Conditional probability

1 Conditional probability

Conditional probability and chain rule

Bayes' rule and total probability law

Conditional probability $P(\cdot|B)$ is also a probability measure

Example - Double Ace

Example - Birthday problem

Example - Monty Hall problem

Example - False positive

2 Joint, marginal, and conditional probabilities

Joint, marginal, conditional probabilities

How to get joint, marginal, conditional from other two

3 Independent, pairwise independent, conditionally independent events

Independent, pairwise independent, conditionally independent events

Dependent events

Example - Pairwise independent but not independent events

4 Gambler's ruin

Gambler's ruin

Simulation of Gambler's ruin sample path

Monte Carlo for Gambler's ruin

Gambler's ruin - First step analysis

`sparse`, `spdiags`, `spdiags_Lee`, `kron`

Example - `spdiags`, `spdiags_Lee`

`A\b` for Gambler's ruin

Gambler's ruin - Linear recurrence relation

Gambler's ruin - $q > 1/2$ case

Gambler's ruin - $q = 1/2$ case

5 Simpson's paradox

Simpson's paradox

6 How to compute the intersection and union probabilities

How to compute $P(\cap_{i=1}^{n} A_i)$

How to compute $P(\cup_{i=1}^{n} A_i)$

Conditional probability and chain rule

Conditional probability - Definition

$$P(B|A) = \frac{P(AB)}{P(A)}$$

Chain rule

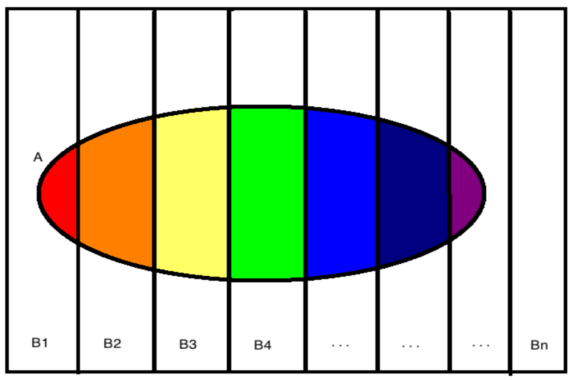
$$\begin{aligned}
P(AB) &= P(A)P(B|A) \\
P(ABC) &= P(A)P(B|A)P(C|AB) \\
P(ABCD) &= P(A)P(B|A)P(C|AB)P(D|ABC) \\
\vdots &= \vdots
\end{aligned}$$

Bayes' rule and total probability law

Bayes' rule

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

Total probability law - Divide and conquer



If Ω can be divided into several disjoint events B_k so that $\Omega = \cup_{k=1}^n B_k$ disjointly or if A can be divided into several disjoint events AB_k so that $A = \cup_{k=1}^n AB_k$ disjointly,

[**Step 1**] (Divide) Divide A into n disjoint AB_k.

[**Step 2**] (Conquer) Compute the probability $P(AB_k)$ using chain rule.

Then, the probability $P(A)$ can be computed as

$$P(A) = \sum_{k=1}^{n} P(AB_k) = \sum_{k=1}^{n} P(B_k)P(A|B_k)$$

Bayes' rule + Total probability law

If $\Omega = \cup_{k=1}^n B_k$ disjointly, then

$$P(B_1|A) \overset{\text{Bayes}}{=} \frac{P(A|B_1)P(B_1)}{P(A)} \overset{\text{TPL}}{=} \frac{P(B_1)P(A|B_1)}{\sum_{k=1}^n P(B_k)P(A|B_k)}$$

Conditional probability $P(\cdot|B)$ is also a probability measure

Conditional probability $P(\cdot|B)$ is also a probability measure and hence it satisfies all the equalities and inequalities that the usual probability measure $P(\cdot)$ satisfies. You just need to add $|B$ at the end of the equalities and inequalities before the right end parenthesis.

(1) $$P(\Omega|B) = 1, \ P(\emptyset|B) = 0$$

(2) $$0 \leq P(A|B) \leq 1 \qquad \text{for any event } A$$

(3) $$P(\cup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} P(A_i|B) \text{ for disjoint } A_i$$

(4) $$P(\cup_{i=1}^{n} A_i|B) = \sum_{i=1}^{n} P(A_i|B) \text{ for disjoint } A_i$$

(5) $$P(A_1|B) \leq P(A_2|B) \qquad \text{for } A_1 \subset A_2$$

(6) $$P(A^c|B) = 1 - P(A|B)$$

(7) $$P(\cup_{i=1}^{n} A_i|B) \leq \sum_{i=1}^{n} P(A_i|B)$$

(7) $$P(\cup_{i=1}^{n} A_i|B) \geq \sum_{i=1}^{n} P(A_i|B) - \sum_{1 \leq i < j \leq n} P(A_iA_j|B)$$

(7) $$P(\cup_{i=1}^{n} A_i|B) \leq \sum_{i=1}^{n} P(A_i|B) - \sum_{1 \leq i < j \leq n} P(A_iA_j|B) + \sum_{1 \leq i < j < k \leq n} P(A_iA_jA_k|B)$$

$\cdots$

(7) $$P(\cup_{i=1}^{n} A_i|B) = \sum_{i=1}^{n} P(A_i|B) - \cdots + (-1)^{n+1} P(A_1A_2 \cdots A_n|B)$$

Example - Double ace

We choose two cards from the ordinary 52 cards deck.

A	An ace is chosen
A_1	Spade ace is chosen
B	Both cards are aces

Calculate $P(B|A)$ and $P(B|A_1)$.

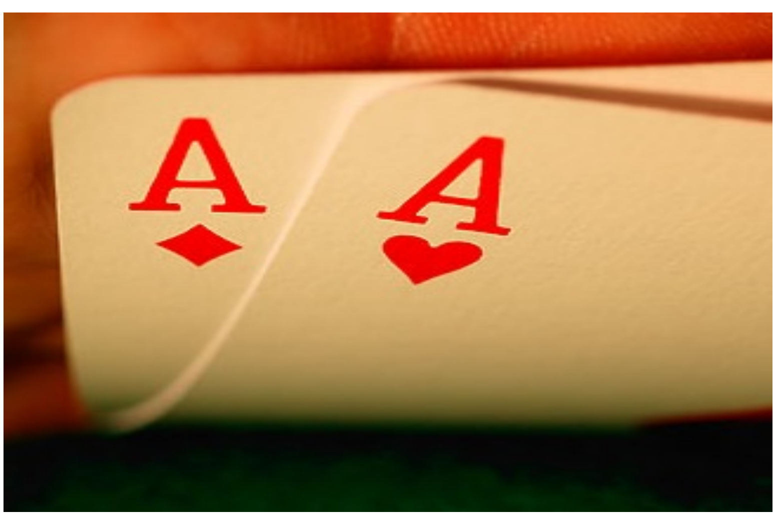
$\boxed{P(B|A_1)}$

Suppose that the spade ace is chosen. Then, there are 51 cards left and among these 51 cards we will chose one cards. To have both aces we have to choose diamond, heart, or club ace. So,

$$P(B|A_1) = \frac{3}{51} = 0.0588$$

$\boxed{P(B|A)}$

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{P(B)}{P(A)} = \frac{P(B)}{1 - P(A^c)} = \frac{\frac{\binom{4}{2}}{\binom{52}{2}}}{1 - \frac{\binom{48}{2}}{\binom{52}{2}}} = 0.0303$$



Example - Birthday problem

Let $p(n)$ be the probability that there is at least one birthday match among n people. Find the minimum number n of people that $p(n)$ is over 50 %.

Approximate computing

B_k First k people have all different birthdays

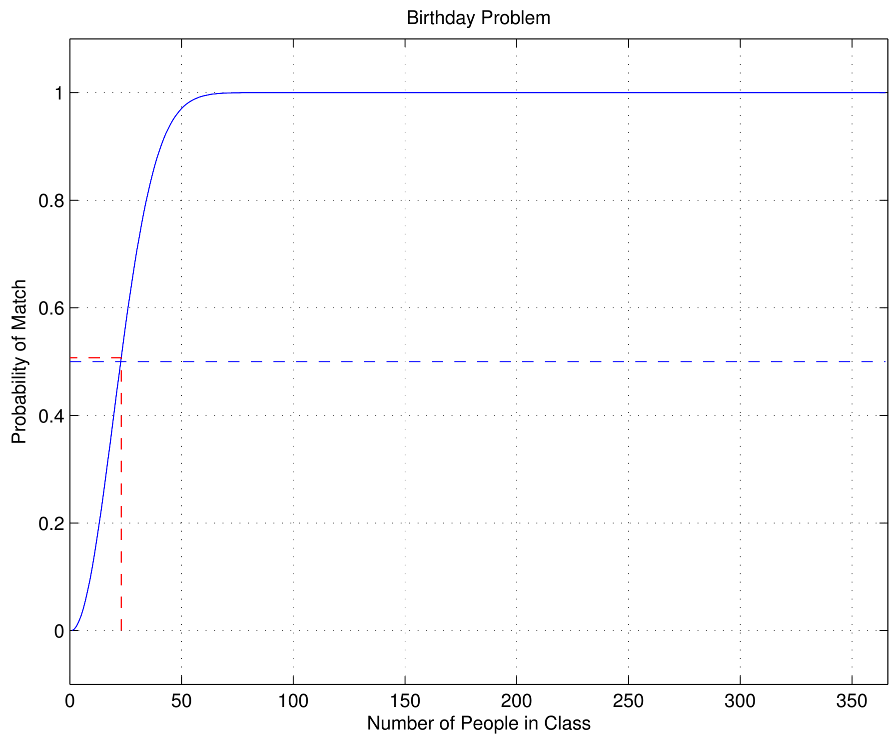
Then, by the chain rule we have

$$\begin{aligned}
1 - p(n) &= P(B_1 B_2 B_3 \cdots B_n) \\
&= P(B_1)P(B_2|B_1)P(B_3|B_1B_2)\cdots P(B_n|B_1B_2\cdots B_{n-1}) \\
&= P(B_1)P(B_2|B_1)P(B_3|B_2)\cdots P(B_n|B_{n-1}) \\
&= (1)\left(\frac{364}{365}\right)\left(\frac{363}{365}\right)\cdots\left(\frac{(365-(n-1))}{365}\right) \\
&= (1)\left(1-\frac{1}{365}\right)\left(1-\frac{2}{365}\right)\cdots\left(1-\frac{n-1}{365}\right) \\
&\approx e^{-\frac{1}{365}}e^{-\frac{2}{365}}\cdots e^{-\frac{n-1}{365}} = e^{-\frac{n(n-1)}{2\times 365}}
\end{aligned}$$

$$e^{-\frac{n(n-1)}{2\times 365}} = 0.5 \quad \Rightarrow \quad n \approx 23$$

Exact computing

$p(22)$	$p(23)$	$p(24)$
0.4757	0.5073	0.5383




```
clear all; close all; clc; rng('default')

% q(n) = P(there is no birthday match among n people)
q=ones(1,365);
for n=2:365
    q(n)=q(n-1)*(1-(n-1)/365);
end

% p(n) = P(there is at least one birthday match among n people)
p=ones(1,365)-q;

% Find the least number n of people with p(n) >= 0.5
MinPeople=find(p>=0.5,1);

plot(1:365,p); grid on; hold on
plot([MinPeople MinPeople 0],[0 p(MinPeople) p(MinPeople)],'--r')
plot([0 365],[0.5 0.5],'--b')
axis([0 366 -0.1 1.1])
title('Birthday Problem')
xlabel('Number of People in Class'); ylabel('Probability of Match')
```

Example - Monty Hall problem

You are on a game show, and you're given the choice of three doors: Behind one door is a car and behind the others are goats. You pick a door, say #1, and the host, who knows what's behind the doors, opens another door, say #3, which has a goat. He then says to you,"Do you want to pick door #2?" Is it to your advantage to switch your choice?

Under the no change strategy the winning probability is 1/3

Under the change strategy the winning probability is 2/3

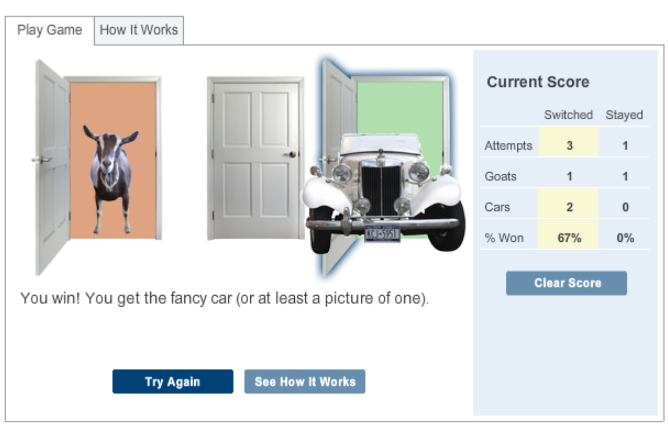
C	Car door is chosen at the first round
G	Goat door is chosen at the first round
W	Win the prize

[Step 1] (Divide) Using the first round choice divide W into WC and WG.

$$W = WC \cup WG \quad \text{disjointly}$$

[Step 2] (Conquer) Compute $P(WC)$ and $P(WG)$ using the chain rule.

$$\begin{aligned} P(WC) &= P(C)P(W|C) = \frac{1}{3} \times 0 = 0 \\ P(WG) &= P(G)P(W|G) = \frac{2}{3} \times 1 = \frac{2}{3} \end{aligned}$$



Kenneth Chang, Sarah Graham, Viktor Koen, Michael Lindsay/The New York Times

Example - False positive

A laboratory blood test is 95% effective in detecting a certain disease when it is, in fact, present. However, the test also yields a "false positive" result for 1% of the healthy persons tested. If 0.01% of the population actually has the disease, what is the probability that a person has the disease given that the test result is positive?

Event

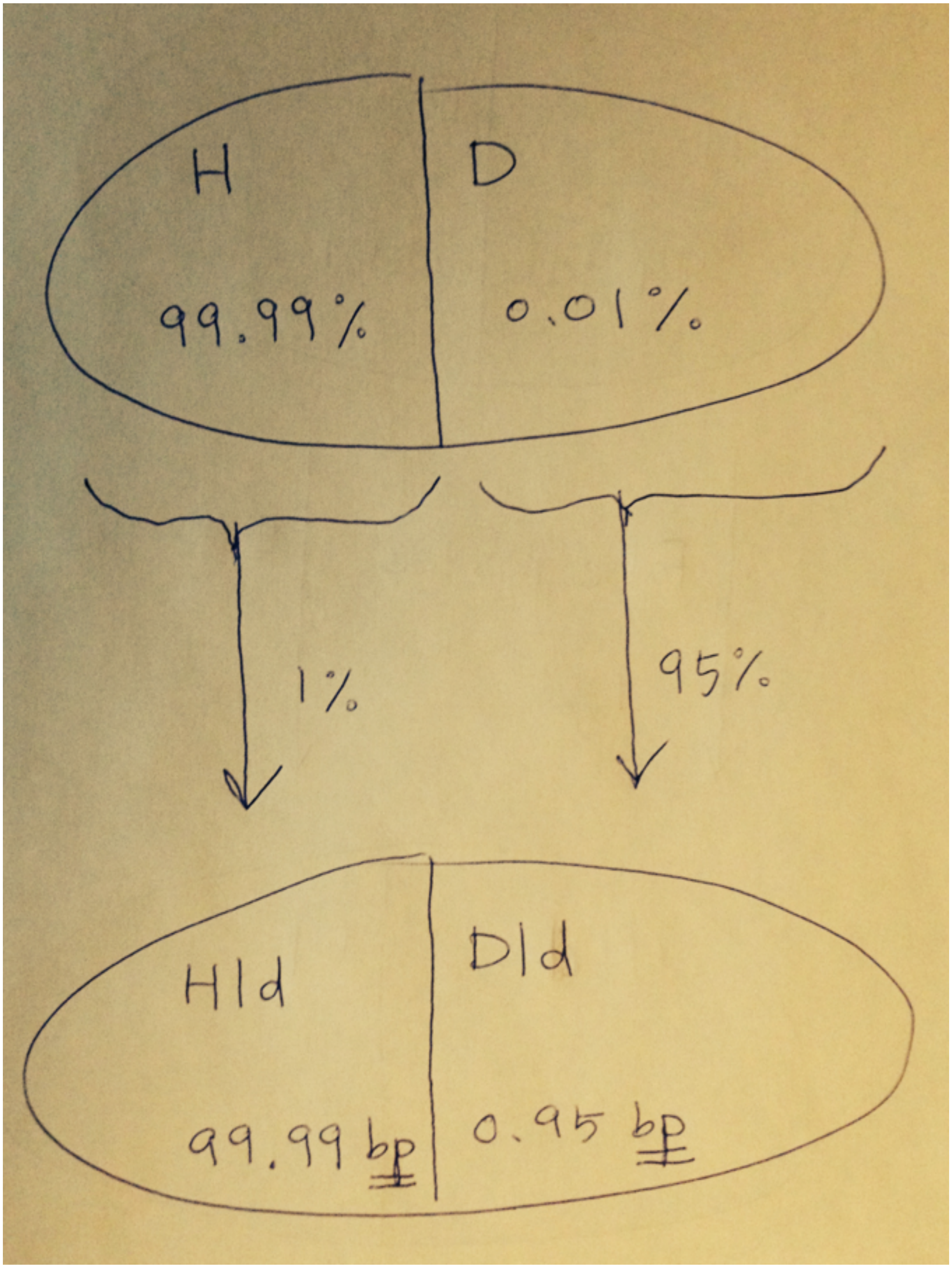
H	A person is **H**ealthy
h	Blood test reports that a person is **h**ealthy
D	A person has the **D**isease
d	Blood test reports that a person has **d**isease

Info

$P(d\|D) = 0.95$	Test is 95 % effective in detecting the disease when it is present
$P(h\|D) = 0.05$	
$P(d\|H) = 0.01$	Test also yields a "false positive" result with 1 % chance
$P(h\|H) = 0.99$	
$P(D) = 0.0001$	0.01 % of the population actually has the disease
$P(H) = 0.9999$	

$P(D|d)$

$$\begin{aligned} P(D|d) &= \frac{P(D)P(d|D)}{P(d)} \\ &= \frac{P(D)P(d|D)}{P(D)P(d|D) + P(H)P(d|H)} \\ &= \frac{(0.0001)(0.95)}{(0.0001)(0.95) + (0.9999)(0.01)} = 0.0094 \end{aligned}$$


H
D
99.99%
0.01%
1%
95%
H|d
D|d
99.99 bp
0.95 bp

Joint, marginal, conditional probabilities

Suppose we decompose the sample space Ω two different ways;

$$\Omega = \cup_{i=1}^{m} A_i \quad \text{disjointly} \quad \text{and} \quad \Omega = \cup_{j=1}^{n} B_j \quad \text{disjointly}$$

Joint $P(A_iB_j)$ and marginal $P(A_i)$, $P(B_j)$

$P(B_4)$	B_4	$P(A_1B_4)$	$P(A_2B_4)$	$P(A_3B_4)$
$P(B_3)$	B_3	$P(A_1B_3)$	$P(A_2B_3)$	$P(A_3B_3)$
$P(B_2)$	B_2	$P(A_1B_2)$	$P(A_2B_2)$	$P(A_3B_2)$
$P(B_1)$	B_1	$P(A_1B_1)$	$P(A_2B_1)$	$P(A_3B_1)$
		A_1	A_2	A_3
		$P(A_1)$	$P(A_2)$	$P(A_3)$

Conditional $P(B_j \mid A_i)$

B_4		$P(B_4 \mid A_2)$	
B_3		$P(B_3 \mid A_2)$	
B_2		$P(B_2 \mid A_2)$	
B_1		$P(B_1 \mid A_2)$	
	A_1	A_2	A_3
		$\Uparrow$	

How to get joint, marginal, conditional from other two

$$\begin{array}{ll} \text{Chain rule} & P(A_iB_j) = P(A_i)P(B_j|A_i) \\ \text{Marginalization} & P(A_i) = \sum_j P(A_iB_j) \\ \text{Conditioning} & P(B_j|A_i) = \dfrac{P(A_iB_j)}{P(A_i)} \end{array}$$

Independent, pairwise independent, conditionally independent events

$\boxed{\text{Independent events}}$

A and B are independent if

$$P(AB) = P(A)P(B)$$

$A_1, \cdots, A_n$ are independent if for any sub collection $A_{i_1}, \cdots, A_{i_m}$ of $A_1, \cdots, A_n$

$$P(A_{i_1}A_{i_2}\cdots A_{i_m}) = P(A_{i_1})P(A_{i_2})P\cdots P(A_{i_m})$$

$\boxed{\text{Pairwise independent events}}$

$A_1, \cdots, A_n$ are pairwise independent if for any pair A_i, A_j from $A_1, \cdots, A_n$

$$P(A_iA_j) = P(A_i)P(A_j)$$

$\boxed{\text{Conditionally independent events}}$

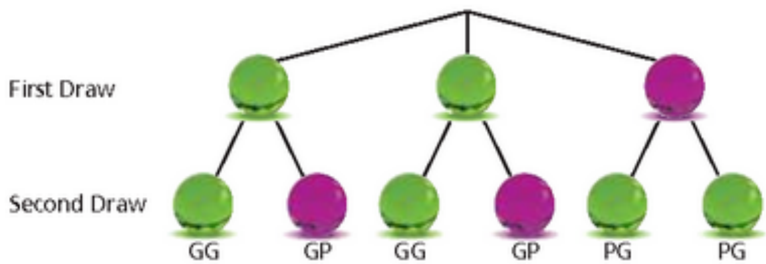
$A_1, \cdots, A_n$ are conditionally independent conditioned on B if for any sub collection $A_{i_1}, \cdots, A_{i_m}$ of $A_1, \cdots, A_n$

$$P(A_{i_1}A_{i_2}\cdots A_{i_m}|B) = P(A_{i_1}|B)P(A_{i_2}|B)P\cdots P(A_{i_m}|B)$$

1 ACTIVITY: Dependent Events

Work with a partner. You have three marbles in a bag. There are two green marbles and one purple marble. You randomly draw two marbles from the bag.

a. Use the tree diagram to find the probability that both marbles are green.

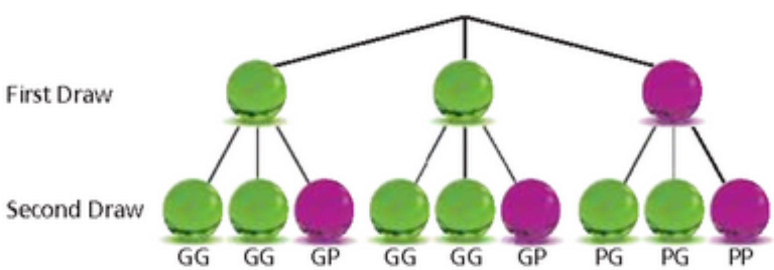


b. In the tree diagram, does the probability of getting a green marble on the second draw *depend* on the color of the first marble? Explain.

2 ACTIVITY: Independent Events

Work with a partner. Using the same marbles from Activity 1, randomly draw a marble from the bag. Then put the marble back in the bag and draw a second marble.

a. Use the tree diagram to find the probability that both marbles are green.



b. In the tree diagram, does the probability of getting a green marble on the second draw *depend* on the color of the first marble? Explain.

Example - Pairwise independent but not independent events

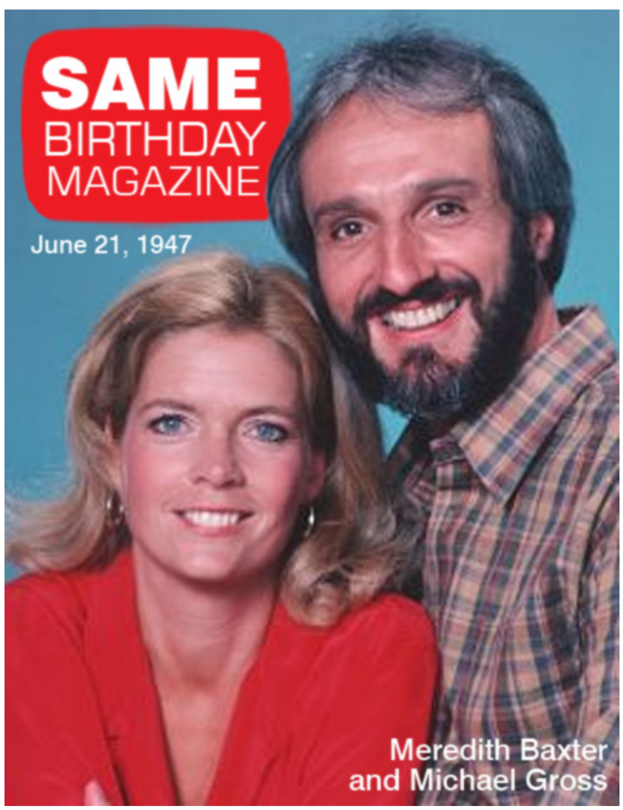
There are n people in the class. Suppose each one choose one's birthday independently and uniformly over the 365 days. For each pair i and j we let A_{ij} be the event that i and j share the common birthday. Show that A_{ij} are not independent but they are pairwise independent.

A_{ij} are not independent

$$P(A_{23}|A_{12}, A_{13}) = 1 \neq P(A_{23}) = \frac{1}{365}$$

A_{ij} are pair-wise independent

$$P(A_{13}|A_{12}) = \frac{P(A_{12}A_{13})}{P(A_{12})} = \frac{\frac{1}{365^2}}{\frac{1}{365}} = \frac{1}{365} = P(A_{13})$$



Gambler's ruin

Suppose you have $\$i$. Each time you are betting \$1 on some gambling that we will win \$1 with probability $p(\leq 1/2)$ and lose \$1 with probability $q := 1 - p$. If you lose all the money, you ruin! If you have $\$N$ in your pocket, you happily quit this game. Let $Q(i)$ be the ruin probability when you start with initial capital $\$i$:

R	Ruin
I	Initial capital
$Q(i) = P(R \mid I = i)$	Ruin probability when you start with initial capital $\$i$

Calculate $Q(i)$.

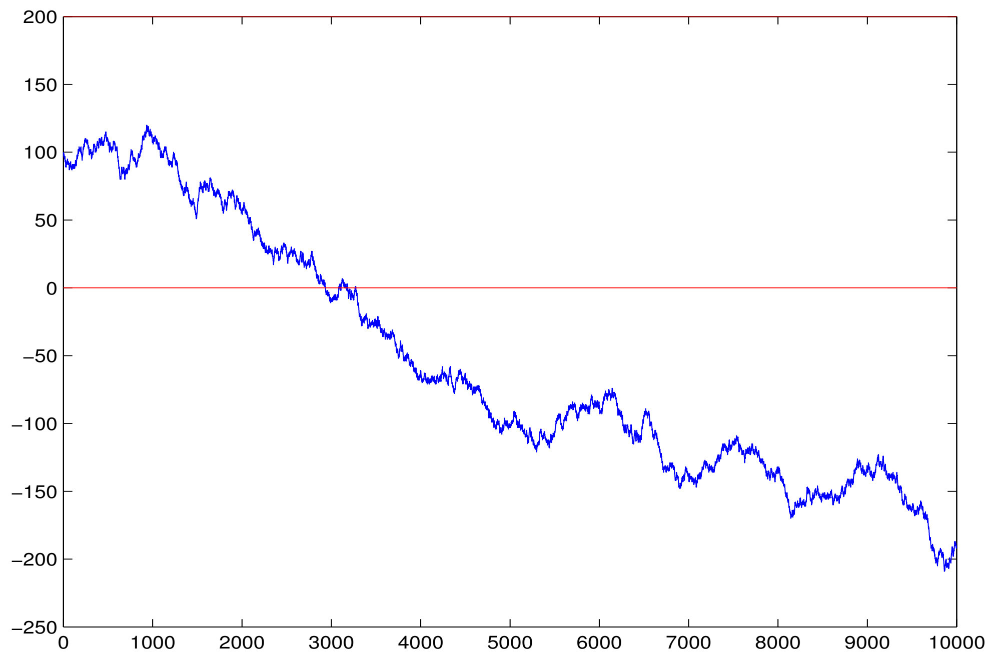


Figure 1: Gambler's ruin sample path with initial capital 100, goal 200, success probability 0.49.

```
clear all; close all; clc; rng('default')

% Parameters
p=0.49;     % Probability of head
IC=100;     % Initial capitial
Goal=200;   % Goal
n=10000;    % Maximum number of steps taken for each simulation path
NumSimu=1; % Number of simulations

% Gambler's ruin path
x=random('Binomial',1*ones(NumSimu,n),p*ones(NumSimu,n));
x=2*x-1;
x=cumsum(x,2);
x=[zeros(NumSimu,1) x];
x=IC+x;
x(:,n+2)=0;     % Make the hitting time of 0 finite
x(:,n+3)=Goal; % Make the hitting time of Goal finite

plot(0:n,x(1:n+1),'-'); hold on
plot([0 n],[0 0],'-r',[0 n],[Goal Goal],'-r')
```

```
clear all; close all; clc; rng('default')

% Parameters
p=0.49;     % Probability of head
IC=100;     % Initial capitial
Goal=200;   % Goal
n=10000;    % Maximum number of steps taken for each simulation path
NumSimu=100; % Number of simulations

% Gambler's ruin path
x=random('Binomial',1*ones(NumSimu,n),p*ones(NumSimu,n));
x=2*x-1;
x=cumsum(x,2);
x=[zeros(NumSimu,1) x];
x=IC+x;
x(:,n+2)=0;     % Make the hitting time of 0 finite
x(:,n+3)=Goal; % Make the hitting time of Goal finite

Ruin_Counter=0;
Undecideded_Counter=0;
Success_Counter=0;
for i=1:NumSimu
    path=x(i,:);
    T_0=find(path==0,1);
    T_Goal=find(path==Goal,1);
    if T_0<=n&&T_0<T_Goal
        Ruin_Counter=Ruin_Counter+1;
    elseif T_Goal<=n&&T_Goal<T_0
        Success_Counter=Success_Counter+1;
    else
        Undecideded_Counter=Undecideded_Counter+1;
    end
end

% Simulation result
Ruin_Counter, Undecideded_Counter, Success_Counter
Ruin_Probability_Estimated=Ruin_Counter/(Ruin_Counter+Success_Counter)

%% Output

Ruin_Counter = 89
Undecideded_Counter = 7
Success_Counter = 4
Ruin_Probability_Estimated = 0.9570
```

Gambler's ruin - First step analysis

Decompose the ruin event R according to the first outcome. Let W be the event that you win the first game. Then

$$\begin{aligned}
Q(i) &= \mathbb{P}(R|I=i) \\
&= \mathbb{P}(RW|I=i) + \mathbb{P}(RW^c|I=i) \\
&= \mathbb{P}(W|I=i)\mathbb{P}(R|I=i,W) + \mathbb{P}(W^c|I=i)\mathbb{P}(R|I=i,W^c) \\
&= p\mathbb{P}(R|I=i,W) + q\mathbb{P}(R|I=i,W^c) \\
&= pQ(i+1) + qQ(i-1)
\end{aligned}$$

Recurrence relation $\quad Q(i) = pQ(i+1) + qQ(i-1)$

Boundary conditions $\quad Q(0) = 1 \quad Q(N) = 0$

sparse, spdiags, spdiags_Lee, kron

sparse, spdiags, spdiags_Lee, kron

`S`	`=`	`sparse`	`(i,` ↑ row indices info	`j,` ↑ column indices info	`s)` ↑ values info
`S`	`=`	`spdiags` ↑ start from left edge	`(D,` ↑ diagonals info	`d,` ↑ positions info	`n,m)` ↑ size of `S`
`S`	`=`	`spdiags_Lee` ↑ start from top left edge	`(D,` ↑ diagonals info	`d,` ↑ positions info	`n,m)` ↑ size of `S`
`S`	`=`	`kron`	`(P,` ↑ kron position	`K)` ↑ kron matrix	

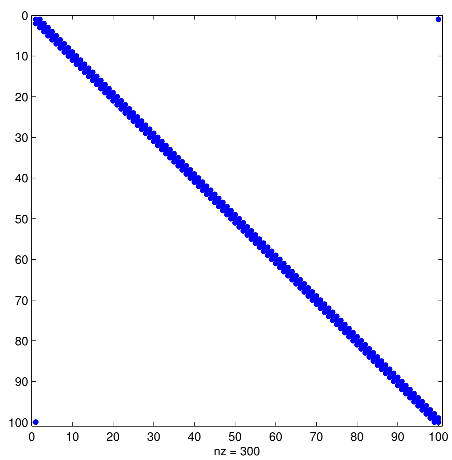
Related functions

	Description
`spy`	Show sparse matrix structure
`nnz`	Count number of non-zeros in sparse matrix
`full`	Turn the sparse matrix in the usual full matrix

Example - spdiags, spdiags_Lee

Construct the following 100×100 matrix as a sparse matrix:

$$\begin{bmatrix} -2 & 1 & & & & & 1 \\ 1 & -4 & 2 & & & & \\ & 2 & -6 & 3 & & & \\ & & \ddots & \ddots & & \ddots & \\ & & & 98 & -198 & & 99 \\ 1 & & & & 99 & & -200 \end{bmatrix}$$




```
clear all; close all; clc;

n = 100;

e1 = (1:n)';                        % should be a column vector
D = [-2*e1 e1 e1    e1      e1]; % diagonals info
d = [    0  1 -1 (n-1) -(n-1)]; % positions info

A1 = spdiags(D,d,n,n);      % start from left edge ---> Not what I want
A2 = spdiags_Lee(D,d,n,n); % start from top left edge ---> Yes I get

B1 = full(A1);
B2 = full(A2);
```

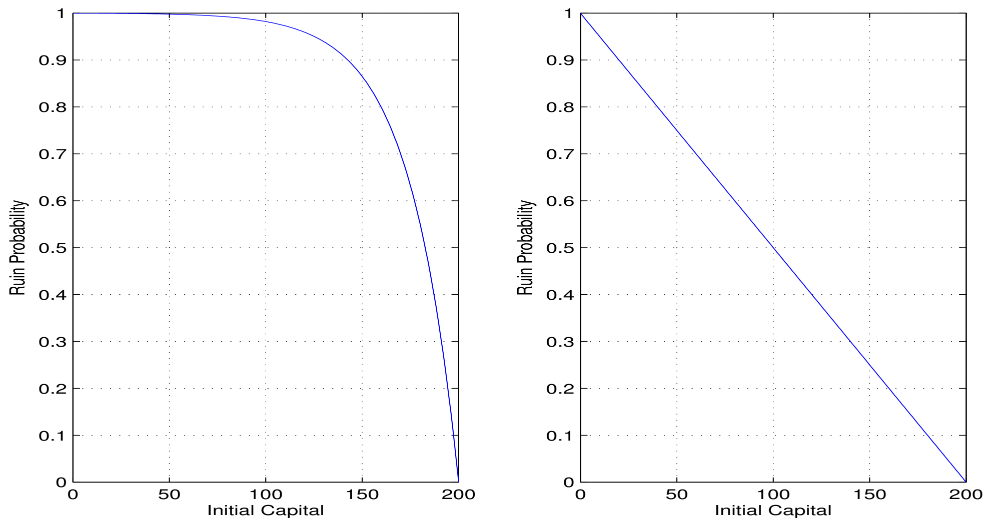


Figure 2: Typically $q > 1/2$ and hence the ruin probability becomes 1 exponentially fast as you lose money a little (left). When $q = 1/2$, the ruin probability becomes 1 lineally as you lose money (right).

```
clear all; close all; clc;

for i=1:2
    if i==1
        p=0.49; q=1-p;
    else
        p=0.50; q=1-p;
    end

    Goal=200;

    d=ones(Goal-1,1);
    A=spdiags([-q*d d -p*d],[-1 0 1],Goal-1,Goal-1);
    b=zeros(Goal-1,1); b(1)=q;
    Ruin_Probab=A\b;
    Ruin_Probab=[1; Ruin_Probab; 0];

    subplot(1,2,i)
    plot(0:Goal,Ruin_Probab); grid on
    xlabel('Initial Capital'); ylabel('Ruin Probability');
end
```

Gambler's ruin - Linear recurrence relation

Linear recurrence relation

$$pQ_1(i+1) + qQ_1(i-1) = Q_1(i) \quad \text{and} \quad pQ_2(i+1) + qQ_2(i-1) = Q_2(i)$$

With $Q := \alpha Q_1 + \beta Q_2$,

$$\begin{aligned}
pQ(i+1) + qQ(i-1) &= p(\alpha Q_1(i+1) + \beta Q_2(i+1)) + q(\alpha Q_1(i-1) + \beta Q_2(i-1)) \\
&= \alpha(pQ_1(i+1) + qQ_1(i-1)) + \beta(pQ_2(i+1) + qQ_2(i-1)) \\
&= \alpha(Q_1(i)) + \beta(Q_2(i)) \\
&= Q(i)
\end{aligned}$$

Characteristic equation

Guessed form of solution	$Q(i) = \lambda^i$
$pQ(i+1) + qQ(i-1) = Q(i) \quad \Rightarrow$	$p\lambda^{i+1} + q\lambda^{i-1} = \lambda^i$
Characteristic equation	$p\lambda^2 + q = \lambda$

Characteristic root

Since $q = 1 - p$,

$$\lambda = p\lambda^2 + 1 - p = p(\lambda+1)(\lambda-1) + 1 \quad \Rightarrow \quad (\lambda - 1)[p(\lambda+1) - 1] \quad \Rightarrow \quad \lambda = 1, \frac{q}{p}$$

Characteristic roots $\quad \lambda = 1 \quad \text{and} \quad \lambda = \frac{q}{p}$

Gambler's ruin - $q > 1/2$ case

Two different characteristic roots and hence two different guessed solutions

$$Q_1(i) = 1 \quad \text{and} \quad Q_2(i) = \left(\frac{q}{p}\right)^i$$

General solution

$$Q(i) = \alpha + \beta \left(\frac{q}{p}\right)^i$$

Two boundary conditions

$$\begin{aligned} &\text{Boundary condition } Q(0) = 1; && \alpha + \beta = 1 \\ &\text{Boundary condition } Q(N) = 0; && \alpha + \beta \left(\tfrac{q}{p}\right)^N = 0 \end{aligned}$$

$$\Rightarrow \quad \alpha = \frac{\left(\frac{q}{p}\right)^N}{\left(\frac{q}{p}\right)^N - 1} \qquad \beta = -\frac{1}{\left(\frac{q}{p}\right)^N - 1}$$

Solution

$$Q(i) = \alpha + \beta \left(\frac{q}{p}\right)^i = \frac{\left(\frac{q}{p}\right)^N - \left(\frac{q}{p}\right)^i}{\left(\frac{q}{p}\right)^N - 1}$$

Why gambler's ruin?

Since $q/p > 1$, $(q/p)^N >>> 1$, and hence

$$Q(i) \approx \frac{\left(\frac{q}{p}\right)^N - \left(\frac{q}{p}\right)^i}{\left(\frac{q}{p}\right)^N} = 1 - \left(\frac{q}{p}\right)^{-(N-i)} = 1 - e^{-(N-i)\log(q/p)}$$

Since $q/p > 1$, $\log(q/p) > 0$ and hence as $i \downarrow 0$

$$\begin{aligned} e^{-(N-i)\log(q/p)} &\downarrow 0 \quad \text{exponentially fast} \\ Q(i) \approx 1 - e^{-(N-i)\log(q/p)} &\uparrow 1 \quad \text{exponentially fast} \end{aligned}$$

Gambler's ruin - $q = 1/2$ case

$\boxed{\text{Double roots and hence one guessed solution}}$

$$Q_1(i) = 1$$

In this case, another solution is given by

$$Q_2(i) = iQ_1(i) = i$$

$\boxed{\text{General solution}}$

$$Q(i) = \alpha + \beta i$$

$\boxed{\text{Two boundary conditions}}$

$$\begin{array}{ll} \text{Boundary condition } Q(0) = 1; & \alpha = 1 \\ \text{Boundary condition } Q(N) = 0; & \alpha + \beta N = 0 \end{array}$$

$$\Rightarrow \quad \alpha = 1 \qquad \beta = -\frac{1}{N}$$

$\boxed{\text{Solution}}$

$$Q(i) = \alpha + \beta i = 1 - \frac{1}{N} i = \frac{N - i}{N}$$

Simpson's paradox

A trend in each group of data disappears when these groups are combined, and the reverse trend appears for the aggregate data.

Good doctor vs bad doctor

Doctor A	Number of successes	Number of fails	Success rate
Easy Operation	10	0	100%
Hard Operation	75	15	83%
Total	85	15	85%

Doctor B	Number of successes	Number of fails	Success rate
Easy Operation	85	5	94%
Hard Operation	1	9	10%
Total	86	14	86%

Berkeley gender bias case

One of the best known real life examples of Simpson's paradox occurred when the University of California, Berkeley was sued for bias against women who had applied for admission to graduate schools there. The admission figures for the fall of 1973 showed that men applying were more likely than women to be admitted, and the difference was so large that it was unlikely to be due to chance.

	Number of applicants	Admitted rate
Men	8442	44 %
Women	4321	35 %

But when examining the individual departments, it appeared that no department was significantly biased against women. In fact, most departments had a "small but statistically significant bias in favor of women". The data from the six largest departments are listed below.

Dept	# M applicants (Admitted rate)	# F applicants (Admitted rate)
A	825 (62 %)	108 (82 %)
B	560 (63 %)	25 (68 %)
C	325 (37 %)	593 (34 %)
D	417 (33 %)	375 (35 %)
E	191 (28 %)	393 (24 %)
F	272 (6 %)	341 (7 %)

How to compute $P(\cap_{i=1}^{n} A_i)$

Independent events - Definition

$$\begin{aligned} P(AB) &= P(A)P(B) \\ P(ABC) &= P(A)P(B)P(C) \\ P(ABCD) &= P(A)P(B)P(C)P(D) \\ \vdots &= \vdots \end{aligned}$$

Dependent events - Chain rule

$$\begin{aligned} P(AB) &= P(A)P(B|A) \\ P(ABC) &= P(A)P(B|A)P(C|AB) \\ P(ABCD) &= P(A)P(B|A)P(C|AB)P(D|ABC) \\ \vdots &= \vdots \end{aligned}$$

How to compute $P(\cup_{i=1}^n A_i)$

Disjoint events

$$\begin{aligned} P\left(\cup_{i=1}^{\infty} A_i\right) &= \sum_{i=1}^{\infty} P\left(A_i\right) \quad \text{for any disjoint events } A_i \\ P\left(\cup_{i=1}^{n} A_i\right) &= \sum_{i=1}^{n} P\left(A_i\right) \quad \text{for any disjoint events } A_i \end{aligned}$$

Non-disjoint events - Inclusion-exclusion principle

$$\begin{aligned} P(\cup_{i=1}^n A_i) &\leq \sum_{i=1}^n P(A_i) \\ P(\cup_{i=1}^n A_i) &\geq \sum_{i=1}^n P(A_i) - \sum_{1\leq i<j\leq n} P(A_i A_j) \\ P(\cup_{i=1}^n A_i) &\leq \sum_{i=1}^n P(A_i) - \sum_{1\leq i<j\leq n} P(A_i A_j) + \sum_{1\leq i<j<k\leq n} P(A_i A_j A_k) \\ &\cdots \\ P(\cup_{i=1}^n A_i) &= \sum_{i=1}^n P(A_i) - \sum_{1\leq i<j\leq n} P(A_i A_j) + \cdots + (-1)^{n+1} P(A_1 A_2 \cdots A_n) \end{aligned}$$

Complement

$$P(\cup_{i=1}^n A_i) = 1 - P(\cap_{i=1}^n A_i^c)$$