Joint, marginal, and conditional distribution

Distribution

Random variable

$$X : \Omega \longrightarrow \mathbb{R}$$

Distribution

Let X be a random variable. We move a brick attached to ω, to $X(\omega)$ in the real line $\mathbb{R}$. In this way we move all the bricks in Ω to $\mathbb{R}$. Then the total weights of the bricks moved into $\mathbb{R}$ is 1. This brick or weight distribution over the real line $\mathbb{R}$ is the distribution of X.

$$\begin{aligned}
\mathbb{P}(X = a) &= \text{Weight of the bricks at } a \\
\mathbb{P}(X \in A) &= \text{Weight of the bricks in } A
\end{aligned}$$

PMF/PDF

$$\begin{aligned}
\text{PMF} \quad p_{x_i} &= \text{Weight of the bricks attached to } x_i \\
\text{PDF} \quad f(x)dx &= \text{Weight of the bricks in } [x, x + dx]
\end{aligned}$$

CDF

$$\begin{aligned}
F(x) &= \mathbb{P}(X \le x) \\
&= \begin{cases} \displaystyle\sum_{x_i \le x} p_{x_i} & \text{if } X \text{ is discrete} \\ \displaystyle\int_{-\infty}^{x} f(s)ds & \text{if } X \text{ is continuous} \end{cases} \\
&= \text{Weight of the bricks cumulatively stacked from } -\infty \text{ up to } x
\end{aligned}$$

Joint distribution

Random vector

$$\mathbf{X} : \Omega \longrightarrow \mathbb{R}^d$$

Joint distribution

Let $\mathbf{X}$ be a random vector. We move a brick attached to ω, to $\mathbf{X}(\omega)$ in $\mathbb{R}^d$. In this way we move all the bricks in Ω to $\mathbb{R}^d$. Then the total weights of the bricks moved into $\mathbb{R}^d$ is 1. This brick or weight distribution over $\mathbb{R}^d$ is the joint distribution of $\mathbf{X}$.

$$\begin{aligned}
\mathbb{P}(\mathbf{X} = \mathbf{a}) &= \text{Weight of the bricks at } \mathbf{a} \\
\mathbb{P}(\mathbf{X} \in A) &= \text{Weight of the bricks in } A
\end{aligned}$$

Joint PMF/PDF

$$\begin{aligned}
\text{Joint PMF} \quad p_{\mathbf{x}} &= \text{Weight of the bricks attached to } \mathbf{x} \\
\text{Joint PDF} \quad f(\mathbf{x})d\mathbf{x} &= \text{Weight of the bricks in } \textstyle\prod_{i=1}^d [x_i, x_i + dx_i]
\end{aligned}$$

Joint CDF

$$\begin{aligned}
F(\mathbf{x}) &= \mathbb{P}(\mathbf{X} \le \mathbf{x}) \\
&= \begin{cases} \displaystyle\sum_{\mathbf{x}_i \le \mathbf{x}} p_{\mathbf{x}_i} & \text{if } \mathbf{X} \text{ is discrete} \\ \displaystyle\int_{-\infty}^{\mathbf{x}} f(\mathbf{s})d\mathbf{s} & \text{if } \mathbf{X} \text{ is continuous} \end{cases} \\
&= \text{Weight of the bricks cumulatively stacked from } -\infty \text{ up to } \mathbf{x}
\end{aligned}$$

Joint, marginal, conditional distribution

Joint, marginal, conditional distribution

Joint distribution $p(x,y) = p(x)p(y|x) = p(y)p(x|y)$

Marginal distribution $p(x) = \sum_y p(x,y), \quad p(y) = \sum_x p(x,y)$

Conditional distribution $p(x|y) = \dfrac{p(x,y)}{p(y)}, \quad p(y|x) = \dfrac{p(x,y)}{p(x)}$

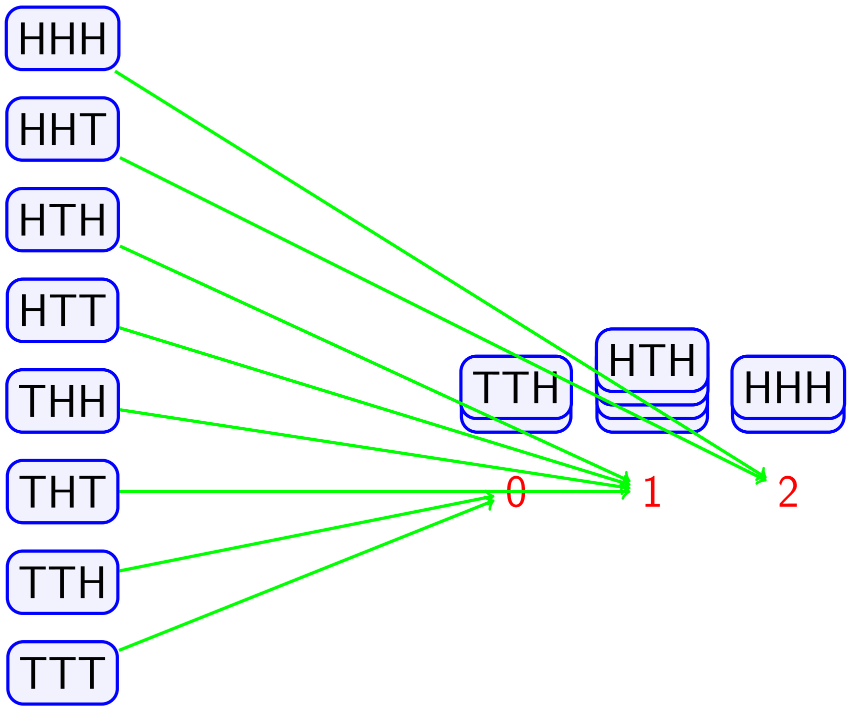
How to get joint, marginal, conditional from other two

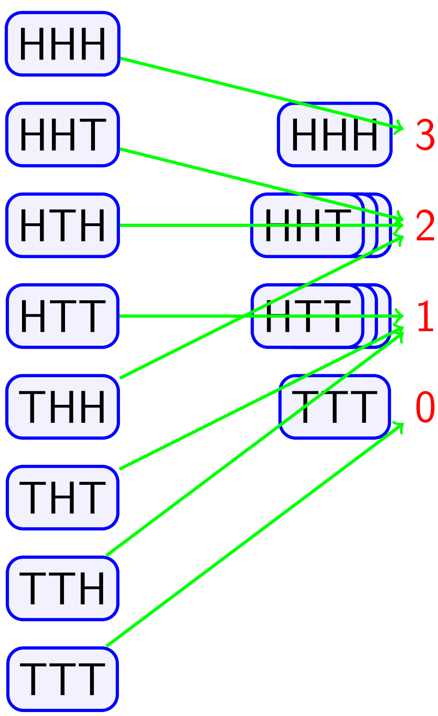
Chain rule $p(x,y) = p(x)p(y|x)$

Marginalization $p(x) = \sum_y p(x,y)$

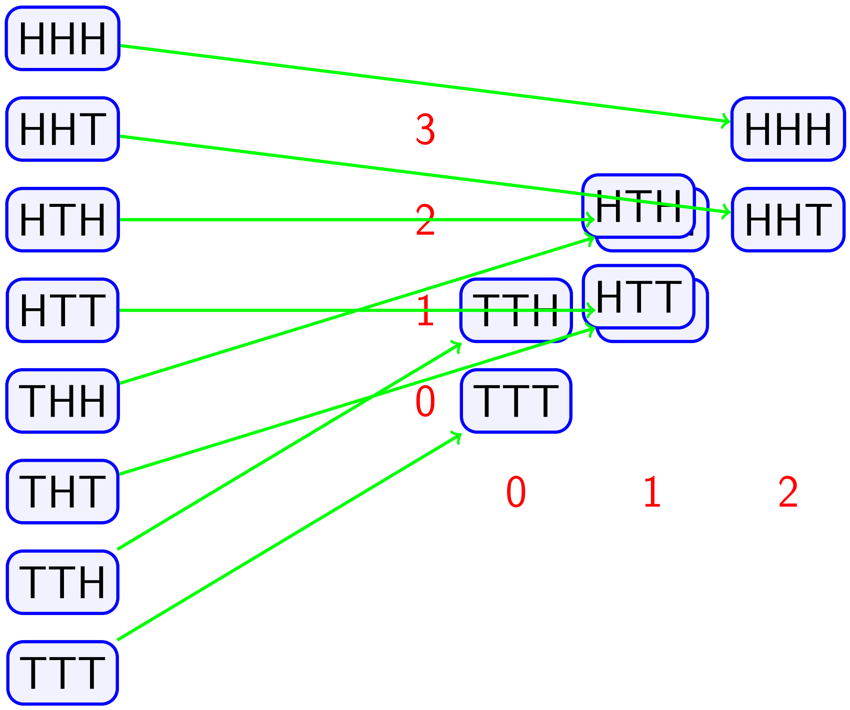
Conditioning $p(y|x) = \dfrac{p(x,y)}{p(x)}$

Example - Number X of heads in first two flips and total number Y of heads

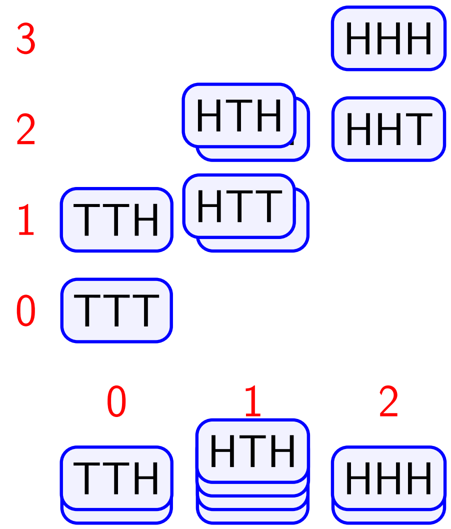


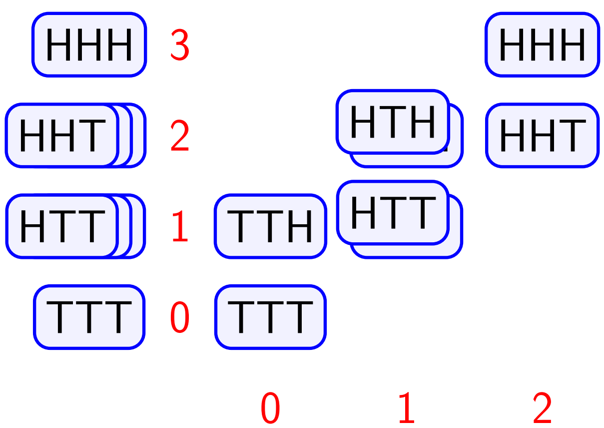


Example - Number of heads in first two flips and total number of heads



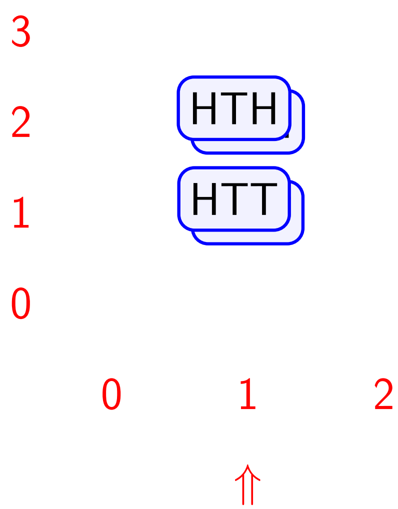
Example - From joint to marginal



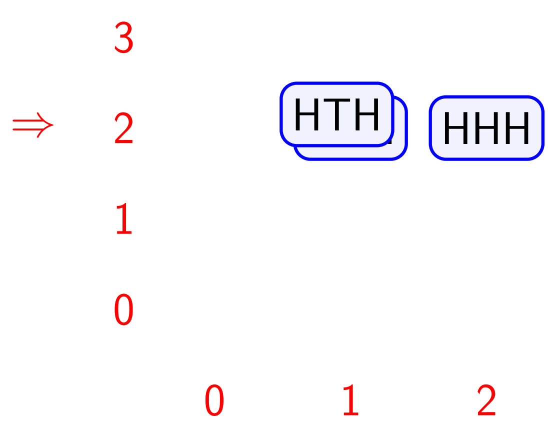


Example - From joint to conditional

Conditional distribution of Y given $X = 1$



Conditional distribution of X given $Y = 2$



Example - From joint to marginal and conditional

The joint PMF of X and Y are given by

y_j				
3	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	
2	$\frac{1}{10}$	0	$\frac{1}{10}$	
1	0	$\frac{2}{10}$	$\frac{1}{10}$	
0	$\frac{1}{10}$	0	$\frac{1}{10}$	
	0	**1**	**2**	x_i

1. Find the marginal PMF of X.
2. Find the marginal PMF of Y.
3. Find the conditional PMF of X given $Y = 1$.
4. Find the conditional PMF of Y given $X = 2$.

1. Do the column sum and get the marginal PMF of X.

y_j				
3	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	
2	$\frac{1}{10}$	0	$\frac{1}{10}$	
1	0	$\frac{2}{10}$	$\frac{1}{10}$	
0	$\frac{1}{10}$	0	$\frac{1}{10}$	
	0	**1**	**2**	x_i
	$\frac{3}{10}$	$\frac{3}{10}$	$\frac{4}{10}$	$\mathbb{P}(X = x_i)$

2. Do the row sum and get the marginal PMF of Y.

$\mathbb{P}(Y = y_j)$	y_j				
$\frac{3}{10}$	**3**	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	
$\frac{2}{10}$	**2**	$\frac{1}{10}$	0	$\frac{1}{10}$	
$\frac{3}{10}$	**1**	0	$\frac{2}{10}$	$\frac{1}{10}$	
$\frac{2}{10}$	**0**	$\frac{1}{10}$	0	$\frac{1}{10}$	
		0	**1**	**2**	x_i

3. To find the conditional PMF of X given $Y = 1$ we remove all the masses except the masses on the line $y = 1$ and then normalize the masses so that the total mass of the remaining masses is 1.

y_j				
3				
2				
1	0	$\frac{2}{10}$	$\frac{1}{10}$	
0				
	0	**1**	**2**	x_i

y_j				
3				
2				
1	0	$\frac{2}{3}$	$\frac{1}{3}$	
0				
	0	**1**	**2**	x_i
	0	$\frac{2}{3}$	$\frac{1}{3}$	$\mathbb{P}(X = x_i \mid Y = 1)$

4. To find the conditional PMF of Y given $X = 2$ we remove all the masses except the masses on the line $x = 2$ and then normalize the masses so that the total mass of the remaining masses is 1.

y_j				
3			$\frac{1}{10}$	
2			$\frac{1}{10}$	
1			$\frac{1}{10}$	
0			$\frac{1}{10}$	
	0	**1**	**2**	x_i

$\mathbb{P}(Y = y_j \mid X = 2)$	y_j				
$\frac{1}{4}$	**3**			$\frac{1}{4}$	
$\frac{1}{4}$	**2**			$\frac{1}{4}$	
$\frac{1}{4}$	**1**			$\frac{1}{4}$	
$\frac{1}{4}$	**0**			$\frac{1}{4}$	
		0	**1**	**2**	x_i

Example - 3 red balls and 1 blue ball

There are 3 red balls and 1 blue ball in the bin.

$$\{r, r, r, b\}$$

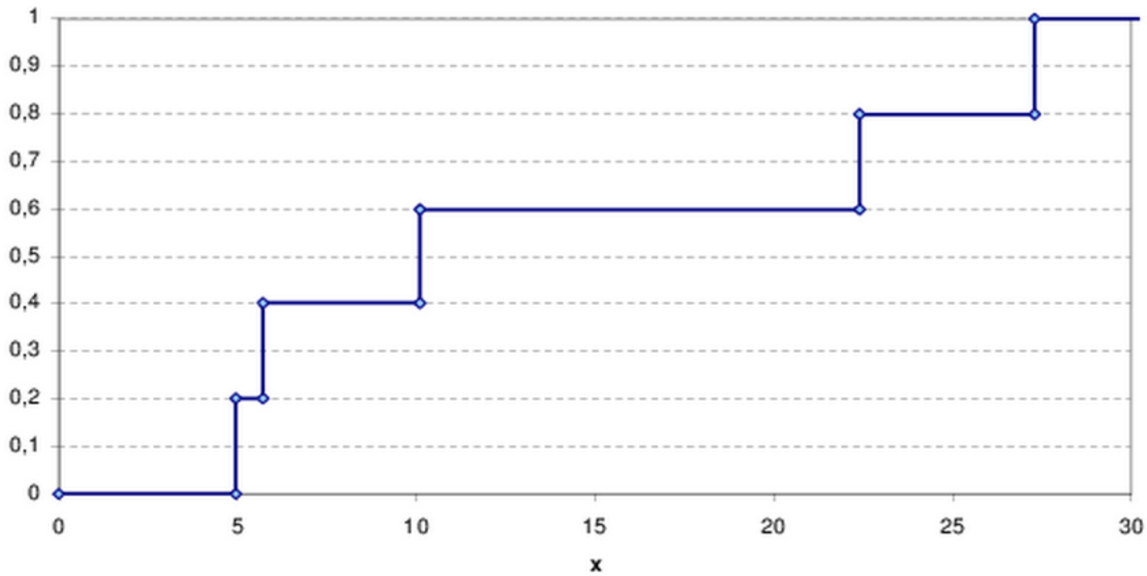
We choose the first ball and record its color. After the record remove the first ball from the bin. Then we choose the second ball and record its color. Find the probability that the first ball is red and the second is blue.

$$X_i = \begin{cases} 1 & \text{if } i\text{-th chosen ball blue} \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{P}(X_1 = 0, X_2 = 1) = \mathbb{P}(X_1 = 0)\mathbb{P}(X_2 = 1 | X_1 = 0) = \frac{3}{4} \times \frac{1}{3} = \frac{1}{4}$$

Example - CDF

Using the below CDF find $\mathbb{P}(X = 5)$ and $\mathbb{P}(X \geq 20)$.



$$\mathbb{P}(X = 5) = F(5) - F(5-) = 0.2 - 0 = 0.2$$

$$\mathbb{P}(X < 20) = F(20-) = 0.6 \quad \Rightarrow \quad \mathbb{P}(X \geq 20) = 1 - \mathbb{P}(X < 20) = 0.4$$

Independent, pairwise independent, conditionally independent random variables

Independent random variables

X and Y are independent if for any x and y

$$p(x, y) = p(x)p(y)$$

$X_1, X_2, \cdots, X_n$ are independent if for any x_1, x_2,$\cdots$, x_n

$$p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2)\cdots p(x_n)$$

Pairwise independent random variables

$X_1, X_2, \cdots, X_n$ are pairwise independent if for any pair X_i, X_j

$$X_i \text{ and } X_j \text{ are independent}$$

Conditionally independent random variables

X_1, $\cdots$, X_n are conditionally independent conditioned on Y if for any x_1, $\cdots$, x_n, y

$$p(x_1, x_2, \ldots, x_n|y) = p(x_1|y)p(x_2|y)\cdots p(x_n|y)$$

Example - Two independent random variables

The marginal PMFs of X and Y are given by

$\mathbb{P}(Y = y_j)$	y_j				
$\frac{3}{10}$	**3**				
$\frac{2}{10}$	**2**				
$\frac{3}{10}$	**1**				
$\frac{2}{10}$	**0**				
		0	**1**	**2**	x_i
		$\frac{2}{10}$	$\frac{3}{10}$	$\frac{5}{10}$	$\mathbb{P}(X = x_i)$

Suppose X and Y are independent. Find the joint PMF of X and Y, i.e., fill up the blank in below table.

y_j				
3				
2				
1				
0				
	0	**1**	**2**	x_i

$\mathbb{P}(Y = y_j)$	y_j				
$\frac{3}{10}$	**3**	$\frac{2}{10} \times \frac{3}{10} = \frac{6}{100}$	$\frac{3}{10} \times \frac{3}{10} = \frac{9}{100}$	$\frac{5}{10} \times \frac{3}{10} = \frac{15}{100}$	
$\frac{2}{10}$	**2**	$\frac{2}{10} \times \frac{2}{10} = \frac{6}{100}$	$\frac{3}{10} \times \frac{2}{10} = \frac{6}{100}$	$\frac{5}{10} \times \frac{2}{10} = \frac{10}{100}$	
$\frac{3}{10}$	**1**	$\frac{2}{10} \times \frac{3}{10} = \frac{6}{100}$	$\frac{3}{10} \times \frac{3}{10} = \frac{9}{100}$	$\frac{5}{10} \times \frac{3}{10} = \frac{15}{100}$	
$\frac{2}{10}$	**0**	$\frac{2}{10} \times \frac{2}{10} = \frac{6}{100}$	$\frac{3}{10} \times \frac{2}{10} = \frac{6}{100}$	$\frac{5}{10} \times \frac{2}{10} = \frac{10}{100}$	
		0	**1**	**2**	x_i
		$\frac{2}{10}$	$\frac{3}{10}$	$\frac{5}{10}$	$\mathbb{P}(X = x_i)$

Example - Two dependent random variables

The joint PMF of X and Y is given by

y_j				
3	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	
2	$\frac{1}{6}$	$\frac{1}{6}$	0	
1	$\frac{1}{6}$	0	0	
0	0	0	0	
	0	**1**	**2**	x_i

Determine whether X and Y are independent.

Conditional PMF X given $Y = 1$ $\quad p_{0|1} = 1$

Conditional PMF X given $Y = 2$ $\quad p_{0|2} = 1/2, \quad p_{1|2} = 1/2$

Conditional PMF X given $Y = 3$ $\quad p_{0|3} = 1/3, \quad p_{1|3} = 1/3, \quad p_{2|3} = 1/3$

The conditional PMF of X given $Y = y_j$ depends on y_j and hence X and Y are dependent.

Conditional PMF Y given $X = 0$ $\quad p_{1|0} = 1/3, \quad p_{2|0} = 1/3, \quad p_{3|0} = 1/3$

Conditional PMF Y given $X = 1$ $\quad p_{2|1} = 1/2, \quad p_{3|1} = 1/2$

Conditional PMF Y given $X = 2$ $\quad p_{3|2} = 1$

The conditional PMF of Y given $X = x_i$ depends on x_i and hence X and Y are dependent.

Distributions related to coin flips

Distribution	Random variable
$B(p)$	Flip a p-coin and check whether we have a head
$B(n,p)$	Flip a p-coin n times and count the number of heads
$Geo(p)$	Flip a p-coin until first head and count the number of flips
$NB(r,p)$	Flip a p-coin until r-th head and count the number of flips

Distribution	Expectation	Variance
$B(p)$	p	pq
$B(n,p)$	np	npq
$G(p)$	$\frac{1}{p}$	$\frac{q}{p^2}$
$NB(r,p)$	$\frac{r}{p}$	$\frac{rq}{p^2}$

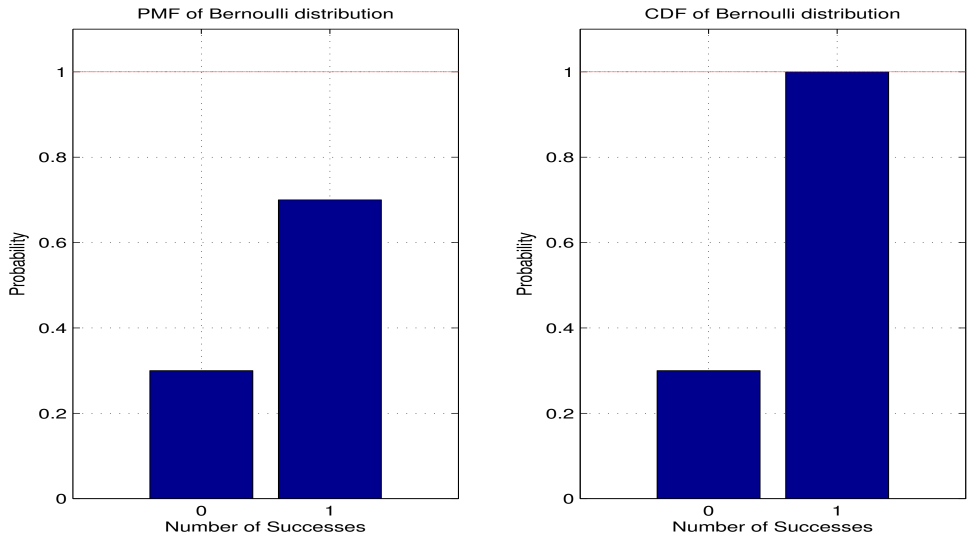


Figure 1: PMF and CDF of Bernoulli distribution $B(0.7)$

```
n=1;
p=0.7;

i=0:n;
pmf=binomial(n,i).*(p.^i).*((1-p).^(n-i));
cdf=cumsum(pmf);

subplot(1,2,1)
bar(i,pmf); grid on; hold on;
plot(-1:0.01:n+1,1,'-r');
axis([-1 n+1 0 1.1]);
xlabel('Number of Successes'); ylabel('Probability');
title('PMF of Bernoulli distribution');

subplot(1,2,2)
bar(i,cdf); grid on; hold on;
plot(-1:0.01:n+1,1,'-r')
axis([-1 n+1 0 1.1]);
xlabel('Number of Successes'); ylabel('Probability');
title('CDF of Bernoulli distribution');
```

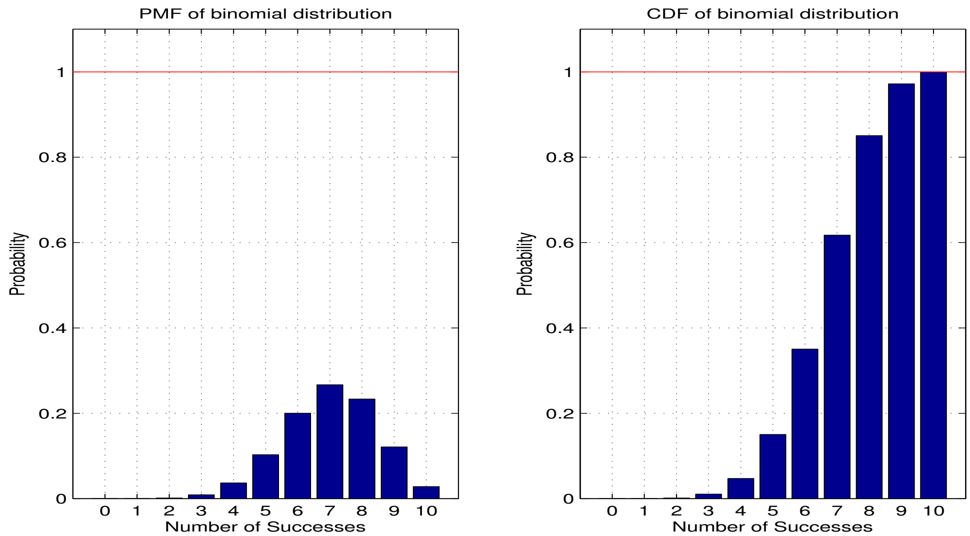


Figure 2: PMF and CDF of binomial distribution $B(10, 0.7)$

```
n=10;
p=0.7;

i=0:n;
pmf=binomial(n,i).*(p.^i).*((1-p).^(n-i));
cdf=cumsum(pmf);

subplot(1,2,1)
bar(i,pmf); grid on; hold on;
plot(-1:0.01:n+1,1,'-r');
axis([-1 n+1 0 1.1])
xlabel('Number of Successes'); ylabel('Probability');
title('PMF of binomial distribution');

subplot(1,2,2)
bar(i,cdf); grid on; hold on;
plot(-1:0.01:n+1,1,'-r');
axis([-1 n+1 0 1.1])
xlabel('Number of Successes'); ylabel('Probability');
title('CDF of binomial distribution');
```

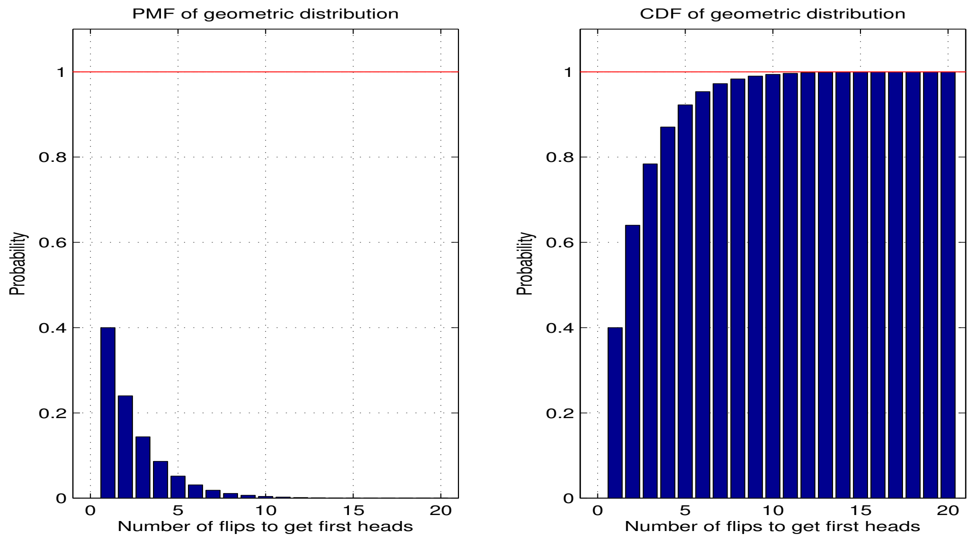


Figure 3: PMF and CDF of geometric distribution $Geo(0.4)$

```
r=1;
p=0.4;
Truncation_Level=20;

i=r:Truncation_Level;
pmf=binomial(i-1,r-1).*p.^r.*(1-p).^(i-r);
cdf=cumsum(pmf);

subplot(1,2,1)
bar(i,pmf); grid on; hold on;
plot(-1:0.01:Truncation_Level+1,1,'-r')
axis([-1 Truncation_Level+1 0 1.1])
xlabel('Number of flips to get first heads'); ylabel('Probability');
title('PMF of geometric distribution');

subplot(1,2,2)
bar(i,cdf); grid on; hold on;
plot(-1:0.01:Truncation_Level+1,1,'-r')
axis([-1 Truncation_Level+1 0 1.1])
xlabel('Number of flips to get first heads'); ylabel('Probability');
title('CDF of geometric distribution');
```

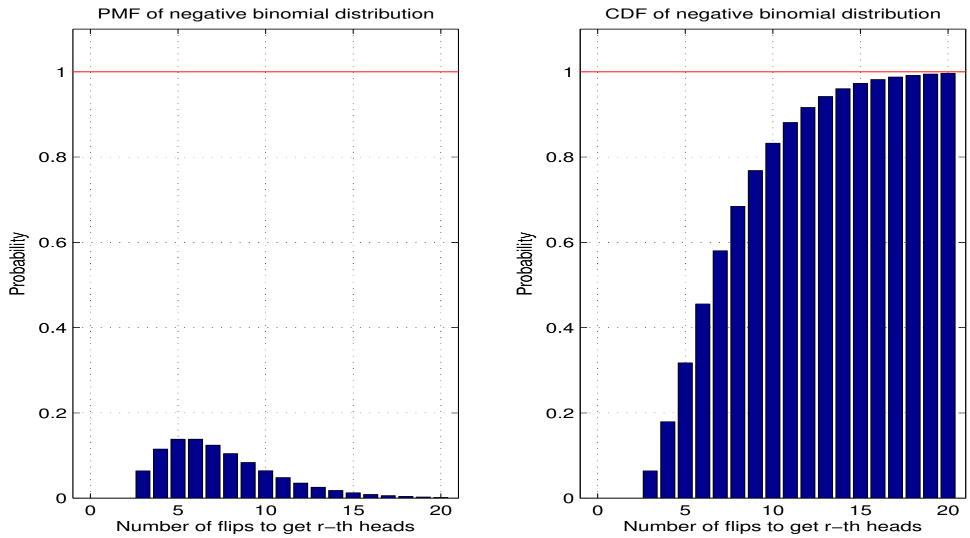


Figure 4: PMF and CDF of negative binomial distribution $NB(3, 0.4)$

```
r=3;
p=0.4;
Truncation_Level=20;

i=r:Truncation_Level;
pmf=binomial(i-1,r-1).*p.^r.*(1-p).^(i-r);
cdf=cumsum(pmf);

subplot(1,2,1)
bar(i,pmf); grid on; hold on;
plot(-1:0.01:Truncation_Level+1,1,'-r')
axis([-1 Truncation_Level+1 0 1.1])
xlabel('Number of flips to get r-th heads'); ylabel('Probability');
title('PMF of negative binomial distribution');

subplot(1,2,2)
bar(i,cdf); grid on; hold on;
plot(-1:0.01:Truncation_Level+1,1,'-r')
axis([-1 Truncation_Level+1 0 1.1])
xlabel('Number of flips to get r-th heads'); ylabel('Probability');
title('CDF of negative binomial distribution');
```

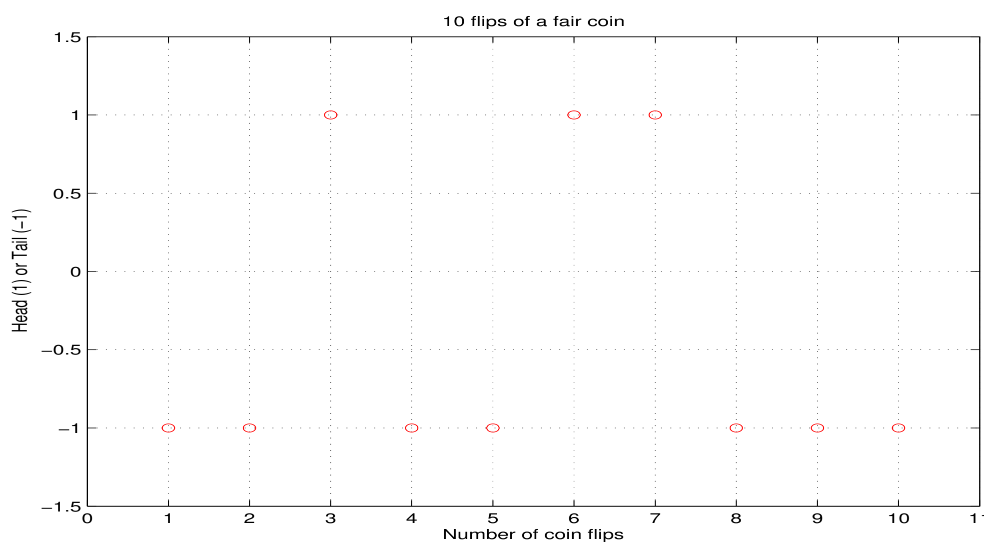
Example - How to generate ± 1 random variables from Bernoulli random variables

If X has a Bernoulli distribution $\mathrm{B}(p)$ with success rate p, then

$$Y := 2X - 1 = \begin{cases} 1 & \text{with probability } p \\ -1 & \text{with probability } 1-p \end{cases}$$

In particular, suppose we flip a fair coin n times independently and we record the i-th flip result as X_i by converting H and T to 1 and 0 respectively. Then X_i is either 1 or 0 equally likely. Now, let $Y_i = 2X_i - 1$. Then

$$X_i \text{ iid with } X_i = \begin{cases} 1 & \text{with prob } 0.5 \\ 0 & \text{with prob } 0.5 \end{cases} \Rightarrow Y_i \text{ iid with } Y_i = \begin{cases} 1 & \text{with prob } 0.5 \\ -1 & \text{with prob } 0.5 \end{cases}$$




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

p=0.5; n=10; % We flip a fair coin n times
x=random('Binomial',1*ones(1,n),p*ones(1,n));
x=2*x-1;

plot(1:n,x,'or'); grid on; axis([0 n+1 -1.5 1.5])
title('10 flips of a fair coin')
xlabel('Number of coin flips'); ylabel('Head (1) or Tail (-1)')
```

$X+Y$ when $X \sim B(n,p)$, $Y \sim B(m,p)$

$X \sim B(n,p)$, $Y \sim B(m,p) \quad \Rightarrow \quad X+Y \sim B(n+m,p)$ Not true in general

$X \sim B(n,p)$, $Y \sim B(m,p) \quad \Rightarrow \quad X+Y \sim B(n+m,p)$ True if X, Y independent

Proof. (Story) We flip the p-coin n times and count the number X of heads. We flip this coin m times additionally and count the number Y of heads in these additional coin flips. Then, $X+Y$ is simply the number of heads in these $n+m$ coin flips. So, $X+Y$ is $B(n+m,p)$.

Proof. (Divide and conquer) Divide:

$$\mathbb{P}(X+Y=k) = \sum_{0 \le l \le n,\ 0 \le k-l \le m} \mathbb{P}(X=l, Y=k-l)$$

Conquer:

$$\begin{aligned}
\mathbb{P}(X=l, Y=k-l) &= \mathbb{P}(X=l)\mathbb{P}(Y=k-l|X=l) \quad (X \text{ and } Y \text{ are independent}) \\
&= \mathbb{P}(X=l)\mathbb{P}(Y=k-l) \quad (X \sim B(n,p),\ Y \sim B(m,p)) \\
&= \binom{n}{l} p^l q^{n-l} \binom{m}{k-l} p^{k-l} q^{m-(k-l)}
\end{aligned}$$

By Vandermonde's identity

$$\begin{aligned}
\mathbb{P}(X+Y=k) &= \sum_{0 \le l \le n, 0 \le k-l \le m} \binom{n}{l} p^l q^{n-l} \binom{m}{k-l} p^{k-l} q^{m-(k-l)} \\
&= \left[\sum_{0 \le l \le n, 0 \le k-l \le m} \binom{n}{l} \binom{m}{k-l} \right] p^k q^{n+m-k} \\
&= \binom{n+m}{k} p^k q^{n+m-k}
\end{aligned}$$

Example - Number of couples with same birthday

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 and let $1_{A_{ij}}$ be its indicator. Then, the number X of common birthday pairs can be represented as

$$X = \sum_{1\leq i<j\leq n}^{n} 1_{A_{ij}}$$

Is X binomial?

The indicator random $1_{A_{ij}}$ is either 1 or 0. So it is a Bernoulli random variable with success rate $p = P(A_{ij}) = 1/365$;

$$1_{A_{ij}} \sim B(p)$$

$$\begin{array}{llll} 1_{A_{ij}} \text{ are independent} & \Rightarrow & X \sim B(m,p), \quad m = \binom{n}{2} & \text{Not true} \\ 1_{A_{ij}} \text{ are not independent} & \Rightarrow & X \text{ not binomial} & \text{True} \end{array}$$

Suppose 1 and 2 share the common birthday and suppose 1 and 3 share the common birthday. Then, of course 2 and 3 share the common birthday. In other word

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

Distributions related to dice rolling

Distribution	Random variable
$Cat(\mathbf{p})$	Roll a $\mathbf{p}$-dice and check the result
$Mul(n, \mathbf{p})$	Roll a $\mathbf{p}$-dice n times and count the number of out comes

Parameter $\mathbf{p} = (p_1, \ldots, p_K)$

$$p_j \geq 0, \qquad \sum_{i=1}^{K} p_j = 1$$

Categorical distribution $Cat(\mathbf{p})$

$$\mathbb{P}(X_1 = 0, \ldots, X_j = 1, \ldots, X_K = 0) = p_j$$

Multinomial distribution $Mul(n, \mathbf{p})$

$$\mathbb{P}(X_1 = n_1, \ldots, X_j = n_j, \ldots, X_K = n_K) = \binom{n}{n_1 \cdots n_K} p_1^{n_1} \cdots p_K^{n_K}$$