n) = \mathrm{Var}(X_1)+\cdots+\mathrm{Var}(X_n)+2\sum_{1\leqslant i<j\leqslant n}\mathrm{Cov}(X_i,X_j)$

Mean and variance of the Binomial Distribution

- A Bernoulli trial is a single coin flip,
$$P(\text{heads}) = p, \qquad P(\text{tails}) = 1 - p = q.$$
- 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 $HTTHT$: $\quad X = 1 + 0 + 0 + 1 + 0 = 2$.
- $X_1, \ldots, X_n$ are independent and have the same pdfs, so they are i.i.d. (independent identically distributed) random variables.

- $$\begin{aligned} E(X_1) &= 0(1-p) + 1p = p \\ E({X_1}^2) &= 0^2(1-p) + 1^2 p = p \\ \mathrm{Var}(X_1) &= E({X_1}^2) - (E(X_1))^2 = p - p^2 = p(1-p) \end{aligned}$$
- $E(X_i) = p$ and $\mathrm{Var}(X_i) = p(1-p)$ 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:

$$\begin{aligned} \mu_X = E(X) &= E(X_1 + \cdots + X_n) \\ &= E(X_1) + \cdots + E(X_n) \\ &= p + \cdots + p = np \quad \text{identically distributed} \end{aligned}$$

Variance:

$$\begin{aligned} {\sigma_X}^2 = \text{Var}(X) &= \text{Var}(X_1 + \cdots + X_n) \\ &= \text{Var}(X_1) + \cdots + \text{Var}(X_n) \quad \text{by independence} \\ &= p(1-p) + \cdots + p(1-p) \quad \text{identically distributed} \\ &= np(1-p) = npq \end{aligned}$$

Standard deviation:

$$\sigma_X = \sqrt{np(1-p)} = \sqrt{npq}$$

Mean and variance of the Binomial Distribution

- For the binomial distribution,

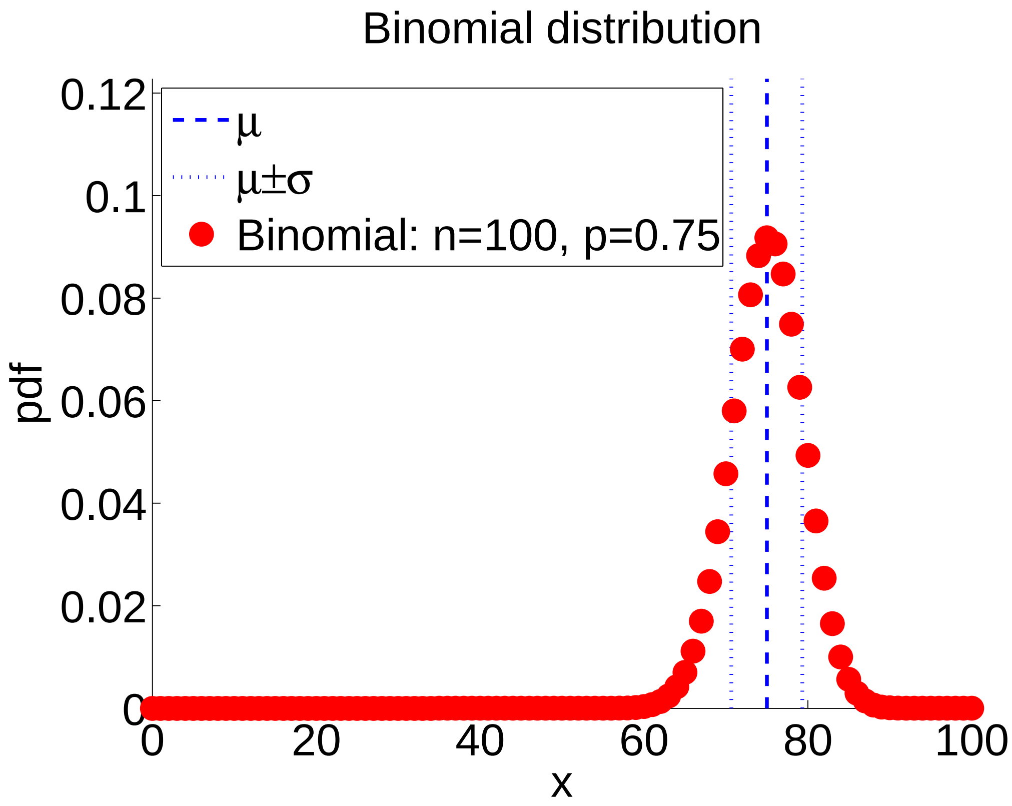
  Mean: $\mu = np$

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

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

- At $n = 100$ and $p = 3/4$:

  $\mu = 100(3/4) = 75$
  $\sigma = \sqrt{100(3/4)(1/4)} \approx 4.33$



- 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.

Geometric Distribution

- Consider a biased coin with probability p of heads.
- Flip it repeatedly (potentially ∞ times).
- Let X be the number of flips until the first head.
- **Example:** $TTTHTTHHT$ 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:** $\mu = \frac{1}{p}$ **Variance:** $\sigma^2 = \frac{1-p}{p^2}$ **Std dev:** $\sigma = \frac{\sqrt{1-p}}{p}$

Negative Binomial Distribution

- Consider a biased coin with probability p of heads.
- Flip it repeatedly (potentially ∞ times).
- Let X be the number of flips until the rth 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;
  - **kth 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 $k = r, r+1, r+2, \ldots$;

$$p_X(k) = 0 \text{ otherwise.}$$

Negative Binomial Distribution – mean and variance

- Consider the sequence of flips $TTTHTHHTTH$.
- Break it up at each heads:

$$\underbrace{TTTH}_{X_1=4} / \underbrace{TH}_{X_2=2} / \underbrace{H}_{X_3=1} / \underbrace{TTH}_{X_4=3}$$

- 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; ...
- The X_i's are i.i.d. geometric random variables with parameter p, and $X = X_1 + \cdots + X_r$.
- **Mean:** $E(X) = E(X_1) + \cdots + E(X_r) = \frac{1}{p} + \cdots + \frac{1}{p} = \frac{r}{p}$

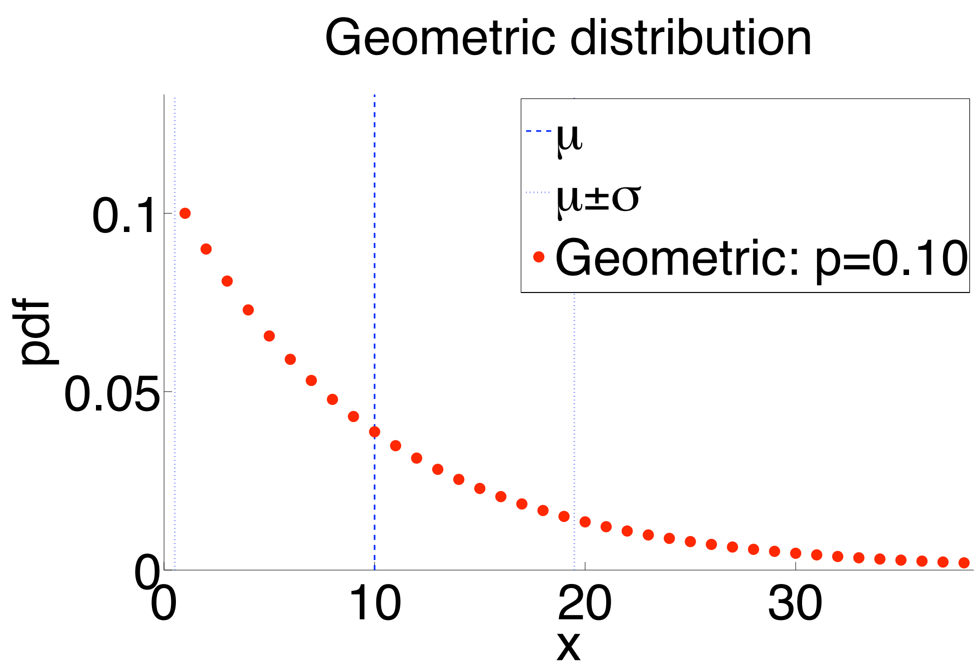
  Variance: $\sigma^2 = \frac{1-p}{p^2} + \cdots + \frac{1-p}{p^2} = \frac{r(1-p)}{p^2}$

  Standard deviation: $\sigma = \frac{\sqrt{r(1-p)}}{p}$

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 $p = .1$:

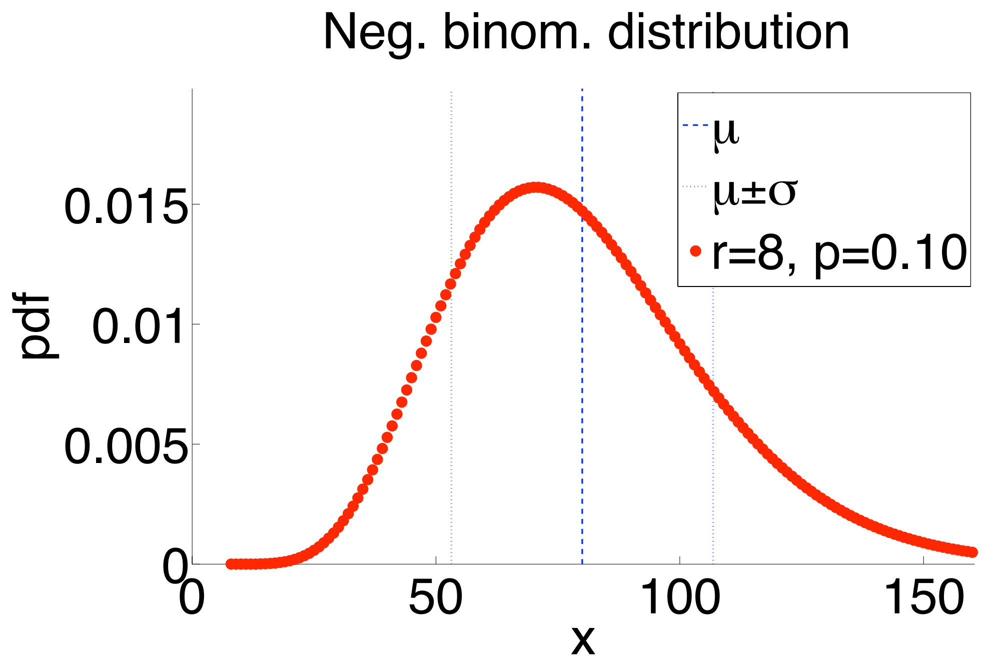
$$p_X(x) = .9^{x-1} \cdot .1 \qquad \text{for } x = 1, 2, 3, \ldots$$



- **Mean:** $\frac{1}{p} = \frac{1}{.1} = 10.$
  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, $r = 8$, $p = .1$.
  $$p_X(x) = \binom{x-1}{8-1}(.1)^8(.9)^{x-8} \qquad \text{for } x = 8, 9, 10, \ldots$$



- **Mean:** $r/p = 8/.1 = 80$.
  Standard deviation: $\dfrac{\sqrt{r(1-p)}}{p} = \dfrac{\sqrt{8(.9)}}{.1} \approx 26.833$.
- **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 do the distribution names come from?

The PDFs correspond to the terms in certain Taylor series

Geometric series

- For real a, x with $|x| < 1$,

$$\frac{a}{1-x} = \sum_{i=0}^{\infty} a\,x^i$$
$$= a + ax + ax^2 + \cdots$$

- Total probability for the geometric distribution:

$$\sum_{k=1}^{\infty} (1-p)^{k-1} p$$
$$= \frac{p}{1-(1-p)}$$
$$= \frac{p}{p} = 1$$

Negative binomial series

- For integer $r > 0$ and real x with $|x| < 1$,

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

- Total probability for the negative binomial distribution:

$$\sum_{k=r}^{\infty} \binom{k-1}{r-1} p^r (1-p)^{k-r}$$
$$= p^r \sum_{k=r}^{\infty} \binom{k-1}{r-1} (1-p)^{k-r}$$
$$= p^r \cdot \frac{1}{(1-(1-p))^r} = 1$$

Geometric and Negative Binomial – versions

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 $Y = 3$.

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

  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 rth heads, so $Y = X - r$.

  $$p_Y(k) = \begin{cases} \binom{k+r-1}{r-1} p^r (1-p)^k & \text{for } k = 0, 1, 2, \ldots; \\ 0 & \text{otherwise} \end{cases}$$
- Versions 3 and 4: switch the roles of heads and tails in the first two versions (so p and $1-p$ are switched).