ILLC Project Course in Information Theory

Crash course

13 January – 17 January 2014
12:00 to 14:00

Student presentations

27 January – 31 January 2014
12:00 to 14:00

Location

ILLC, room F1.15,
Science Park 107, Amsterdam

Materials

informationtheory.weebly.com

Contact

Mathias Winther Madsen
mathias.winther@gmail.com

Monday

Probability theory
Uncertainty and coding

Tuesday

The weak law of large numbers
The source coding theorem

Wednesday

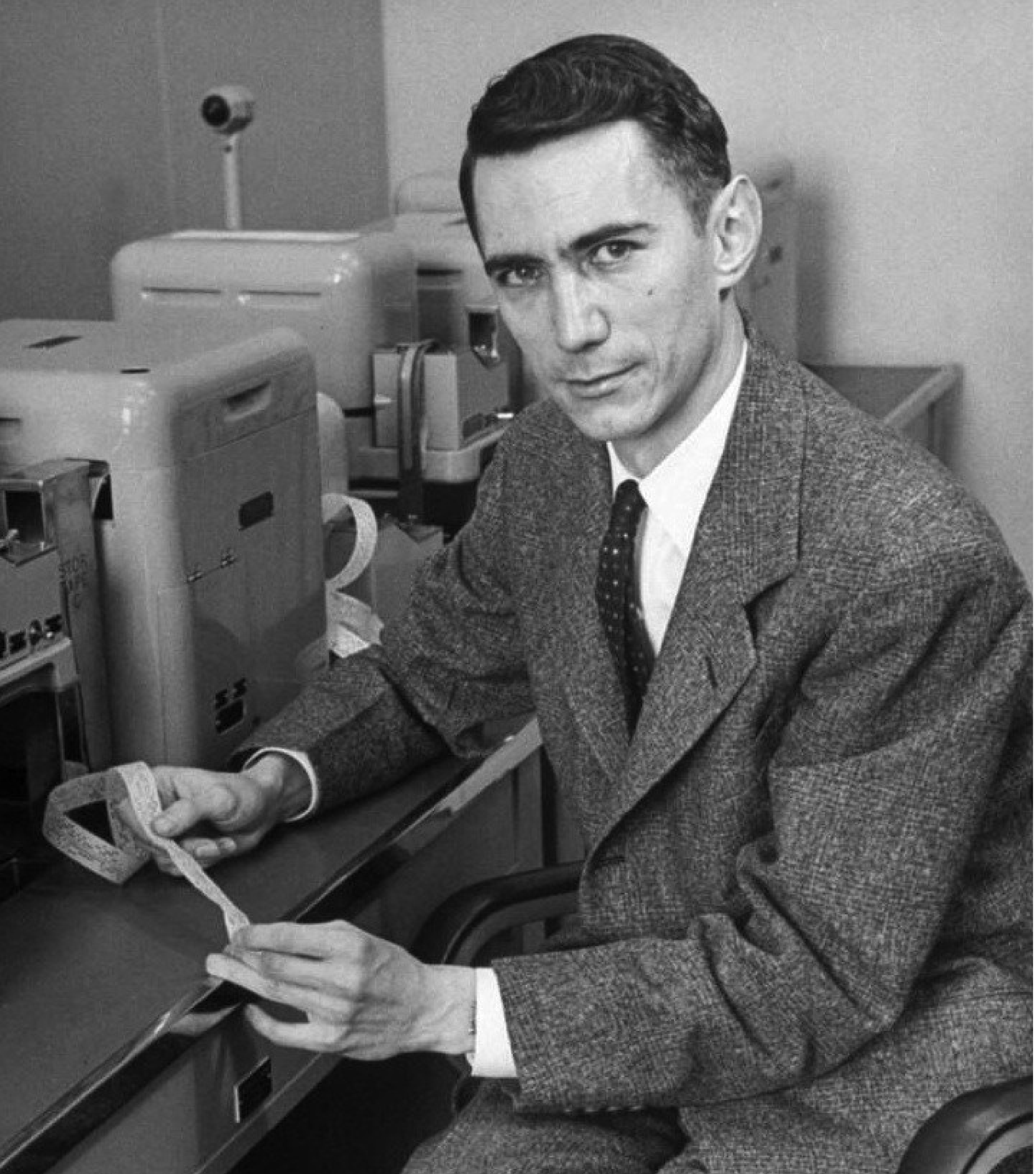
Random processes
Arithmetic coding

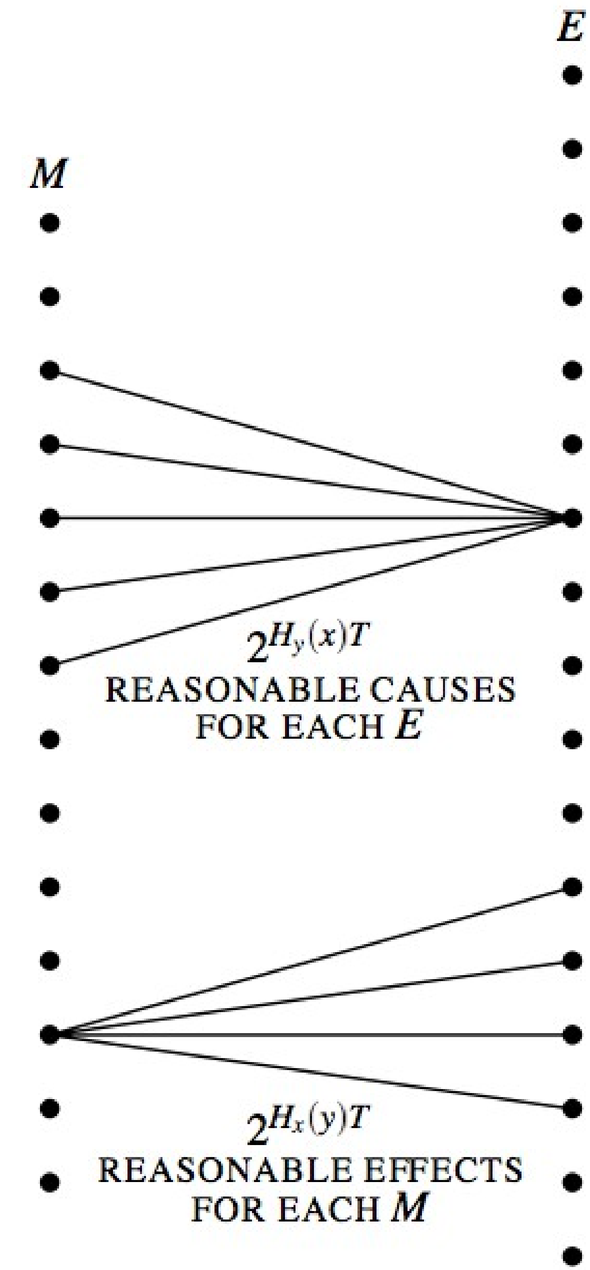
Thursday

Divergence
Kelly Gambling

Friday

Kolmogorov Complexity
The limits of statistics



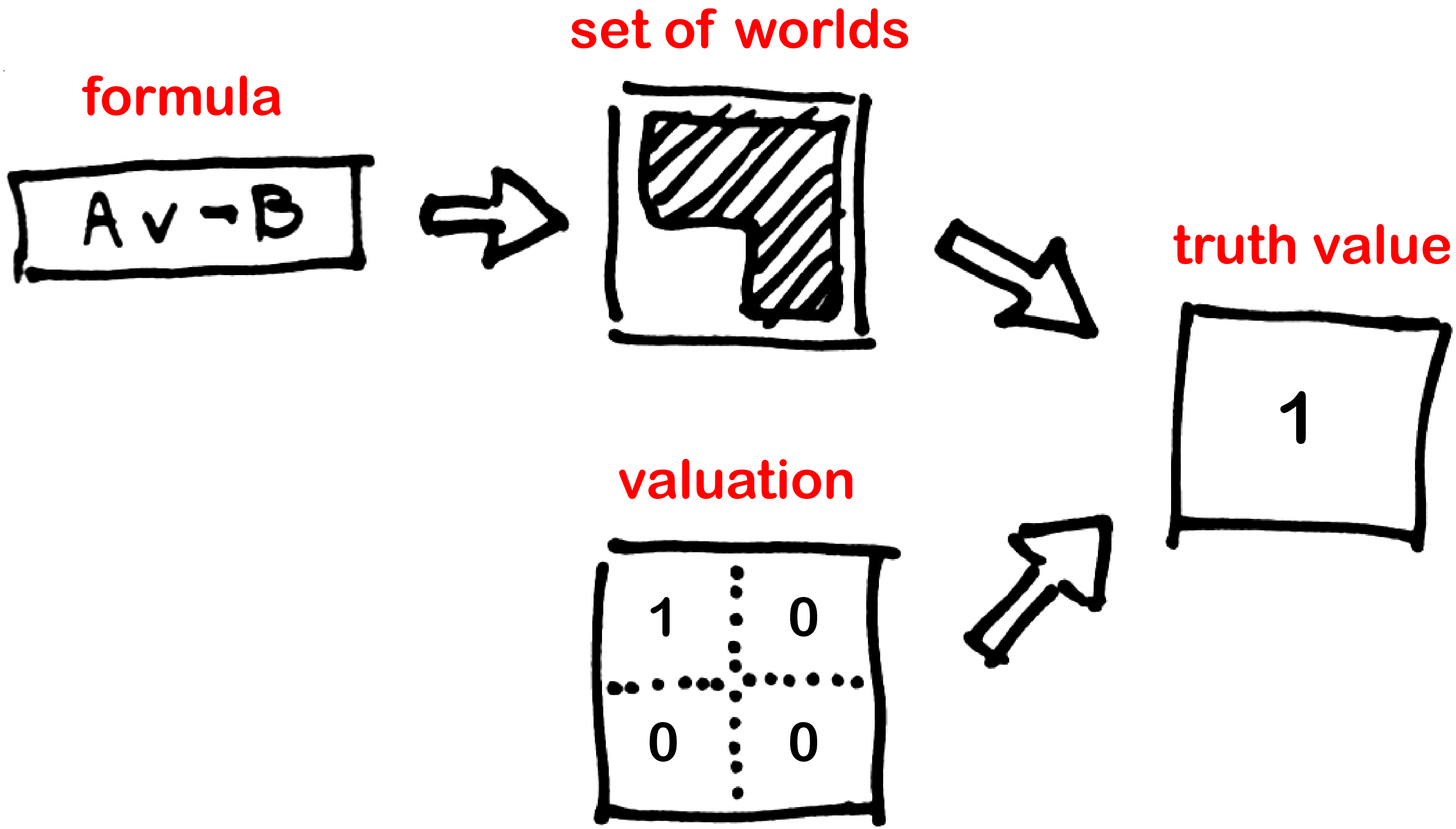


M
E
$2^{H_y(x)T}$
REASONABLE CAUSES
FOR EACH E
$2^{H_x(y)T}$
REASONABLE EFFECTS
FOR EACH M

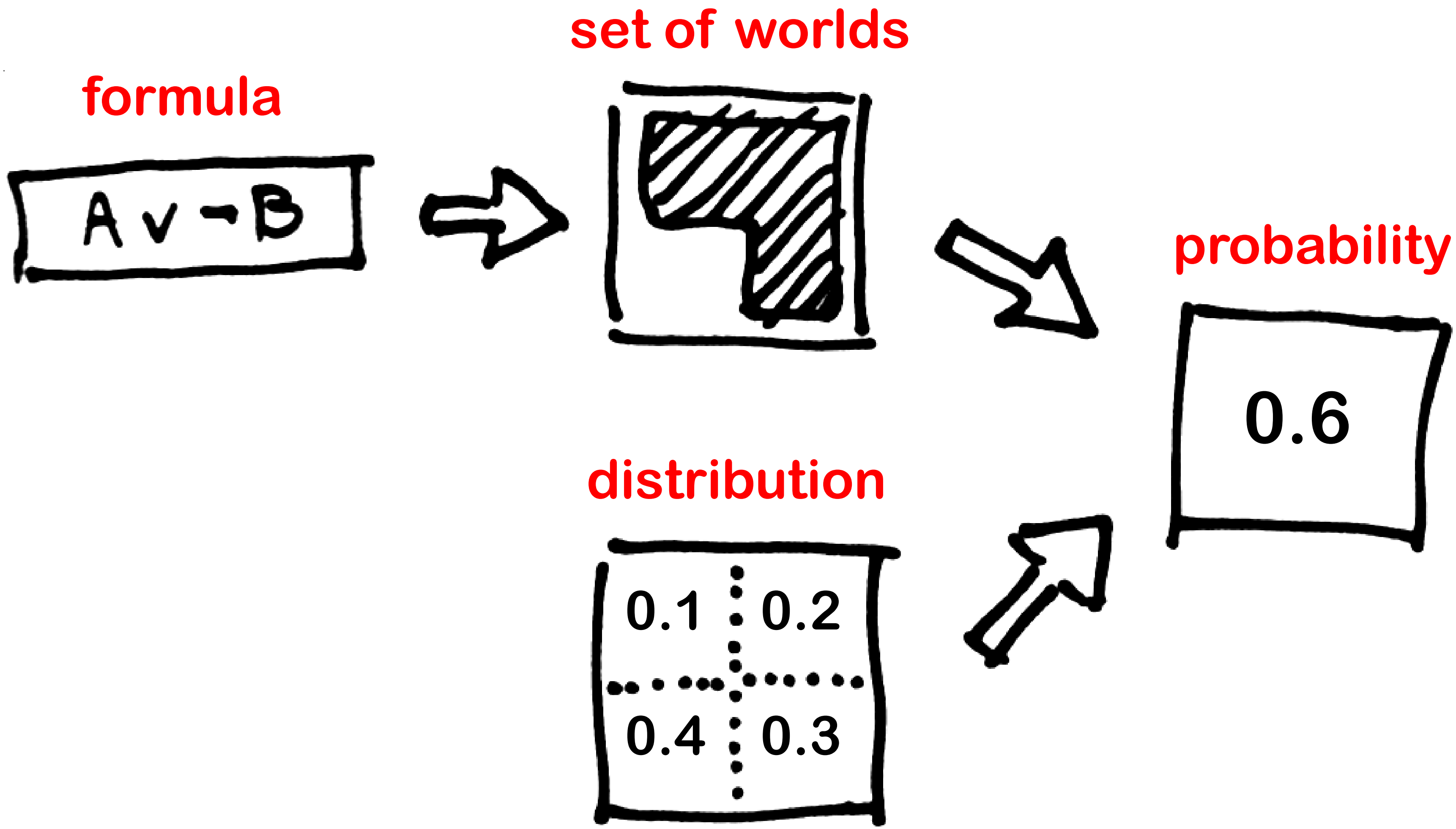
Probability theory:

Core concepts	Distributions Random variables Independence
Operations	Marginalization Conditionalization
Theorems	The chain rule

Propositional logic:



Probability theory:



Additivity axiom: The probabilities of disjoint sets add up

w	w_1	w_2	w_3
$\Pr\{w\}$	0.1	0.4	0.5

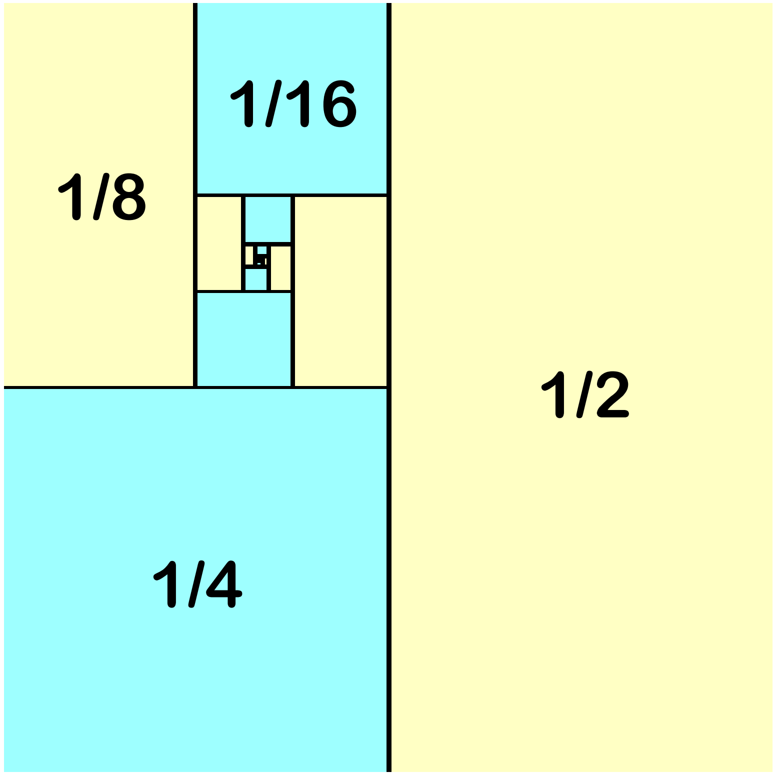
$$\begin{aligned} \Pr\{w_1, w_2\} &= \Pr\{w_1\} + \Pr\{w_2\} \\ &= 0.1 + 0.4 \\ &= 0.5 \end{aligned}$$

Example: A geometric distribution

w	w_1	w_2	w_3	w_4	...
$\Pr\{w\}$	1/2	1/4	1/8	1/16	...

$$\Pr\{w_1, w_3, w_5, \ldots\} = ?$$

$$\Pr\{w_2, w_4, w_6, \ldots\} = ?$$



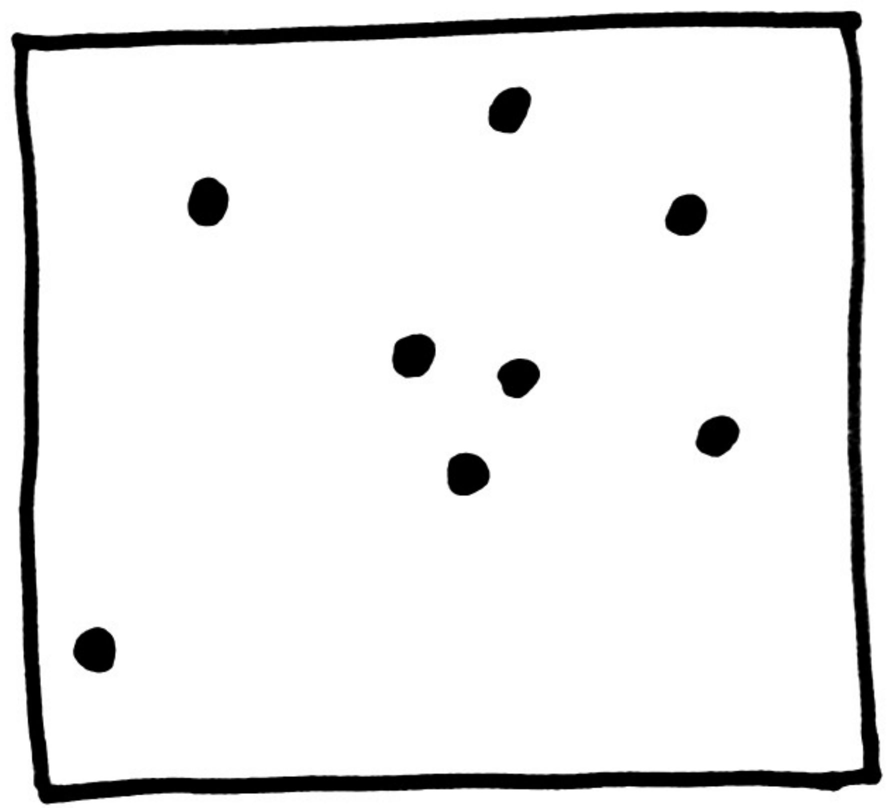
“Irrational” distributions?

“The sum of two rolls of a die is 2, 3, …, 11, or 12, each of which have probability 1/11.”

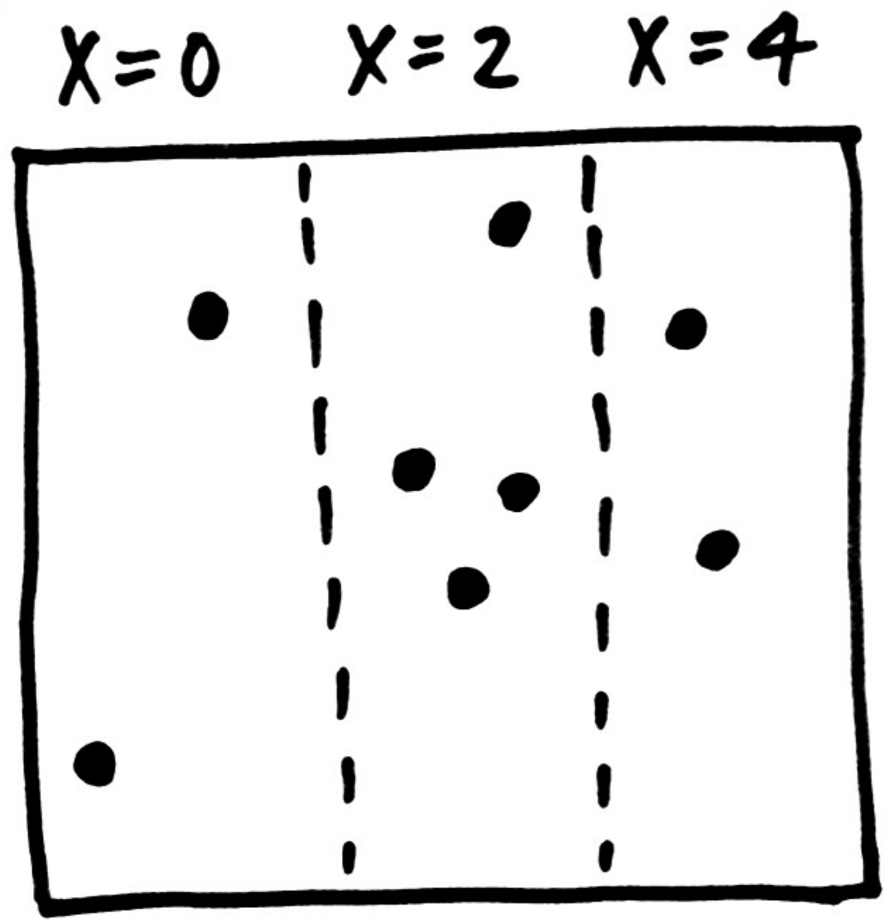
“The probability that the 100th digit of π is 5 is 1/10.”

Cf. plato.stanford.edu/entries/dutch-book

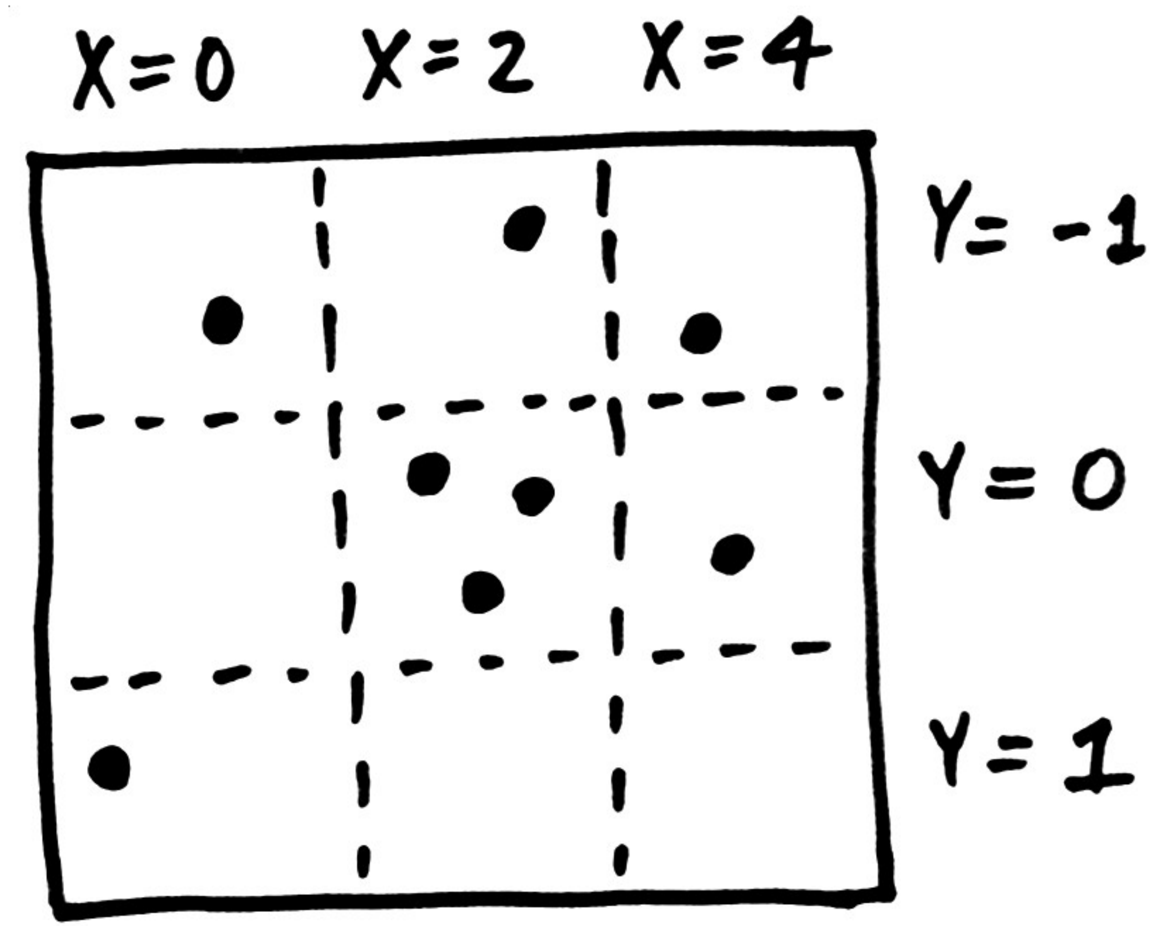
Random variables

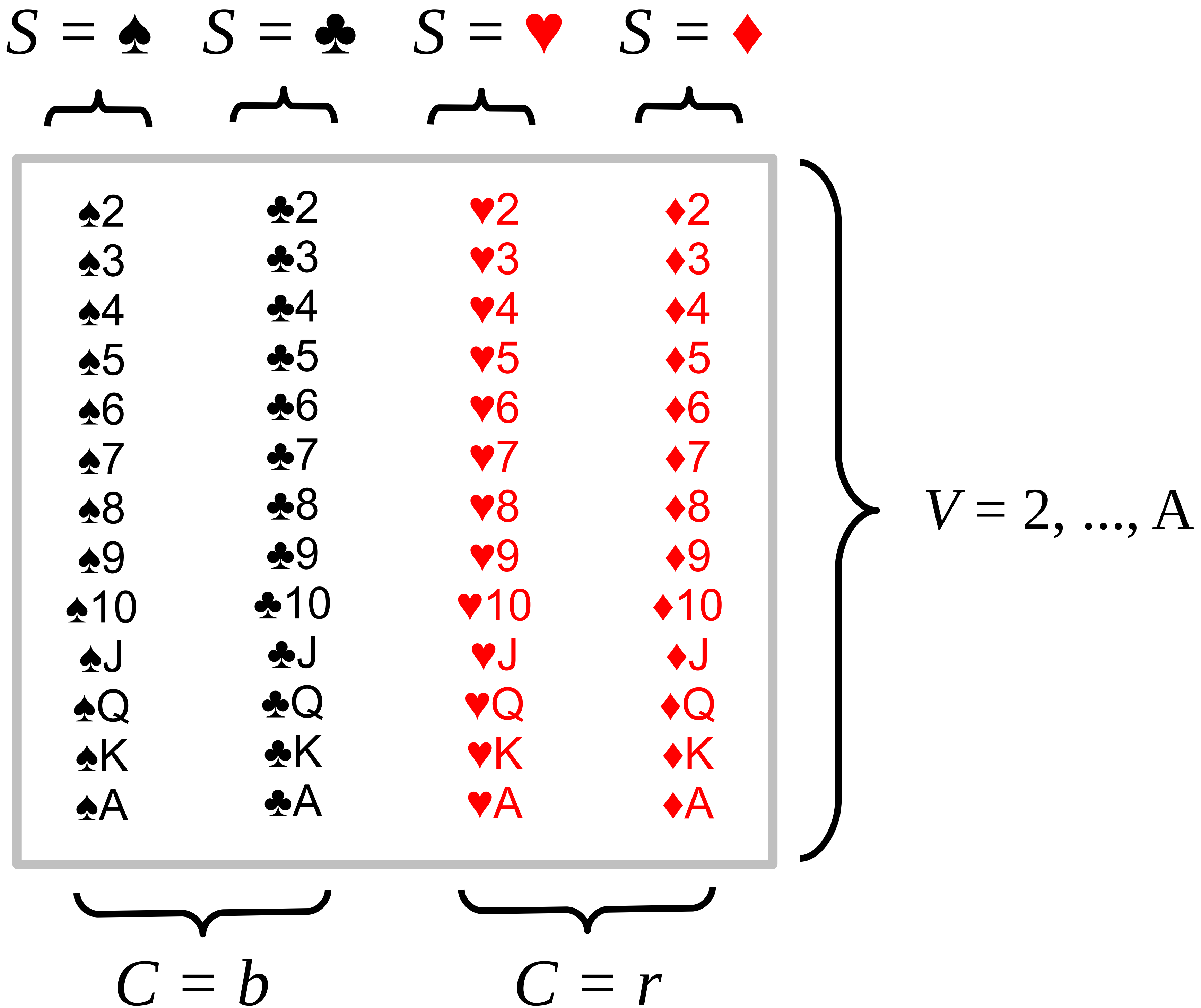


Random variables

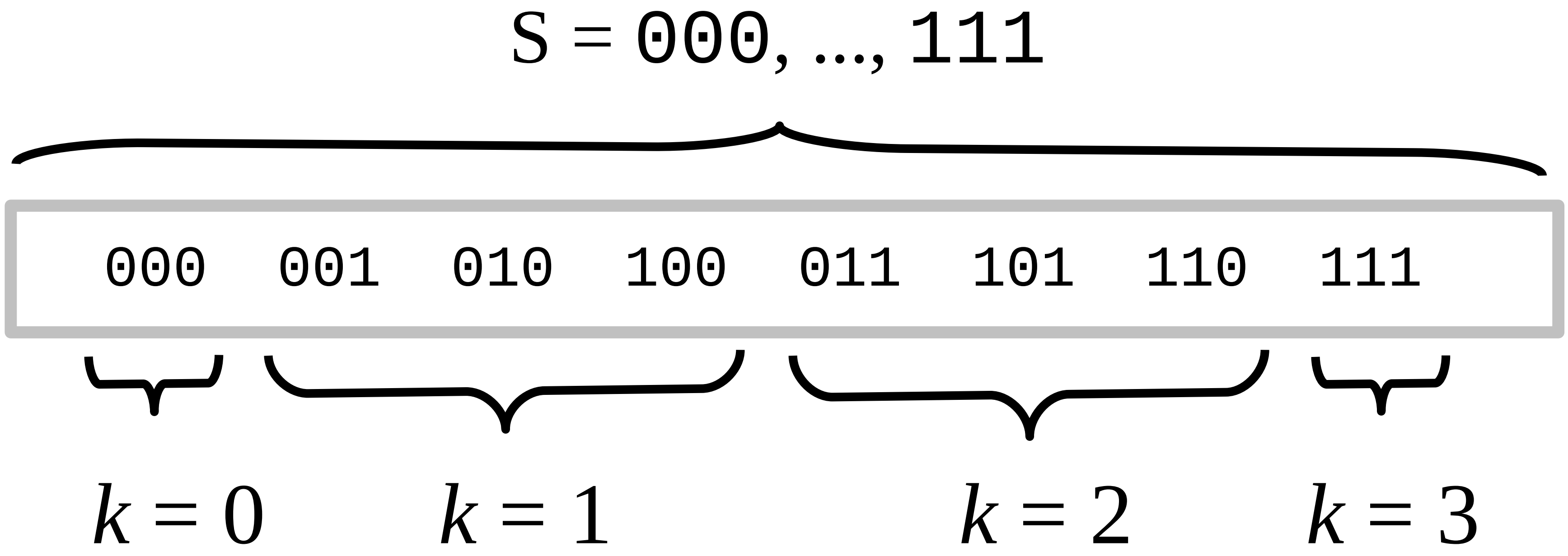


Random variables





S = ♠
S = ♣
S = ♥
S = ♦
♠2 ♠3 ♠4 ♠5 ♠6 ♠7 ♠8 ♠9 ♠10 ♠J ♠Q ♠K ♠A
♣2 ♣3 ♣4 ♣5 ♣6 ♣7 ♣8 ♣9 ♣10 ♣J ♣Q ♣K ♣A
♥2 ♥3 ♥4 ♥5 ♥6 ♥7 ♥8 ♥9 ♥10 ♥J ♥Q ♥K ♥A
♦2 ♦3 ♦4 ♦5 ♦6 ♦7 ♦8 ♦9 ♦10 ♦J ♦Q ♦K ♦A
V = 2, …, A
C = b
C = r


S = 000, ..., 111
000 001 010 100 011 101 110 111
k = 0
k = 1
k = 2
k = 3

	X = 0	X = 1	Σ
Y = 0	0.40	0.30	0.70
Y = 1	0.20	0.10	0.30
Σ	0.60	0.40	1.00

Dependent

	X = 0	X = 1	Σ
Y = 0	0.42	0.28	0.70
Y = 1	0.18	0.12	0.30
Σ	0.60	0.40	1.00

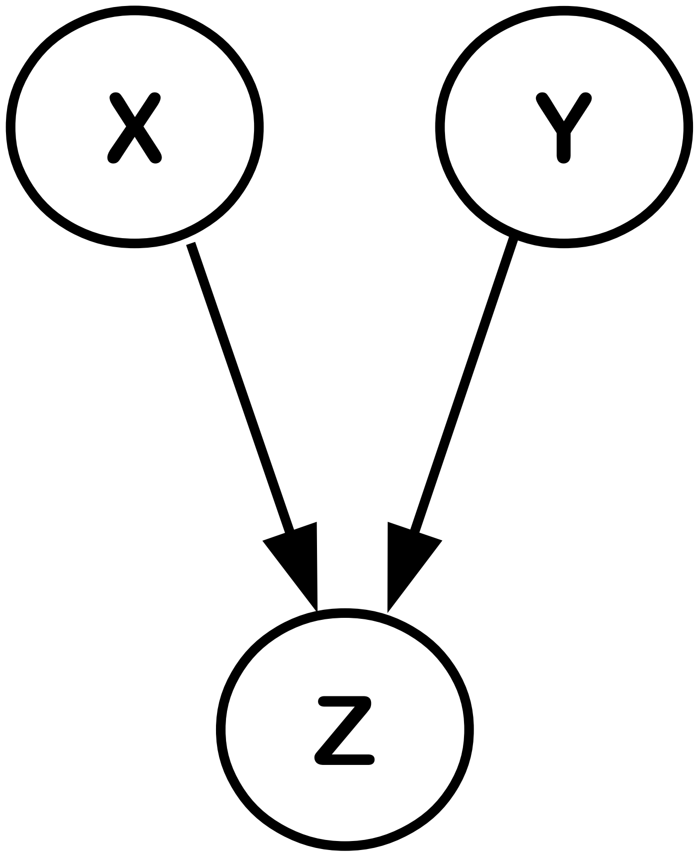
Independent

X and Y are **independent** if

$$\Pr(X = x \text{ and } Y = y) = \Pr(X = x)\,\Pr(Y = y)$$

for all x and y.

Joint distributions = random programs




```
X = flip()
Y = flip()
Z = X + Z
```

Joint distributions = random programs

```
X = 0, Y = 0, Z = 0
X = 0, Y = 0, Z = 0
X = 1, Y = 0, Z = 1
X = 1, Y = 0, Z = 1
X = 1, Y = 1, Z = 2
X = 1, Y = 0, Z = 1
X = 1, Y = 1, Z = 2
X = 0, Y = 0, Z = 0
X = 1, Y = 0, Z = 1
X = 1, Y = 1, Z = 2
X = 0, Y = 0, Z = 0
X = 0, Y = 1, Z = 1
X = 1, Y = 1, Z = 2
 ...    ...    ...
```

```
X = flip()
Y = flip()
Z = X + Z
```

Joint distributions = random programs

X	Y	Z	Pr
0	0	0	1/4
0	0	1	0
0	0	2	0
0	1	0	0
0	1	1	1/4
0	1	2	0
1	0	0	0
1	0	1	1/4
1	0	2	0
1	1	0	0
1	1	1	0
1	1	2	1/4

```
X = flip()
Y = flip()
Z = X + Z
```

Marginalization

	X = 0	X = 1	X = 2	Σ
Y = 0	0.1	0.1	0.0	0.2
Y = 1	0.2	0.2	0.1	0.5
Y = 2	0.0	0.3	0.0	0.3
Σ	0.3	0.6	0.1	1.00

$$\Pr(X = x) == \Sigma_y \Pr(X = x \text{ and } Y = y)$$

Shorthand: $\Pr(x) == \Sigma_y \Pr(x, y)$

Joint distributions = random programs

X	Y	Z	Pr
0	0	0	1/4
0	0	1	0
0	0	2	0
0	1	0	0
0	1	1	1/4
0	1	2	0
1	0	0	0
1	0	1	1/4
1	0	2	0
1	1	0	0
1	1	1	0
1	1	2	1/4

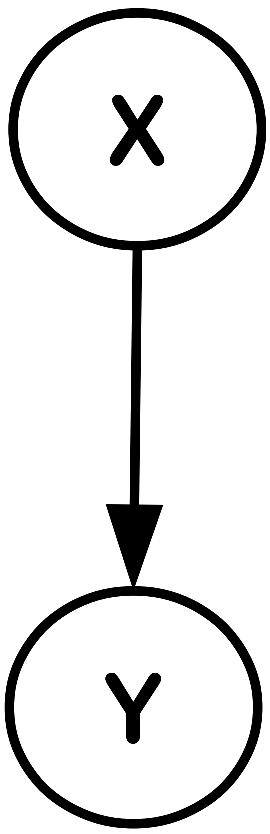
```
X = flip()
Y = flip()
Z = X + Z
```

$$\Pr(Z = 0) = 1/4$$

$$\Pr(Z = 1) = 1/2$$

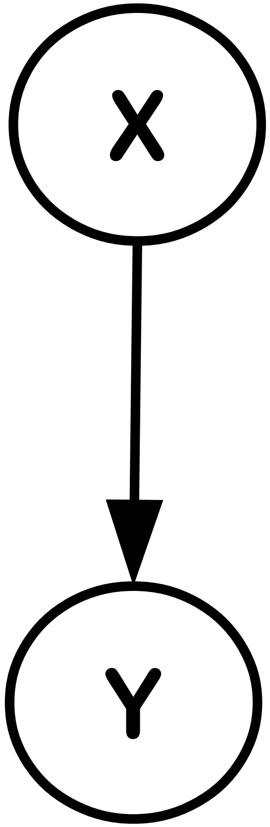
$$\Pr(Z = 2) = 1/4$$

Joint distributions = random programs



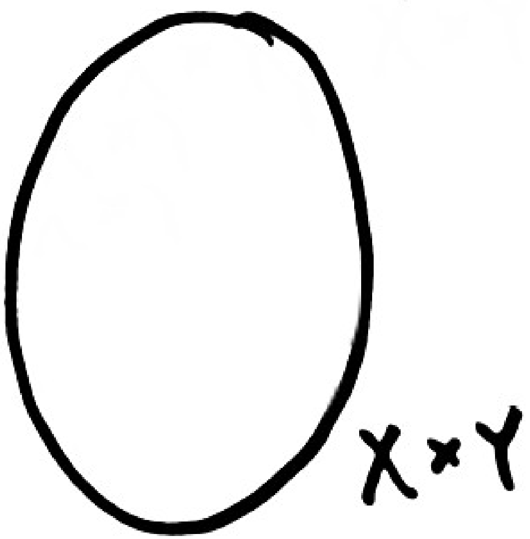

```
X = randint(0, 6)
Y = (X - 3)^2
```

Joint distributions = random programs




```
X = 1
while flip():
     X = X + 1
Y = randint(1, X)
```

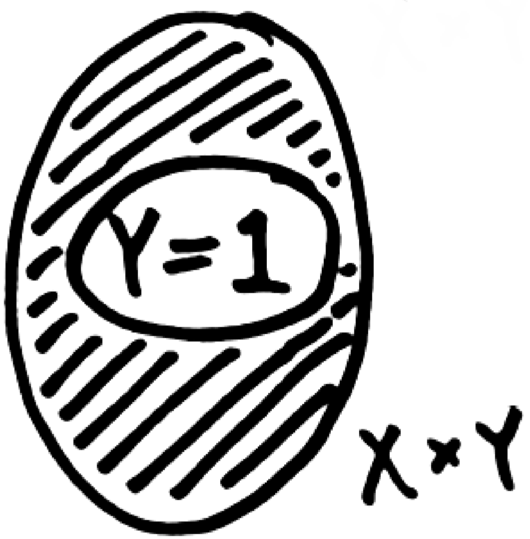
Conditional probability



	X = 0	X = 1	X = 2	Σ
Y = 0	0.1	0.1	0.0	0.2
Y = 1	0.2	0.2	0.1	0.5
Y = 2	0.0	0.3	0.0	0.3
Σ	0.3	0.6	0.1	1.00

$$\Pr(X = x \text{ and } Y = y)$$

Conditional probability



	X = 0	X = 1	X = 2	Σ
Y = 0	0	0	0	0
Y = 1	0.2	0.2	0.1	0.5
Y = 2	0	0	0	0
Σ	0.2	0.1	0.1	0.4

$$\Pr(X = x \text{ and } Y = 1)$$

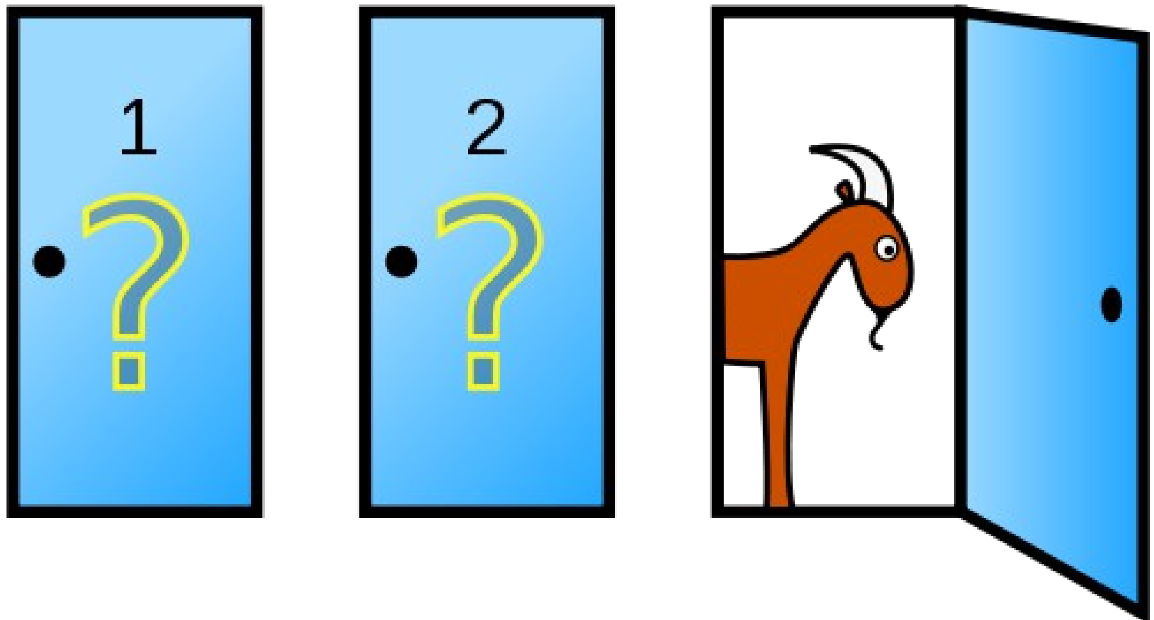
Conditional probability

Y=1 X

	X = 0	X = 1	X = 2	Σ
Y = 0	0	0	0	0
Y = 1	0.4	0.4	0.2	1.0
Y = 2	0	0	0	0
Σ	0.4	0.4	0.2	1.0

$$\Pr(X = x \mid Y = 1) \;==\; \frac{\Pr(X = x \text{ and } Y = 1)}{\Pr(Y = 1)}$$

Example: The Monty Hall Problem



$$\Pr(X = 1) == ?$$

$$\Pr(X = 1 \mid \neg(X = 3)) == ?$$

(en.wikipedia.org/wiki/Monty_Hall_problem)

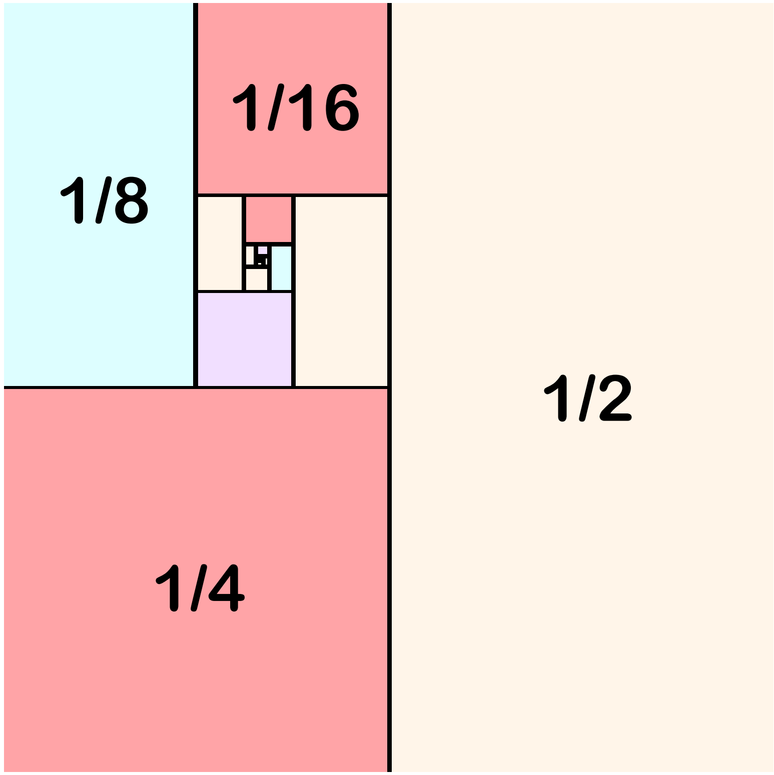
Example: A geometric distribution

x	1	2	2	...
$\Pr(X = x)$	1/2	1/4	1/8	...

$\Pr(\,T \mid E\,)$, where

$T = X$ is divisible by 3

$E = X$ is divisible by 2

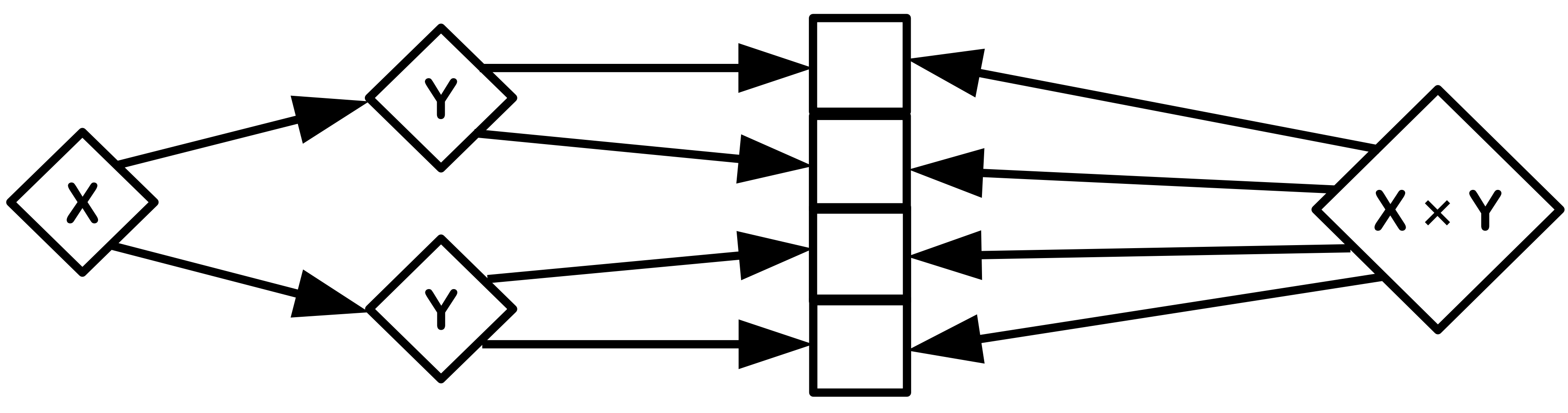


Conditional probability: The Chain Rule

$$\Pr(X \text{ and } Y) == \Pr(X \mid Y)\,\Pr(Y)$$

Verbose form: For all x and y,

$$\Pr(X = x \text{ and } Y = y) == \Pr(X = x \mid Y = y)\,\Pr(Y = y)$$



Distributions
Random variables
Independence

Marginalization
Conditionalization

The chain rule

Now:
Exercises

Colorblindness Colorblindness is caused by a genetic defect which is present on approximately 8% of all X chromosomes. Since men only have one X chromosome, about 8% of the male population is colorblind.

1. Women have two X chromosomes. What percentage of the female population is colorblind? (You can check your answer against the actual figures.)
2. Suppose that the genetic defect occurred more frequently than 8% of the time. How common would it have to be in order for 50% of the female population to be colorblind?

Forwards and backwards prediction Consider the following two tasks:

- Guessing the next letter of a text given the preceding ones:

```
... re particularly impr_
```

- Guessing the previous letter of a text given the following ones:

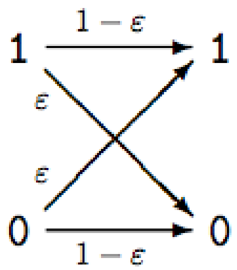
```
_onth following the c ...
```

In general, which task is the more difficult — from statistical perspective, and from a cognitive? Why?

Chinese whisper The binary symmetric channel is a communication channel which transmits 0s and 1s, but occasionally outputs the wrong symbol.

Suppose we have a binary symmetric channel with error probability 0.05, and that we send the string $X = 0000$ through this channel. We then send the output Y back through the channel again, ending up with a third string Z.

What's the probability that $X = Z$?



Coin flipping You generate two sequences by flipping a coin three times. What's the probability that the two sequences are identical? What's the probability that the have the same number of heads?