Random variable $X$ and event $E$ are independent if
\[
\forall x \ P(\{X = x\} \land E) = P(\{X=x\}) \cdot P(E)
\]
Ex 1: Roll a fair die to obtain a random number $1 \leq X \leq 6$, then flip a fair coin $X$ times. Let $E$ be the event that the number of heads is even. $P(\{X=x\}) = 1/6$ for any $1 \leq x \leq 6$. $P(E) = 1/2$. $P(\{X=x\} \land E) = 1/12$, so they are independent.

Ex 2: as above, and let $F$ be the event that the total number of heads $= 6$. $P(F) = 2^{4/6} > 0$, and considering, say, $X=4$, we have $P(X=4) = 1/6 > 0$ (as above), but $P(\{X=4\} \land F) = 0$, since you can't see 6 heads in 4 flips. So $X$ & $F$ are dependent. (Knowing that $X$ is small renders $F$ impossible; knowing that $F$ happened means $X$ must be 6.)
r.v.s and independence

**Defn:** Random variable $X$ and event $E$ are independent if the event $E$ is independent of the event $\{X=x\}$ (for any fixed $x$), i.e.
\[
\forall x \ P(\{X = x\} \& E) = P(\{X=x\}) \cdot P(E)
\]

**Defn:** Two random variables $X$ and $Y$ are independent if the events $\{X=x\}$ and $\{Y=y\}$ are independent (for any fixed $x, y$), i.e.
\[
\forall x, y \ P(\{X = x\} \& \{Y=y\}) = P(\{X=x\}) \cdot P(\{Y=y\})
\]

Intuition as before: knowing $X$ doesn't help you guess $Y$ or $E$ and vice versa.

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 $E[X^2] \neq E[X]^2$

properties of variance

In general:
\[
\text{Var}[X+Y] \neq \text{Var}[X] + \text{Var}[Y]
\]

**Not linear**

**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 $\text{Var}[X+X] = 2\text{Var}[X]$?

variance of independent r.v.s is additive

**Theorem:** If $X$ & $Y$ are independent, then $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$  
**Alternate Proof:**
\[
\text{Var}[X + Y]
\]
\[
= E[(X + Y)^2] - (E[X + Y])^2
\]
\[
= E[X^2 + 2XY + Y^2] - (E[X] + E[Y])^2
\]
\[
\]
\[
= E[X^2] - (E[X])^2 + E[Y^2] - (E[Y])^2 + 2(E[XY] - E[X]E[Y])
\]
\[
= \text{Var}[X] + \text{Var}[Y] + 2(E[XY] - E[X]E[Y])
\]
\[
= \text{Var}[X] + \text{Var}[Y]
\]
Difference between independent and dependent

Dependent r.v.s can reinforce/cancel/correlate in arbitrary ways.
Independent r.v.s are …… independent.

Example:
\[ Z = X_1 + X_2 + \ldots + X_n \]
\[ X_i \text{ is indicator r.v. with probability } 1/2 \text{ of being 1.} \]

versus
\[ W = n X_i \]

One more linearity of expectation practice problem

Given a DNA sequence of length \( n \)
e.g. AAATGAATGAATCC…….
where each position is
A with probability \( p_A \)
T with probability \( p_T \)
G with probability \( p_G \)
C with probability \( p_C \).

What is the expected number of occurrences of the substring AATGAAT?

\[ \text{AAATGAATGAATCC} \quad \text{AAATGAATGAATCC} \]

discrete uniform random variables

A discrete random variable \( X \) equally likely to take any (integer) value between integers \( a \) and \( b \), inclusive, is uniform.

Notation: \( X \sim \text{Unif}(a,b) \)

Probability: \( P(X = i) = \frac{1}{b - a + 1} \)

Mean, Variance: \( E[X] = \frac{a + b}{2}, \ Var[X] = \frac{(b - a)(b - a + 2)}{12} \)

Example: value shown on one roll of a fair die is Unif(1,6):
\( P(X=i) = 1/6 \)
\( E[X] = 7/2 \)
\( \text{Var}[X] = 35/12 \)
Bernoulli random variables

An experiment results in “Success” or “Failure”

$X$ is an indicator random variable ($1 = $ success, $0 = $ failure)

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

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

$E[X] = p$

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

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

---

Binomial random variables

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

$E[X] = np$

$\text{Var}(X) = np(1-p)$
mean, variance of the binomial (II)

If $Y_1, Y_2, \ldots, Y_n \sim \text{Ber}(p)$ and independent, then $X = \sum_{i=1}^{n} Y_i \sim \text{Bin}(n, p)$.

$$E[X] = np$$

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

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

$$\text{Var}[X] = \text{Var} \left[ \sum_{i=1}^{n} Y_i \right] = \sum_{i=1}^{n} \text{Var}[Y_i] = n\text{Var}[Y_1] = np(1-p)$$

models & reality

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 ($\text{length } n \approx 10^4$)

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

# bit errors in file written to disk

# of typos in a book

# of elements in particular bucket of large hash table

# of server crashes per day in giant data center

# 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.

\[\begin{array}{ll}
\lambda = 0.5 \\
\lambda = 3
\end{array}\]

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

\(X\) 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:

\[
e^\lambda = \frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \cdots = \sum_{i=0}^{\infty} \frac{\lambda^i}{i!}
\]

So

\[
\sum_{0 \leq i} P(X = i) = \sum_{0 \leq i} e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} \sum_{0 \leq i} \frac{\lambda^i}{i!} = e^{-\lambda} e^\lambda = 1
\]

\[\begin{array}{ll}
E[X] &= \sum_{0 \leq i} i \cdot e^{-\lambda} \frac{\lambda^i}{i!} \\
&= \sum_{1 \leq i} i \cdot e^{-\lambda} \frac{\lambda^i}{i!} \\
&= \lambda e^{-\lambda} \sum_{1 \leq i} \frac{\lambda^{i-1}}{(i-1)!} \\
&= \lambda e^{-\lambda} \sum_{0 \leq j} \frac{\lambda^j}{j!}
\end{array}\]

\[\begin{array}{ll}
&= \lambda e^{-\lambda} e^\lambda \\
&= \lambda
\end{array}\]

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

\(\text{Var}[X] = \lambda\), 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,” e.g.
- \( n > 20 \) 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{Bin}(n, p) \Rightarrow \text{Pois} which in the limit is \( \lambda \) as \( n \to \infty \) (equivalently, \( p \to 0 \)) while holding \( n \).

Handy: Poisson has only 1 parameter—the expected # of successes

\[
X \sim \text{Bin}(n, p)
\]

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

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

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

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

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

I.e., Binomial \( \approx \) Poisson for large \( n \), small \( p \), moderate \( i, \lambda \).

**sending data on a network**

Consider 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?

\[
P(X = 0) = e^{-\lambda} \frac{\lambda^0}{0!} = e^{-0.01} \frac{0.01^0}{0!} \approx 0.990049829
\]

Using \( Y \sim \text{Bin}(10^4, 10^{-6}) \):

\[
P(Y=0) \approx 0.990049829
\]

I.e., Poisson approximation (here) is accurate to \( \approx 5 \) parts per billion

**binomial vs poisson**

- Binomial(10, 0.3)
- Binomial(100, 0.03)
- Poisson(3)

![Graph](image-url)
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.

geometric distribution

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 SAT 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

\[
P(Y=k) = (1-p)^{k-1}p; \quad \text{Mean } 1/p; \quad \text{Variance } (1-p)/p^2
\]

interlude: more on conditioning

Recall: conditional probability
\[
P(X=x \mid A) = \frac{P(X=x \text{ \& } A)}{P(A)}
\]
Conditional probability is a probability, i.e.
1. it’s nonnegative
2. it’s normalized
3. it’s happy with the axioms, etc.

Define: The conditional expectation of \( X \)
\[
E[X \mid A] = \sum_x x \cdot p(X = x \mid A)
\]
I.e., the value of r.v. \( X \) averaged over outcomes where I know event \( A \) happened

total expectation

Recall: the law of total probability
\[
p(X=x) = p(X=x \mid A) \cdot P(A) + p(X=x \mid A^c) \cdot P(A^c)
\]
I.e., unconditional probability is the weighted average of conditional probabilities, weighted by the probabilities of the conditioning events

The Law of Total Expectation
\[
E[X] = E[X \mid A] \cdot P(A) + E[X \mid A^c] \cdot P(A^c)
\]
I.e., unconditional expectation is the weighted average of conditional expectations, weighted by the probabilities of the conditioning events
Proof of the Law of Total Expectation:

\[ E[X] = \sum_x x p(x) \]
\[ = \sum_x x (p(x \mid A) P(A) + p(x \mid \overline{A}) P(\overline{A})) \]
\[ = \sum_x x p(x \mid A) P(A) + \sum_x x p(x \mid \overline{A}) P(\overline{A}) \]
\[ = \left( \sum_x x p(x \mid A) \right) P(A) + \left( \sum_x x p(x \mid \overline{A}) \right) P(\overline{A}) \]
\[ = E[X \mid A] P(A) + E[X \mid \overline{A}] P(\overline{A}) \]

\[ X \sim \text{Geo}(p) \]

\[ E[X] = E[X \mid X=1] \cdot P(X=1) + E[X \mid X>1] \cdot P(X>1) \]

\[ E[X] = \frac{1}{p} \]

E.g., if \( p = 1/2 \), expect to wait 2 flips for 1st head; \( p = 1/10 \), expect to wait 10 flips.

(Similar derivation for variance: \((1-p)/p^2\))

memorylessness: after flipping one tail, remaining waiting time until 1st head is exactly the same as starting from scratch

balls in urns – the hypergeometric distribution

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: \( \binom{n}{k} = 0 \) if \( k < 0 \) or \( k > n \)]

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

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

the \( X_i \) are dependent, but \( E[X] = \sum E[X_i] = \sum E[X] = dp \)

\[ \text{Var}[X] = dp(1-p)(1-(d-1)/(N-1)) \]
data mining

N ≈ 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 (and 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?

---

joint distributions

Often, several random variables are simultaneously observed

\( X = \text{height} \) and \( Y = \text{weight} \)

\( X = \text{cholesterol} \) and \( Y = \text{blood pressure} \)

\( X_1, X_2, X_3 = \text{work loads on servers} A, B, C \)

*Joint* probability mass function:

\[
f_{XY}(x, y) = P(\{X = x\} \cap \{Y = y\})
\]

*Joint* cumulative distribution function:

\[
F_{XY}(x, y) = P(\{X \leq x\} \cap \{Y \leq y\})
\]

---

examples

Two joint PMFs

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

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

\[
P(W = Z) = 3 \times 2/24 = 6/24
\]

\[
P(X = Y) = (4 + 3 + 2)/24 = 9/24
\]

Can look at arbitrary relationships among variables this way
sampling from a joint distribution

- Flip n fair coins
  - X = #Heads seen in first n/2+k
  - Y = #Heads seen in last n/2+k

another example

joint, marginals and independence

Repeating the Definition: Two random variables X and Y are independent if the events \{X=x\} and \{Y=y\} are independent (for any fixed x, y), i.e.

$$\forall x, y \ P(\{X=x \} \& \{Y=y\}) = P(\{X=x\}) \cdot P(\{Y=y\})$$

Equivalent Definition: Two random variables X and Y are independent if their joint probability mass function is the product of their marginal distributions, i.e.

$$\forall x, y \ f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$$

Exercise: Show that this is also true of their cumulative distribution functions
expectation of a function of 2 r.v.'s

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. E.g., if \( g \) is linear:
\[
E[g(X, Y)] = E[a X + b Y + c] = a E[X] + b E[Y] + c
\]

Example:
\[
g(X, Y) = 2X - Y
\]
\[
E[g(X, Y)] = 72/24 = 3
\]
\[
E[g(X, Y)] = 2 \cdot E[X] - E[Y] = 2 \cdot 2.5 - 2 = 3
\]

recall both marginals are uniform
Important Examples:

Uniform(a,b): \( P(X = i) = \frac{1}{b - a + 1} \)
\[ \mu = \frac{a + b}{2}, \quad \sigma^2 = \frac{(b - a)(b - a + 2)}{12} \]

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

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) \( \approx \text{Poi}(\lambda) \) 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, negative binomial, …