6. random variables

let $X$ = index of

T T T T T H T H H
A *random variable* is some (usually numeric) *function* of the outcome, not in the outcome itself.

**Ex.**

Let $H$ be the number of Heads when 20 coins are tossed

Let $T$ be the total of 2 dice rolls

Let $X$ be the number of coin tosses needed to see 1\textsuperscript{st} head

Note; even if the underlying experiment has “equally likely outcomes,” the associated random variable may not

<table>
<thead>
<tr>
<th>Outcome</th>
<th>$H$</th>
<th>$P(H)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TT</td>
<td>0</td>
<td>$P(H=0) = 1/4$</td>
</tr>
<tr>
<td>TH</td>
<td>1</td>
<td>${P(H=1) = 1/2$</td>
</tr>
<tr>
<td>HT</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>HH</td>
<td>2</td>
<td>$P(H=2) = 1/4$</td>
</tr>
</tbody>
</table>
20 balls numbered 1, 2, ..., 20
Draw 3 without replacement
Let $X = \text{the maximum of the numbers on those 3 balls}$

What is $P(X \geq 17)$

\[
P(X = 20) = \begin{pmatrix} 19 \\ 2 \end{pmatrix} / \begin{pmatrix} 20 \\ 3 \end{pmatrix} = \frac{3}{20} = 0.150
\]
\[
P(X = 19) = \begin{pmatrix} 18 \\ 2 \end{pmatrix} / \begin{pmatrix} 20 \\ 3 \end{pmatrix} = \frac{18 \cdot 17/2!}{20 \cdot 19 \cdot 18/3!} \approx 0.134
\]
\[
\vdots
\]
\[
\sum_{i=17}^{20} P(X = i) \approx 0.508
\]

Alternatively:

\[
P(X \geq 17) = 1 - P(X < 17) = 1 - \begin{pmatrix} 16 \\ 3 \end{pmatrix} / \begin{pmatrix} 20 \\ 3 \end{pmatrix} \approx 0.508
\]
Flip a (biased) coin repeatedly until 1\textsuperscript{st} head observed

How many flips? Let $X$ be that number.

- $P(X=1) = P(H) = p$
- $P(X=2) = P(TH) = (1-p)p$
- $P(X=3) = P(TTH) = (1-p)^2p$

... 

Check that it is a valid probability distribution:

$$P \left( \bigcup_{i \geq 1} \{X = i\} \right) = \sum_{i \geq 1} (1-p)^{i-1} p = p \sum_{i \geq 0} (1-p)^i = p \frac{1}{1 - (1-p)} = 1$$
A *discrete* random variable is one taking on a countable number of possible values.

**Ex:**

- $X = \text{sum of 3 dice, } 3 \leq X \leq 18, X \in \mathbb{N}$
- $Y = \text{number of 1st head in seq of coin flips, } 1 \leq Y, Y \in \mathbb{N}$
- $Z = \text{largest prime factor of } (1+Y), \quad Z \in \{2, 3, 5, 7, 11, \ldots\}$

If $X$ is a discrete random variable taking on values from a countable set $T \subseteq \mathbb{R}$, then

$$p(a) = \begin{cases} P(X = a) & \text{for } a \in T \\ 0 & \text{otherwise} \end{cases}$$

is called the *probability mass function*. Note: $\sum_{a \in T} p(a) = 1$
Let $X$ be the number of heads observed in $n$ coin flips

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \text{ where } p = P(H)$$

Probability mass function:
The **cumulative distribution function** for a random variable $X$ is the function $F: \mathbb{R} \rightarrow [0,1]$ defined by

$$F(a) = P[X \leq a]$$

**Ex:** if $X$ has **probability mass function** given by:

$$p(1) = \frac{1}{4} \quad p(2) = \frac{1}{2} \quad p(3) = \frac{1}{8} \quad p(4) = \frac{1}{8}$$

$$F(a) = \begin{cases} 0 & a < 1 \\ \frac{1}{4} & 1 \leq a < 2 \\ \frac{3}{4} & 2 \leq a < 3 \\ \frac{7}{8} & 3 \leq a < 4 \\ 1 & 4 \leq a \end{cases}$$

**NB:** for discrete random variables, be careful about “$\leq$” vs “$<$”
For a discrete r.v. $X$ with p.m.f. $p(•)$, the *expectation of $X$*, aka *expected value* or *mean*, is

$$E[X] = \sum_x xp(x)$$

For the equally-likely outcomes case, this is just the average of the possible random values of $X$

For *unequally-likely* outcomes, it is again the average of the possible random values of $X$, *weighted by their respective probabilities*.

**Ex 1:** Let $X =$ value seen rolling a fair die $p(1), p(2), \ldots, p(6) = 1/6$

$$E[X] = \sum_{i=1}^{6} ip(i) = \frac{1}{6} (1 + 2 + \cdots + 6) = \frac{21}{6} = 3.5$$

**Ex 2:** Coin flip; $X = +1$ if H (win $\$1$), $-1$ if T (lose $\$1$)

$$E[X] = (+1)p(+1) + (-1)p(-1) = 1 \cdot (1/2) + (-1) \cdot (1/2) = 0$$
For a discrete r.v. \( X \) with p.m.f. \( p(\cdot) \), the \textit{expectation of} \( X \), aka \textit{expected value} or \textit{mean}, is

\[
E[X] = \sum_x x p(x)
\]

Another view: A 2-person gambling game. If \( X \) is how much you win playing the game once, how much would you expect to win, on average, per game when repeatedly playing?

\begin{enumerate}
\item Ex 1: Let \( X \) = value seen rolling a fair die \( p(1), p(2), \ldots, p(6) = 1/6 \)
If you win \( X \) dollars for that roll, how much do you expect to win?
\[
E[X] = \sum_{i=1}^{6} ip(i) = \frac{1}{6} (1 + 2 + \cdots + 6) = \frac{21}{6} = 3.5
\]
\item Ex 2: Coin flip; \( X = +1 \) if H (win $1), -1 if T (lose $1)
\[
E[X] = (+1)\cdot p(+1) + (-1)\cdot p(-1) = 1\cdot(1/2) +(-1)\cdot(1/2) = 0
\]
“a fair game”: in repeated play you expect to win as much as you lose. Long term net gain/loss = 0.
Let $X$ be the number of flips up to & including 1$^{st}$ head observed in repeated flips of a biased coin. If I pay you $1 per flip, how much money would you expect to make?

$$P(H) = p; \quad P(T) = 1 - p = q$$

$$p(i) = pq^{i-1}$$

$$E(x) = \sum_{i \geq 1} ip(i) = \sum_{i \geq 1} ipq^{i-1} = p \sum_{i \geq 1} iq^{i-1} \quad (*)$$

A calculus trick:

$$\sum_{i \geq 1} iy^{i-1} = \sum_{i \geq 1} \frac{d}{dy} y^i = \sum_{i \geq 0} \frac{d}{dy} y^i = \frac{d}{dy} \sum_{i \geq 0} y^i = \frac{d}{dy} \frac{1}{1 - y} = \frac{1}{(1 - y)^2}$$

So $(*)$ becomes:

$$E[X] = p \sum_{i \geq 1} iq^{i-1} = \frac{p}{(1 - q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

E.g.:

- $p = 1/2$; on average head every 2$^{nd}$ flip
- $p = 1/10$; on average, head every 10$^{th}$ flip.

How much would you pay to play?
Calculating $E[g(X)]$:

$Y = g(X)$ is a new r.v. Calc $P[Y = j]$, then apply defn:

- $X = \text{sum of 2 dice rolls}$
- $Y = g(X) = X \mod 5$

<table>
<thead>
<tr>
<th>$i$</th>
<th>$p(i) = P[X=i]$</th>
<th>$i \cdot p(i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>$\frac{1}{36}$</td>
<td>$\frac{2}{36}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{2}{36}$</td>
<td>$\frac{6}{36}$</td>
</tr>
<tr>
<td>4</td>
<td>$\frac{3}{36}$</td>
<td>$\frac{12}{36}$</td>
</tr>
<tr>
<td>5</td>
<td>$\frac{4}{36}$</td>
<td>$\frac{20}{36}$</td>
</tr>
<tr>
<td>6</td>
<td>$\frac{5}{36}$</td>
<td>$\frac{30}{36}$</td>
</tr>
<tr>
<td>7</td>
<td>$\frac{6}{36}$</td>
<td>$\frac{42}{36}$</td>
</tr>
<tr>
<td>8</td>
<td>$\frac{5}{36}$</td>
<td>$\frac{40}{36}$</td>
</tr>
<tr>
<td>9</td>
<td>$\frac{4}{36}$</td>
<td>$\frac{36}{36}$</td>
</tr>
<tr>
<td>10</td>
<td>$\frac{3}{36}$</td>
<td>$\frac{30}{36}$</td>
</tr>
<tr>
<td>11</td>
<td>$\frac{2}{36}$</td>
<td>$\frac{22}{36}$</td>
</tr>
<tr>
<td>12</td>
<td>$\frac{1}{36}$</td>
<td>$\frac{12}{36}$</td>
</tr>
</tbody>
</table>

$E[X] = \sum_i i \cdot p(i) = \frac{252}{36} = 7$

$Y = g(X) = X \mod 5$

<table>
<thead>
<tr>
<th>$j$</th>
<th>$q(j) = P[Y = j]$</th>
<th>$j \cdot q(j)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\frac{4}{36} + \frac{3}{36} = \frac{7}{36}$</td>
<td>$0 \cdot \frac{7}{36}$</td>
</tr>
<tr>
<td>1</td>
<td>$\frac{5}{36} + \frac{2}{36} = \frac{7}{36}$</td>
<td>$7 \cdot \frac{7}{36}$</td>
</tr>
<tr>
<td>2</td>
<td>$\frac{1}{36} + \frac{6}{36} + \frac{1}{36} = \frac{8}{36}$</td>
<td>$16 \cdot \frac{8}{36}$</td>
</tr>
<tr>
<td>3</td>
<td>$\frac{2}{36} + \frac{5}{36} = \frac{7}{36}$</td>
<td>$21 \cdot \frac{7}{36}$</td>
</tr>
<tr>
<td>4</td>
<td>$\frac{3}{36} + \frac{4}{36} = \frac{7}{36}$</td>
<td>$28 \cdot \frac{7}{36}$</td>
</tr>
</tbody>
</table>

$E[Y] = \sum_j j \cdot q(j) = \frac{72}{36} = 2$
Calculating $E[g(X)]$: Another way – *add in a different order*, using $P[X=...]$ instead of calculating $P[Y=...]$.

$X = \text{sum of 2 dice rolls}$

$Y = g(X) = X \mod 5$

<table>
<thead>
<tr>
<th>$i$</th>
<th>$p(i) = P[X=i]$</th>
<th>$g(i) \cdot p(i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1/36</td>
<td>2/36</td>
</tr>
<tr>
<td>3</td>
<td>2/36</td>
<td>6/36</td>
</tr>
<tr>
<td>4</td>
<td>3/36</td>
<td>12/36</td>
</tr>
<tr>
<td>5</td>
<td>4/36</td>
<td>0/36</td>
</tr>
<tr>
<td>6</td>
<td>5/36</td>
<td>5/36</td>
</tr>
<tr>
<td>7</td>
<td>6/36</td>
<td>12/36</td>
</tr>
<tr>
<td>8</td>
<td>5/36</td>
<td>15/36</td>
</tr>
<tr>
<td>9</td>
<td>4/36</td>
<td>16/36</td>
</tr>
<tr>
<td>10</td>
<td>3/36</td>
<td>0/36</td>
</tr>
<tr>
<td>11</td>
<td>2/36</td>
<td>2/36</td>
</tr>
<tr>
<td>12</td>
<td>1/36</td>
<td>2/36</td>
</tr>
</tbody>
</table>

$E[Y] = \Sigma_i j \cdot q(j) = \frac{72}{36} = 2$

$E[g(X)] = \Sigma_i g(i) \cdot p(i) = \frac{72}{36} = 2$
Above example is not a fluke.

**Theorem:** if $Y = g(X)$, then $E[Y] = \sum_i g(x_i)p(x_i)$, where $x_i, i = 1, 2, ...$ are all possible values of $X$.

**Proof:** Let $y_j, j = 1, 2, ...$ be all possible values of $Y$.

$$\sum_i g(x_i)p(x_i) = \sum_j \sum_{i: g(x_i)=y_j} g(x_i)p(x_i)$$

$$= \sum_j \sum_{i: g(x_i)=y_j} y_jp(x_i)$$

$$= \sum_j y_j \sum_{i: g(x_i)=y_j} p(x_i)$$

$$= \sum_j y_j P\{g(X) = y_j\}$$

$$= E[g(X)]$$

Note that $S_j = \{ x_i \mid g(x_i)=y_j \}$ is a partition of the domain of $g$. 

**expectation of a function of a random variable**
properties of expectation

A & B each bet $1, then flip 2 coins:

<table>
<thead>
<tr>
<th>HH</th>
<th>A wins $2</th>
</tr>
</thead>
<tbody>
<tr>
<td>HT</td>
<td>Each takes back $1</td>
</tr>
<tr>
<td>TH</td>
<td>back $1</td>
</tr>
<tr>
<td>TT</td>
<td>B wins $2</td>
</tr>
</tbody>
</table>

Let $X$ be A’s net gain: +1, 0, -1, resp.:

<table>
<thead>
<tr>
<th>$P(X = +1) = 1/4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(X = 0) = 1/2$</td>
</tr>
<tr>
<td>$P(X = -1) = 1/4$</td>
</tr>
</tbody>
</table>

What is $E[X]$?

$$E[X] = 1 \cdot 1/4 + 0 \cdot 1/2 + (-1) \cdot 1/4 = 0$$

What is $E[X^2]$?

$$E[X^2] = 1^2 \cdot 1/4 + 0^2 \cdot 1/2 + (-1)^2 \cdot 1/4 = 1/2$$

Note: $E[X^2] \neq E[X]^2$
properties of expectation

Linearity of expectation, I

For any constants $a, b$: $E[aX + b] = aE[X] + b$

Proof:

$$E[aX + b] = \sum_x (ax + b) \cdot p(x)$$

$$= a \sum_x xp(x) + b \sum_x p(x)$$

$$= aE[X] + b$$

Example:

Q: In the 2-person coin game above, what is $E[2X+1]$?

A: $E[2X+1] = 2E[X]+1 = 2\cdot0 + 1 = 1$
Linearity, II

Let $X$ and $Y$ be two random variables derived from outcomes of a single experiment. Then

$$E[X+Y] = E[X] + E[Y]$$

True even if $X, Y$ dependent

Proof: Assume the sample space $S$ is countable. (The result is true without this assumption, but I won’t prove it.) Let $X(s), Y(s)$ be the values of these r.v.’s for outcome $s \in S$.

Claim: $E[X] = \sum_{s \in S} X(s) \cdot p(s)$

Proof: similar to that for “expectation of a function of an r.v.,” i.e., the events “$X=x$” partition $S$, so sum above can be rearranged to match the definition of $E[X] = \sum_x x \cdot P(X = x)$

Then:

$$E[X+Y] = \sum_{s \in S} (X[s] + Y[s]) \cdot p(s)$$

$$= \sum_{s \in S} X[s] \cdot p(s) + \sum_{s \in S} Y[s] \cdot p(s) = E[X] + E[Y]$$
Example

$X = \# \text{ of heads in one coin flip, where } P(X=1) = p.$

What is $E(X)$?

$$E[X] = 1 \cdot p + 0 \cdot (1-p) = p$$

Let $X_i, 1 \leq i \leq n$, be $\# \text{ of H in flip of coin with } P(X_i=1) = p_i$

What is the expected number of heads when all are flipped?

$$E[\sum_i X_i] = \sum_i E[X_i] = \sum_i p_i$$

Special case: $p_1 = p_2 = \ldots = p$:

$$E[\# \text{ of heads in } n \text{ flips}] = pn$$
properties of expectation

Note:

Linearity is special!

It is not true in general that

\[ E[X \cdot Y] \neq E[X] \cdot E[Y] \]
\[ E[X^2] \neq E[X]^2 \]
\[ E[X/Y] \neq E[X] / E[Y] \]
\[ E[asinh(X)] \neq asinh(E[X]) \]
Alice & Bob are gambling (again). $X = $ Alice’s gain per flip:

$$X = \begin{cases} 
+1 & \text{if Heads} \\
-1 & \text{if Tails}
\end{cases}$$

$E[X] = 0$

... Time passes ...

Alice (yawning) says “let’s raise the stakes”

$$Y = \begin{cases} 
+1000 & \text{if Heads} \\
-1000 & \text{if Tails}
\end{cases}$$

$E[Y] = 0$, as before.

Are you (Bob) equally happy to play the new game?
E[X] measures the “average” or “central tendency” of X. What about its variability?

If E[X] = μ, then E[|x-μ|] seems like a natural quantity to look at: how much do we expect X to deviate from its average. Unfortunately, it’s a bit inconvenient mathematically; following is easier/more common.

**Definition**

The *variance* of a random variable X with mean E[X] = μ is Var[X] = E[(X-μ)^2], often denoted σ^2.

The *standard deviation* of X is σ = √Var[X]
what does variance tell us?

The variance of a random variable $X$ with mean $E[X] = \mu$ is $\text{Var}[X] = E[(X-\mu)^2]$, often denoted $\sigma^2$.

I:

Square always $\geq 0$, and exaggerated as $X$ moves away from $\mu$, so $\text{Var}[X]$ emphasizes deviation from the mean.

II:

Numbers vary a lot depending on exact distribution of $X$, but typically $X$ is

- within $\mu \pm \sigma \sim 66\%$ of the time, and
- within $\mu \pm 2\sigma \sim 95\%$ of the time.

(We’ll see the reasons for this soon.)
mean and variance

$\mu = E[X]$ is about location; $\sigma = \sqrt{\text{Var}(X)}$ is about spread
Alice & Bob are gambling (again). $X = \text{Alice’s gain per flip}$:

$$X = \begin{cases} +1 & \text{if Heads} \\ -1 & \text{if Tails} \end{cases}$$

$E[X] = 0$ \hspace{2cm} $\text{Var}[X] = 1$

... Time passes ... 

Alice (yawning) says “let’s raise the stakes”

$$Y = \begin{cases} +1000 & \text{if Heads} \\ -1000 & \text{if Tails} \end{cases}$$

$E[Y] = 0$, as before. \hspace{2cm} $\text{Var}[Y] = 1,000,000$

Are you (Bob) equally happy to play the new game?
Two games:

a) flip 1 coin, win $Y = \$100$ if heads, $-\$100$ if tails

b) flip 100 coins, win $Z = (#(\text{heads}) - #(\text{tails}))$ dollars

Same expectation in both: $E[Y] = E[Z] = 0$

Same extremes in both: max gain = $\$100$; max loss = $\$100$

But variability is very different:
\[ \text{Var}(X) = E[X^2] - (E[X])^2 \]

\[ \text{Var}(X) = E[(X - \mu)^2] \]
\[ = \sum_x (x - \mu)^2 p(x) \]
\[ = \sum_x (x^2 - 2\mu x + \mu^2) p(x) \]
\[ = \sum_x x^2 p(x) - 2\mu \sum_x xp(x) + \mu^2 \sum_x p(x) \]
\[ = E[X^2] - 2\mu^2 + \mu^2 \]
\[ = E[X^2] - \mu^2 \]
Example:

What is $\text{Var}[X]$ when $X$ is outcome of one fair die?

\[
E[X^2] = 1^2 \left( \frac{1}{6} \right) + 2^2 \left( \frac{1}{6} \right) + 3^2 \left( \frac{1}{6} \right) + 4^2 \left( \frac{1}{6} \right) + 5^2 \left( \frac{1}{6} \right) + 6^2 \left( \frac{1}{6} \right) = \left( \frac{1}{6} \right)(91)
\]

$E[X] = 7/2$, so

\[
\text{Var}(X) = \frac{91}{6} - \left( \frac{7}{2} \right)^2 = \frac{35}{12}
\]
properties of variance

\[ \text{Var}[aX+b] = a^2 \text{Var}[X] \]

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

Ex:

\[
X = \begin{cases} 
+1 & \text{if Heads} \\
-1 & \text{if Tails}
\end{cases}
\]

\[
E[X] = 0 \\
\text{Var}[X] = 1
\]

\[
Y = 1000X
\]

\[
E[Y] = E[1000X] = 1000 E[X] = 0 \\
\text{Var}[Y] = \text{Var}[1000X] \\
= 10^6 \text{Var}[X] = 10^6
\]
In general:

\[ \text{Var}[X+Y] \neq \text{Var}[X] + \text{Var}[Y] \]

Ex 1:

Let \( X = \pm 1 \) based on 1 coin flip

As shown above, \( E[X] = 0, \text{Var}[X] = 1 \)

Let \( Y = -X; \) then \( \text{Var}[Y] = (-1)^2 \text{Var}[X] = 1 \)

But \( X+Y = 0, \) always, so \( \text{Var}[X+Y] = 0 \)

Ex 2:

As another example, is \( \text{Var}[X+X] = 2\text{Var}[X]? \)
a zoo of (discrete) random variables
An experiment results in “Success” or “Failure”

X is a random *indicator variable* (1=succes, 0=failure)

\[ P(X=1) = p \quad \text{and} \quad P(X=0) = 1-p \]

X is called a *Bernoulli* random variable: \( X \sim \text{Ber}(p) \)

\[ E[X] = E[X^2] = p \]

\[ \text{Var}(X) = E[X^2] - (E[X])^2 = p - p^2 = p(1-p) \]

Examples:

- coin flip
- random binary digit
- whether a disk drive crashed

Jacob (aka James, Jacques) Bernoulli, 1654 – 1705
Consider \( n \) independent random variables \( Y_i \sim \text{Ber}(p) \)

\( X = \sum_i Y_i \) is the number of successes in \( n \) trials

\( X \) is a Binomial random variable: \( X \sim \text{Bin}(n,p) \)

\[
P(X = i) = \binom{n}{i} p^i (1-p)^{n-i} \quad i = 0, 1, \ldots, n
\]

By Binomial theorem,

\[
\sum_{i=0}^{n} P(X = i) = 1
\]

Examples

- # of heads in \( n \) coin flips
- # of 1’s in a randomly generated length \( n \) bit string
- # of disk drive crashes in a 1000 computer cluster

\[
E[X] = pn
\]

\[
\text{Var}(X) = p(1-p)n
\]

\( \leftarrow \) (proof below, twice)
PMF for $X \sim \text{Bin}(10, 0.5)$

$P(X = k)$

PMF for $X \sim \text{Bin}(10, 0.25)$

$P(X = k)$

binomial pmfs
binomial pmfs

PMF for $X \sim \text{Bin}(30, 0.5)$

PMF for $X \sim \text{Bin}(30, 0.1)$
mean and variance of the binomial

\[ E[X^k] = \sum_{i=0}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i} \]

\[ = \sum_{i=1}^{n} i^k \binom{n}{i} p^i (1-p)^{n-i} \]

\[ E[X^k] = np \sum_{i=1}^{n} i^{k-1} \binom{n-1}{i-1} p^{i-1} (1-p)^{n-i} \]

\[ = np \sum_{j=0}^{n-1} (j + 1)^{k-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} \]

\[ = np E[(Y+1)^{k-1}] \]

where \( Y \) is a binomial random variable with parameters \( n-1, p \).

k=1 gives: \( E[X] = np \); k=2 gives \( E[X^2] = np[(n-1)p+1] \)

hence:

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

\[ = np[(n-1)p + 1] - (np)^2 \]

\[ = np(1-p) \]
products of independent r.v.s

Theorem: If $X$ & $Y$ are independent, then $E[X \cdot Y] = E[X] \cdot E[Y]$.

Proof:

Let $x_i, y_i$, $i = 1, 2, \ldots$ be the possible values of $X, Y$.

$$E[X \cdot Y] = \sum_{i} \sum_{j} x_i \cdot y_j \cdot P(X = x_i \land Y = y_j)$$

$$= \sum_{i} \sum_{j} x_i \cdot y_j \cdot P(X = x_i) \cdot P(Y = y_j)$$

$$= \sum_{i} x_i \cdot P(X = x_i) \cdot \left( \sum_{j} y_j \cdot P(Y = y_j) \right)$$

$$= E[X] \cdot E[Y]$$

Note: NOT true in general; see earlier example $E[X^2] \neq E[X]^2$.
Theorem: If $X$ & $Y$ are independent, then

$$\text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y]$$

Proof: Let $\hat{X} = X - E[X]$ and $\hat{Y} = Y - E[Y]$.

$$E[\hat{X}] = 0 \quad E[\hat{Y}] = 0$$

$$\text{Var}[\hat{X}] = \text{Var}[X] \quad \text{Var}[\hat{Y}] = \text{Var}[Y]$$

$$\text{Var}[X + Y] = \text{Var}[\hat{X} + \hat{Y}]$$

$$= E[(\hat{X} + \hat{Y})^2] - (E[\hat{X} + \hat{Y}])^2$$

$$= E[\hat{X}^2 + 2\hat{X}\hat{Y} + \hat{Y}^2] - 0$$

$$= E[\hat{X}^2] + 2E[\hat{X}\hat{Y}] + E[\hat{Y}^2]$$

$$= \text{Var}[\hat{X}] + 0 + \text{Var}[\hat{Y}]$$

$$= \text{Var}[X] + \text{Var}[Y]$$

variance of independent r.v.s is additive

(Bienaymé, 1853)
If \( Y_1, Y_2, \ldots, Y_n \sim \text{Ber}(p) \) and independent,

then \( X = \sum_{i=1}^{n} Y_i \sim \text{Bin}(n, p) \).

\[
E[X] = E[\sum_{i=1}^{n} Y_i] = nE[Y_1] = np
\]

\[
\text{Var}[X] = \text{Var}[\sum_{i=1}^{n} Y_i] = n\text{Var}[Y_1] = np(1 - p)
\]
A RAID-like disk array consists of $n$ drives, each of which will fail independently with probability $p$. Suppose it can operate effectively if at least one-half of its components function, e.g., by “majority vote.” For what values of $p$ is a 5-component system more likely to operate effectively than a 3-component system?

$X_5 = \# \text{ failed in 5-component system} \sim \text{Bin}(5, p)$

$X_3 = \# \text{ failed in 3-component system} \sim \text{Bin}(3, p)$
$X_5 = \# \text{ failed in 5-component system} \sim \text{Bin}(5, p)$

$X_3 = \# \text{ failed in 3-component system} \sim \text{Bin}(3, p)$

$P(\text{5 component system effective}) = P(X_5 < 5/2)$

$$= \binom{5}{0}p^0(1-p)^5 + \binom{5}{1}p^1(1-p)^4 + \binom{5}{2}p^2(1-p)^3$$

$P(\text{3 component system effective}) = P(X_3 < 3/2)$

$$= \binom{3}{0}p^0(1-p)^3 + \binom{3}{1}p^1(1-p)^2$$

**Calculation:**

5-component system is better iff $p < 1/2$
Goal: send a 4-bit message over a noisy communication channel. Say, 1 bit in 10 is flipped in transit, independently.

What is the probability that the message arrives correctly?

Let \( X \) = # of errors; \( X \sim \text{Bin}(4, 0.1) \)

\[
P(\text{correct message received}) = P(X=0)
\]

\[
P(X = 0) = \binom{4}{0} (0.1)^0 (0.9)^4 = 0.6561
\]

Can we do better? Yes: error correction via redundancy.

E.g., send every bit in triplicate; use majority vote.

Let \( Y \) = # of errors in one trio; \( Y \sim \text{Bin}(3, 0.1) \); \( P(\text{a trio is OK}) = \)

\[
P(Y \leq 1) = \binom{3}{0} (0.1)^0 (0.9)^3 + \binom{3}{1} (0.1)^1 (0.9)^2 = 0.972
\]

If \( X' \) = # errors in triplicate msg, \( X' \sim \text{Bin}(4, 0.028) \), and

\[
P(X' = 0) = \binom{4}{0} (0.028)^0 (0.972)^4 = 0.8926168
\]
The Hamming(7,4) code:
Have a 4-bit string to send over the network (or to disk)
Add 3 “parity” bits, and send 7 bits total
If bits are $b_1b_2b_3b_4$ then the three parity bits are
\[ \text{parity}(b_1b_2b_3), \text{parity}(b_1b_3b_4), \text{parity}(b_2b_3b_4) \]
Each bit is independently corrupted (flipped) in transit with probability 0.1
\[ Z = \text{number of bits corrupted} \sim \text{Bin}(7, 0.1) \]
The Hamming code allow us to correct all 1 bit errors.
(E.g., if $b_1$ flipped, 1st 2 parity bits, but not 3rd, will look wrong; the only single bit error causing this symptom is $b_1$. Similarly for any other single bit being flipped. Some, but not all, multi-bit errors can be detected, but not corrected.)
\[ P(\text{correctable message received}) = P(Z \leq 1) \]
Using Hamming error-correcting codes: \( Z \sim \text{Bin}(7, 0.1) \)

\[
P(Z \leq 1) = \binom{7}{0}(0.1)^0(0.9)^7 + \binom{7}{1}(0.1)^1(0.9)^6 \approx 0.8503
\]

Recall, uncorrected success rate is

\[
P(X = 0) = \binom{4}{0}(0.1)^0(0.9)^4 = 0.6561
\]

And triplicate code error rate is:

\[
P(X' = 0) = \binom{4}{0}(0.028)^0(0.972)^4 = 0.8926168
\]

Hamming code is nearly as reliable as the triplicate code, with \( 5/12 \approx 42\% \) fewer bits. (& better with longer codes.)
Sending a bit string over the network

- $n = 4$ bits sent, each corrupted with probability $0.1$
- $X = \# \text{ of corrupted bits}, X \sim \text{Bin}(4, 0.1)$

In real networks, large bit strings (length $n \approx 10^4$)

Corruption probability is very small: $p \approx 10^{-6}$

- $X \sim \text{Bin}(10^4, 10^{-6})$ is unwieldy to compute

Extreme $n$ and $p$ values arise in many cases

- $\# \text{ bit errors in file written to disk}$
- $\# \text{ of typos in a book}$
- $\# \text{ of elements in particular bucket of large hash table}$
- $\# \text{ of server crashes per day in giant data center}$
- $\# \text{ facebook login requests sent to a particular server}$
Suppose “events” happen, independently, at an average rate of \( \lambda \) per unit time. Let \( X \) be the actual number of events happening in a given time unit. Then \( X \) is a Poisson r.v. with parameter \( \lambda \) (denoted \( X \sim \text{Poi}(\lambda) \)) and has distribution (PMF):

\[
P(X = i) = e^{-\lambda} \frac{\lambda^i}{i!}
\]

Examples:

- # of alpha particles emitted by a lump of radium in 1 sec.
- # of traffic accidents in Seattle in one year
- # of babies born in a day at UW Med center
- # of visitors to my web page today

See B&T Section 6.2 for more on theoretical basis for Poisson.
X is a Poisson r.v. with parameter $\lambda$ if it has PMF:

$$P(X = i) = e^{-\lambda} \frac{\lambda^i}{i!}$$

Is it a valid distribution? Recall Taylor series:

$$e^{\lambda} = \frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \cdots = \sum_{0 \leq i} \frac{\lambda^i}{i!}$$

So

$$\sum_{0 \leq i} P(X = i) = \sum_{0 \leq i} e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} \sum_{0 \leq i} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1$$
expected value of Poisson r.v.s

\[
E[X] = \sum_{0 \leq i} i \cdot e^{-\lambda} \frac{\lambda^i}{i!}
\]

\[
= \sum_{1 \leq i} i \cdot e^{-\lambda} \frac{\lambda^i}{i!}
\]

\[
= \lambda e^{-\lambda} \sum_{1 \leq i} \frac{\lambda^{i-1}}{(i-1)!}
\]

\[
= \lambda e^{-\lambda} \sum_{0 \leq j} \frac{\lambda^j}{j!}
\]

\[
= \lambda e^{-\lambda} e^\lambda
\]

\[
= \lambda
\]

As expected, given definition in terms of “average rate \( \lambda \)”

\( \text{(Var}[X] = \lambda, \text{too; proof similar, see B&T example 6.20)} \)
binomial random variable is Poisson in the limit

Poisson approximates binomial when $n$ is large, $p$ is small, and $\lambda = np$ is “moderate”

Different interpretations of “moderate”
- $n > 20$ and $p < 0.05$
- $n > 100$ and $p < 0.1$

Formally, Binomial is Poisson in the limit as $n \to \infty$ (equivalently, $p \to 0$) while holding $np = \lambda$
\( X \sim \text{Binomial}(n, p) \)

\[
P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i}
\]

\[
= \frac{n!}{i!(n - i)!} \left( \frac{\lambda}{n} \right)^i \left( 1 - \frac{\lambda}{n} \right)^{n-i}, \text{ where } \lambda = pn
\]

\[
= \frac{n(n-1) \cdots (n-i+1)}{n^i} \left( \frac{\lambda}{n} \right)^i \frac{\lambda^i}{i!} \left( 1 - \frac{\lambda}{n} \right)^n
\]

\[
\approx 1 \cdot \frac{\lambda^i}{i!} \cdot e^{-\lambda}
\]

I.e., Binomial \( \approx \) Poisson for large \( n \), small \( p \), moderate \( i, \lambda \).
Recall example of sending bit string over a network
Send bit string of length \( n = 10^4 \)
Probability of (independent) bit corruption is \( p = 10^{-6} \)
\( X \sim \text{Poi}(\lambda = 10^4 \cdot 10^{-6} = 0.01) \)
What is probability that message arrives uncorrupted?
\[
P(X = 0) = e^{-\lambda} \frac{\lambda^0}{0!} = e^{-0.01} \frac{0.01^0}{0!} \approx 0.990049834
\]
Using \( Y \sim \text{Bin}(10^4, 10^{-6}) \):
\[
P(Y=0) \approx 0.990049829
\]
Binomial vs Poisson

- Binomial(10, 0.3)
- Binomial(100, 0.03)
- Poisson(3)
expectation and variance of a poisson

Recall: if $Y \sim \text{Bin}(n,p)$, then:

\[ E[Y] = pn \]
\[ \text{Var}[Y] = np(1-p) \]

And if $X \sim \text{Poi}(\lambda)$ where $\lambda = np \ (n \to \infty, p \to 0)$ then

\[ E[X] = \lambda = np = E[Y] \]
\[ \text{Var}[X] = \lambda \approx \lambda(1-\lambda/n) = np(1-p) = \text{Var}[Y] \]

Expectation and variance of Poisson are the same ($\lambda$)
Expectation is the same as corresponding binomial
Variance almost the same as corresponding binomial

Note: when two different distributions share the same mean & variance, it suggests (but doesn’t prove) that one may be a good approximation for the other.
Suppose a server can process 2 requests per second. Requests arrive at random at an average rate of 1/sec. Unprocessed requests are held in a buffer.

**Q. How big a buffer do we need to avoid ever dropping a request?**

**A. Infinite**

**Q. How big a buffer do we need to avoid dropping a request more often than once a day?**

**A. (approximate)** If \( X \) is the number of arrivals in a second, then \( X \) is Poisson (\( \lambda = 1 \)). We want \( b \) s.t.

\[
P(X > b) < \frac{1}{24*60*60} \approx 1.2 \times 10^{-5}
\]

\[
P(X = b) = e^{-1}/b! \quad \sum_{i \geq 8} P(X=i) \approx P(X=8) \approx 10^{-5}
\]

Also look at probability of 10 arrivals in 2 seconds, 12 in 3 seconds, etc.
In a series $X_1, X_2, \ldots$ of Bernoulli trials with success probability $p$, let $Y$ be the index of the first success, i.e.,

$$X_1 = X_2 = \ldots = X_{Y-1} = 0 \quad \text{and} \quad X_Y = 1$$

Then $Y$ is a geometric random variable with parameter $p$.

Examples:
- Number of coin flips until first head
- Number of blind guesses on LSAT until I get one right
- Number of darts thrown until you hit a bullseye
- Number of random probes into hash table until empty slot
- Number of wild guesses at a password until you hit it

$$P(Y=k) = (1-p)^{k-1}p; \quad \text{Mean } 1/p; \quad \text{Variance } (1-p)/p^2$$
balls in urns – the hypergeometric distribution

B&T, exercise 1.61

Draw \(d\) balls (without replacement) from an urn containing \(N\), of which \(w\) are white, the rest black.

Let \(X\) = number of white balls drawn

\[
P(X = i) = \binom{w}{i} \binom{N-w}{d-i} \binom{N}{d}, \quad i = 0, 1, \ldots, d
\]

(note: \(n\) choose \(k\) = 0 if \(k < 0\) or \(k > n\))

\[E[X] = dp, \quad \text{where } p = w/N \text{ (the fraction of white balls)}\]

proof: Let \(X_j\) be 0/1 indicator for \(j\)-th ball is white, \(X = \sum X_j\)

The \(X_j\) are dependent, but \(E[X] = E[\sum X_j] = \sum E[X_j] = dp\)

\[\text{Var}[X] = dp(1-p)(1-(d-1)/(N-1))\]
\( N \approx 22500 \) human genes, many of unknown function

Suppose in some experiment, \( d = 1588 \) of them were observed (say, they were all switched on in response to some drug)

**A big question: What are they doing?**

One idea: The Gene Ontology Consortium ([www.geneontology.org](http://www.geneontology.org)) has grouped genes with known functions into categories such as “muscle development” or “immune system.” Suppose 26 of your \( d \) genes fall in the “muscle development” category.

- Just chance?
- Or call Coach & see if he wants to dope some athletes?

**Hypergeometric:** GO has 116 genes in the muscle development category. If those are the white balls among 22500 in an urn, what is the probability that you would see 26 of them in 1588 draws?
A differentially bound peak was associated to the closest gene (unique Entrez ID) measured by distance to TSS within CTCF flanking domains. OR: ratio of predicted to observed number of genes within a given GO category. Count: number of genes with differentially bound peaks. Size: total number of genes for a given functional group. Ont: the Geneontology. BP = biological process, MF = molecular function, CC = cellular component.

### Table 2. Gene Ontology Analysis on Differentially Bound Peaks in Myoblasts versus Myotubes

<table>
<thead>
<tr>
<th>GOID</th>
<th>Term</th>
<th>P Value</th>
<th>OR(^a)</th>
<th>Count(^b)</th>
<th>Size(^c)</th>
<th>Ont(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0005856</td>
<td>cytoskeleton</td>
<td>2.05E-11</td>
<td>2.40</td>
<td>94</td>
<td>490</td>
<td>CC</td>
</tr>
<tr>
<td>GO:0043292</td>
<td>contractile fiber</td>
<td>6.98E-09</td>
<td>5.85</td>
<td>22</td>
<td>58</td>
<td>CC</td>
</tr>
<tr>
<td>GO:0030016</td>
<td>myofibril</td>
<td>1.96E-08</td>
<td>5.74</td>
<td>21</td>
<td>56</td>
<td>CC</td>
</tr>
<tr>
<td>GO:0044449</td>
<td>contractile fiber part</td>
<td>2.58E-08</td>
<td>5.97</td>
<td>20</td>
<td>52</td>
<td>CC</td>
</tr>
<tr>
<td>GO:0030017</td>
<td>sarcomere</td>
<td>4.95E-08</td>
<td>6.04</td>
<td>19</td>
<td>49</td>
<td>CC</td>
</tr>
<tr>
<td>GO:0008092</td>
<td>skeletal muscle development</td>
<td>2.50E-06</td>
<td>4.13</td>
<td>20</td>
<td>65</td>
<td>BP</td>
</tr>
<tr>
<td>GO:0015629</td>
<td>actin cytoskeleton</td>
<td>4.73E-06</td>
<td>3.08</td>
<td>27</td>
<td>111</td>
<td>CC</td>
</tr>
<tr>
<td>GO:0003779</td>
<td>actin binding</td>
<td>3.01E-04</td>
<td>3.06</td>
<td>159</td>
<td></td>
<td>MF</td>
</tr>
<tr>
<td>GO:0006936</td>
<td>cytoskeleton part</td>
<td>3.50E-04</td>
<td>2.68</td>
<td>294</td>
<td></td>
<td>CC</td>
</tr>
<tr>
<td>GO:0031674</td>
<td>I band</td>
<td>2.27E-05</td>
<td>5.67</td>
<td>12</td>
<td>32</td>
<td>CC</td>
</tr>
<tr>
<td>GO:0003012</td>
<td>muscle system process</td>
<td>2.54E-05</td>
<td>4.11</td>
<td>16</td>
<td>52</td>
<td>BP</td>
</tr>
<tr>
<td>GO:0030029</td>
<td>actin filament-based process</td>
<td>2.89E-05</td>
<td>2.73</td>
<td>27</td>
<td>119</td>
<td>BP</td>
</tr>
<tr>
<td>GO:0007517</td>
<td>muscle development</td>
<td>5.06E-05</td>
<td>2.69</td>
<td>26</td>
<td>116</td>
<td>BP</td>
</tr>
</tbody>
</table>

The probability of seeing this many genes from a set of this size by chance according to the hypergeometric distribution. E.g., if you draw 1588 balls from an urn containing 490 white balls and \( \approx 22000 \) black balls, \( P(94 \text{ white}) \approx 2.05 \times 10^{-11} \).
Joint distributions

Often care about 2 (or more) random variables simultaneously measured $X$ = height and $Y$ = weight
$X$ = cholesterol and $Y$ = blood pressure
$X_1, X_2, X_3$ = work loads on servers A, B, C

Joint probability mass function:

$$f_{XY}(x, y) = P(X = x \& Y = y)$$

Joint cumulative distribution function:

$$F_{XY}(x, y) = P(X \leq x \& Y \leq y)$$
Two joint PMFs

<table>
<thead>
<tr>
<th>W</th>
<th>Z</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>2/24</td>
<td>2/24</td>
<td>2/24</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>2/24</td>
<td>2/24</td>
<td>2/24</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>2/24</td>
<td>2/24</td>
<td>2/24</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>2/24</td>
<td>2/24</td>
<td>2/24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>4/24</td>
<td>1/24</td>
<td>1/24</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0</td>
<td>3/24</td>
<td>3/24</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>0</td>
<td>4/24</td>
<td>2/24</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>4/24</td>
<td>0</td>
<td>2/24</td>
</tr>
</tbody>
</table>

\[
P(W = Z) = 3 \times \frac{2}{24} = \frac{6}{24}
\]

\[
P(X = Y) = \frac{(4 + 3 + 2)}{24} = \frac{9}{24}
\]

Can look at arbitrary relationships between variables this way
Two joint PMFs

<table>
<thead>
<tr>
<th>W</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>$f_w(w)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2/24</td>
<td>2/24</td>
<td>2/24</td>
<td>6/24</td>
</tr>
<tr>
<td>2</td>
<td>2/24</td>
<td>2/24</td>
<td>2/24</td>
<td>6/24</td>
</tr>
<tr>
<td>3</td>
<td>2/24</td>
<td>2/24</td>
<td>2/24</td>
<td>6/24</td>
</tr>
<tr>
<td>4</td>
<td>2/24</td>
<td>2/24</td>
<td>2/24</td>
<td>6/24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Z</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>$f_z(z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8/24</td>
<td>8/24</td>
<td>8/24</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8/24</td>
<td>8/24</td>
<td>8/24</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>8/24</td>
<td>8/24</td>
<td>8/24</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Y</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>$f_y(y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8/24</td>
<td>8/24</td>
<td>8/24</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8/24</td>
<td>8/24</td>
<td>8/24</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>8/24</td>
<td>8/24</td>
<td>8/24</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>$f_x(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4/24</td>
<td>1/24</td>
<td>1/24</td>
<td>6/24</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>3/24</td>
<td>3/24</td>
<td>6/24</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>4/24</td>
<td>2/24</td>
<td>6/24</td>
</tr>
<tr>
<td>4</td>
<td>4/24</td>
<td>0</td>
<td>2/24</td>
<td>6/24</td>
</tr>
</tbody>
</table>

**Marginal distribution of one r.v.:**

$$f_Y(y) = \Sigma_x f_{XY}(x,y)$$

$$f_X(x) = \Sigma_y f_{XY}(x,y)$$

**Question:** Are $W$ & $Z$ independent? Are $X$ & $Y$ independent?
sampling from a (continuous) joint distribution

var(x)=1, var(y)=1, cov=0, n=1000

var(x)=1, var(y)=3, cov=0, n=1000

var(x)=1, var(y)=3, cov=0, n=100

var(x)=1, var(y)=3, cov=0.8, n=1000

var(x)=1, var(y)=3, cov=1.5, n=1000

var(x)=1, var(y)=3, cov=1.7, n=1000
A function $g(X,Y)$ defines a new random variable.

Its expectation is:

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

Expectation is linear. I.e., if $g$ is linear:

$$E[g(X, Y)] = E[aX + bY + c] = aE[X] + bE[Y] + c$$

Example:

$$g(X,Y) = 2X - Y$$

$$E[g(X,Y)] = \frac{72}{24} = 3$$

$$E[g(X,Y)] = 2 \cdot 2.5 - 2 = 3$$

<table>
<thead>
<tr>
<th>$X$</th>
<th>$Y$</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1\cdot4/24</td>
<td>0\cdot1/24</td>
<td>-1\cdot1/24</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3\cdot0/24</td>
<td>2\cdot3/24</td>
<td>1\cdot3/24</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>5\cdot0/24</td>
<td>4\cdot4/24</td>
<td>3\cdot2/24</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>7\cdot4/24</td>
<td>6\cdot0/24</td>
<td>5\cdot2/24</td>
</tr>
</tbody>
</table>
random variables – summary

**RV:** a numeric function of the outcome of an experiment

**Probability Mass Function** $p(x)$: prob that $RV = x; \sum p(x)=1$

**Cumulative Distribution Function** $F(x)$: probability that $RV \leq x$

Concepts generalize to **joint** distributions

**Expectation:**

- of a random variable: $E[X] = \sum x \cdot p(x)$
- of a function: if $Y = g(X)$, then $E[Y] = \sum g(x) \cdot p(x)$

**linearity:**

$E[aX + b] = aE[X] + b$

$E[X+Y] = E[X] + E[Y]$; even if dependent

this interchange of “order of operations” is quite special to linear combinations. E.g. $E[XY] \neq E[X] \cdot E[Y]$, in general (but see below)
random variables – summary

Variance:

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

Standard deviation: \( \sigma = \sqrt{\text{Var}[X]} \)

\[ \text{Var}[aX+b] = a^2 \text{Var}[X] \]

If \( X \) & \( Y \) are independent, then

\[ E[X \cdot Y] = E[X] \cdot E[Y]; \]

\[ \text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] \]

(These two equalities hold for indp rv's; but not in general.)
random variables – summary

Important Examples:

Bernoulli: \( P(X=1) = p \) and \( P(X=0) = 1-p \) \( \mu = p, \quad \sigma^2 = p(1-p) \)

Binomial: \( P(X = i) = \binom{n}{i} p^i (1-p)^{n-i} \) \( \mu = np, \quad \sigma^2 = np(1-p) \)

Poisson: \( P(X = i) = e^{-\lambda} \frac{\lambda^i}{i!} \) \( \mu = \lambda, \quad \sigma^2 = \lambda \)

Bin(n,p) \( \approx \) Poi(\( \lambda \)) where \( \lambda = np \) fixed, \( n \to \infty \) (and so \( p=\lambda/n \to 0 \))

Geometric \( P(X=k) = (1-p)^{k-1}p \) \( \mu = 1/p, \quad \sigma^2 = (1-p)/p^2 \)

Many others, e.g., hypergeometric