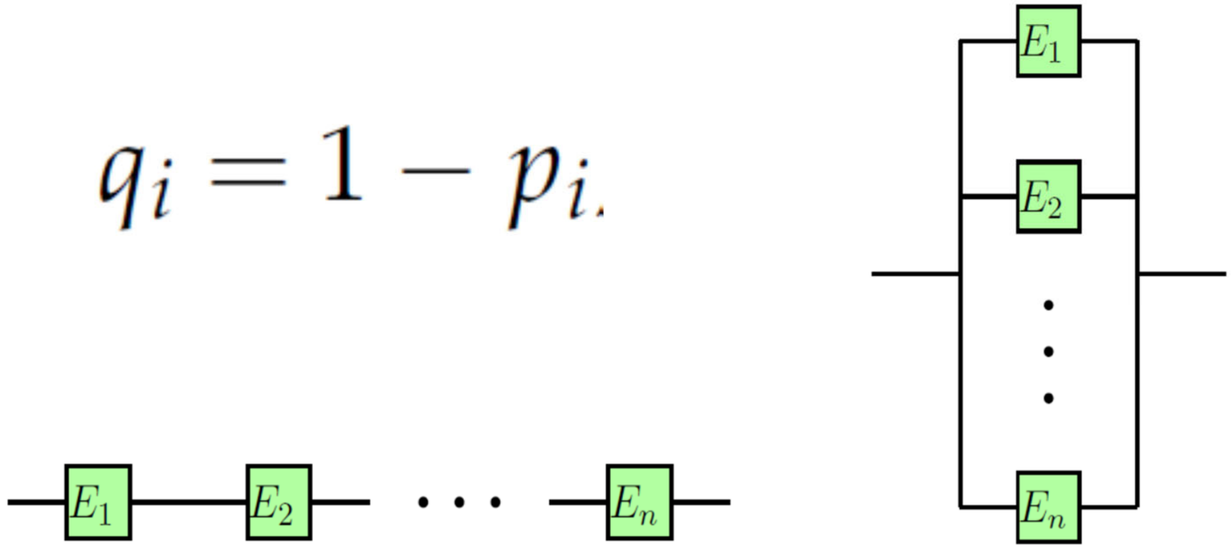
Duality between parallel and series circuits

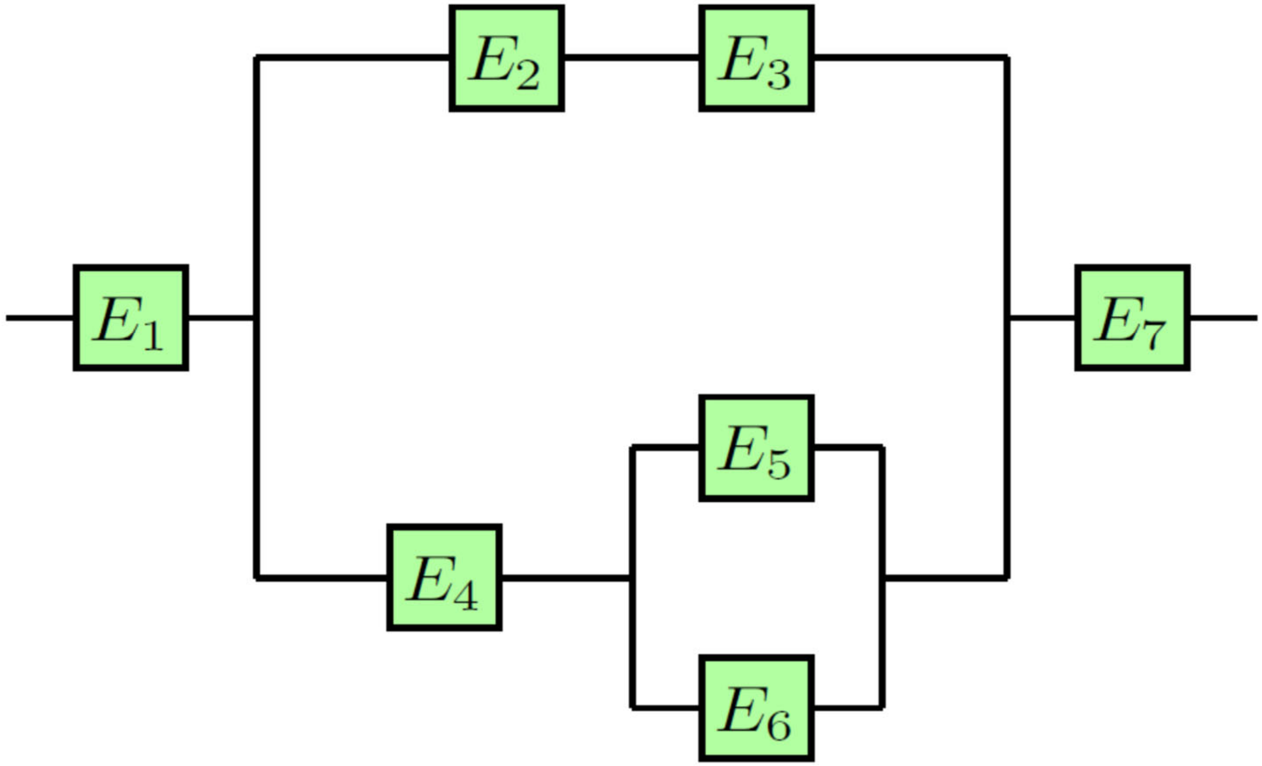
$$q_i = 1 - p_i.$$



(a) (b)

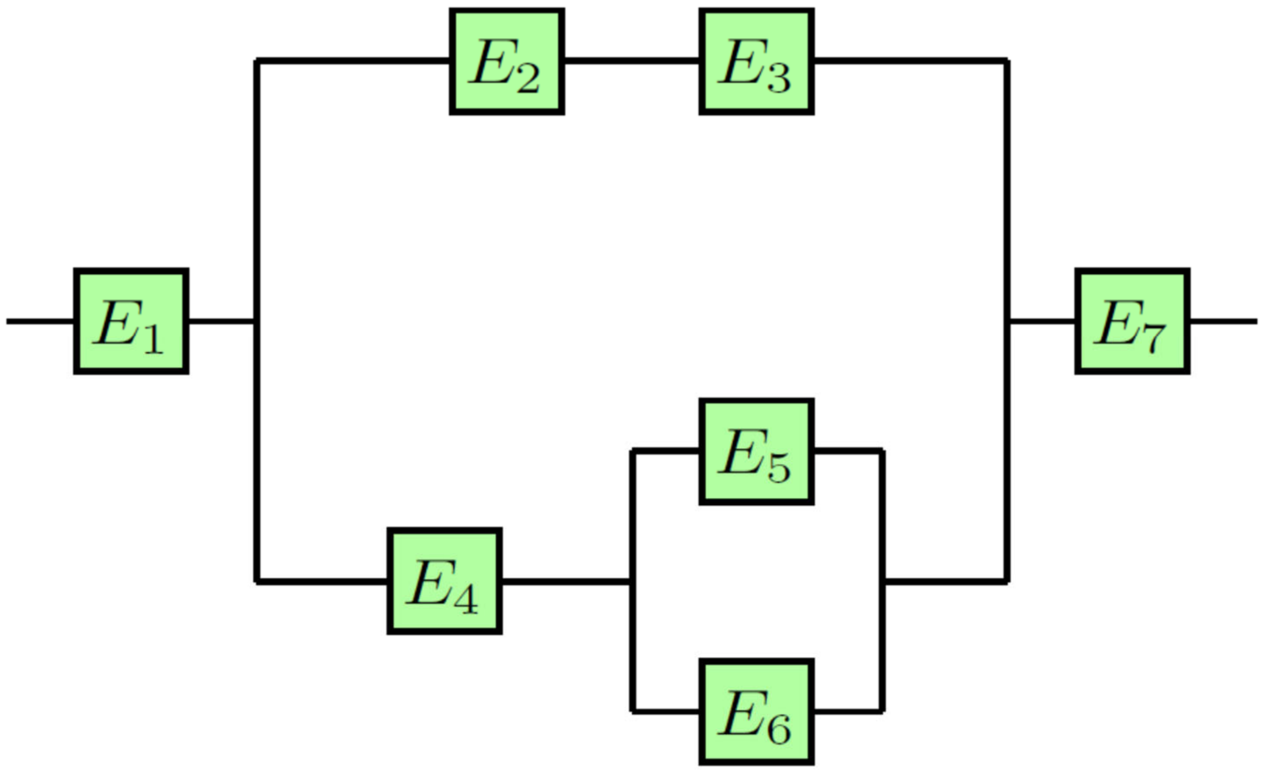
Connection	Notation	Works with prob	Fails with prob
Serial	$E_1 \cap E_2 \cap \cdots \cap E_n$	$p_1 p_2 \ldots p_n$	$1 - p_1 p_2 \ldots p_n$
Parallel	$E_1 \cup E_2 \cup \cdots \cup E_n$	$1 - q_1 q_2 \ldots q_n$	$q_1 q_2 \ldots q_n$

Circuit → Set equation



Component	E_1	E_2	E_3	E_4	E_5	E_6	E_7
Probability of functioning well	0.9	0.5	0.3	0.1	0.4	0.5	0.8

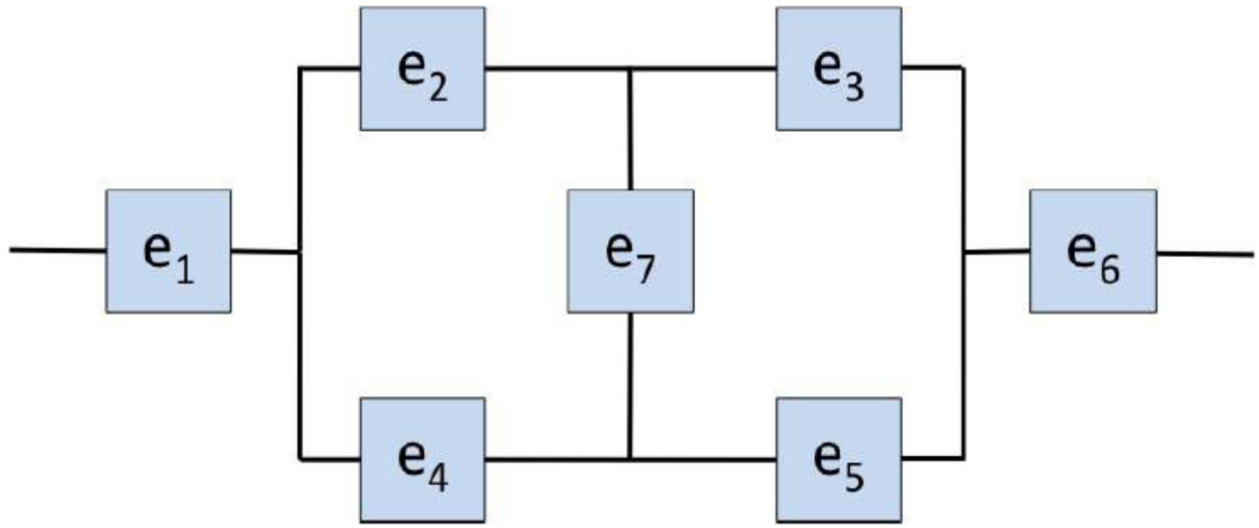
Circuit → Set equation



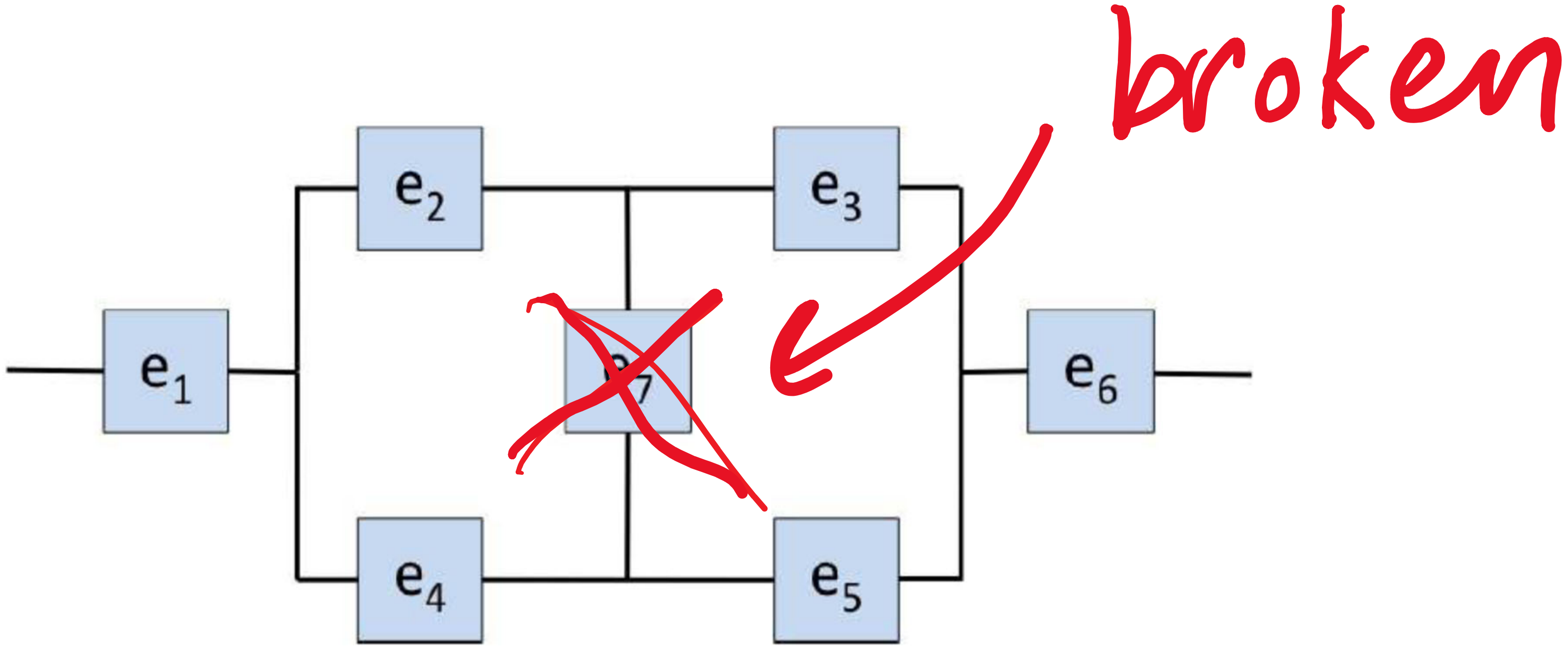
Component	E_1	E_2	E_3	E_4	E_5	E_6	E_7
Probability of functioning well	0.9	0.5	0.3	0.1	0.4	0.5	0.8

P(Works) = 0.9.*(1-(1-0.5.*0.3).*(1-0.1.*(1-0.6.*0.5))).*0.8=0.15084

Let’s check it in Matlab
Simulate 1 million random circuits
and count the fraction that work



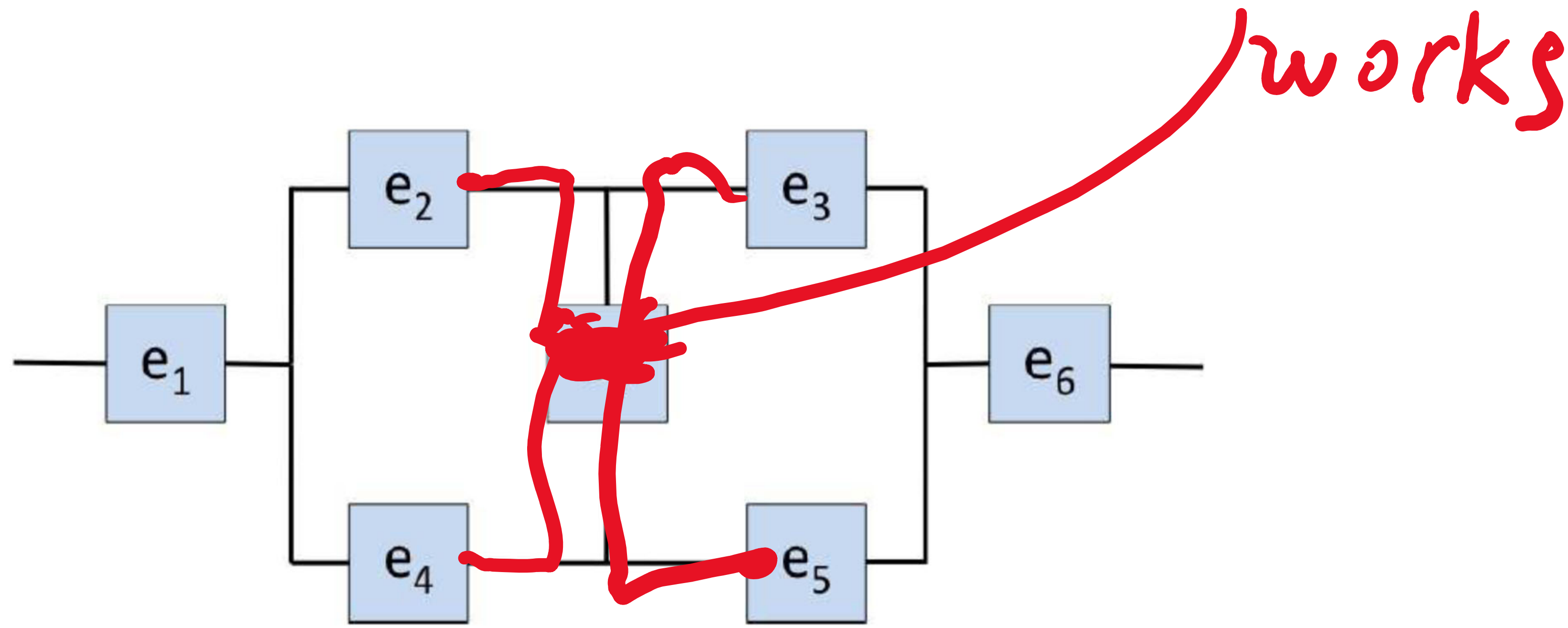
Component	e_1	e_2	e_3	e_4	e_5	e_6	e_7
Probability of component working	0.3	0.8	0.2	0.2	0.5	0.6	0.4



Component	e_1	e_2	e_3	e_4	e_5	e_6	e_7
Probability of component working	0.3	0.8	0.2	0.2	0.5	0.6	0.4

P(circuit works |e7 is broken)=P(e1 works)*
[1-(1-P(e2 works)*P(e3 works))*(1-P(e4 works)*P(e5 works))]*
P(e6 works)=0.3*(1-(1-0.8*0.2)*(1-0.2*0.5))*0.6=0.0439

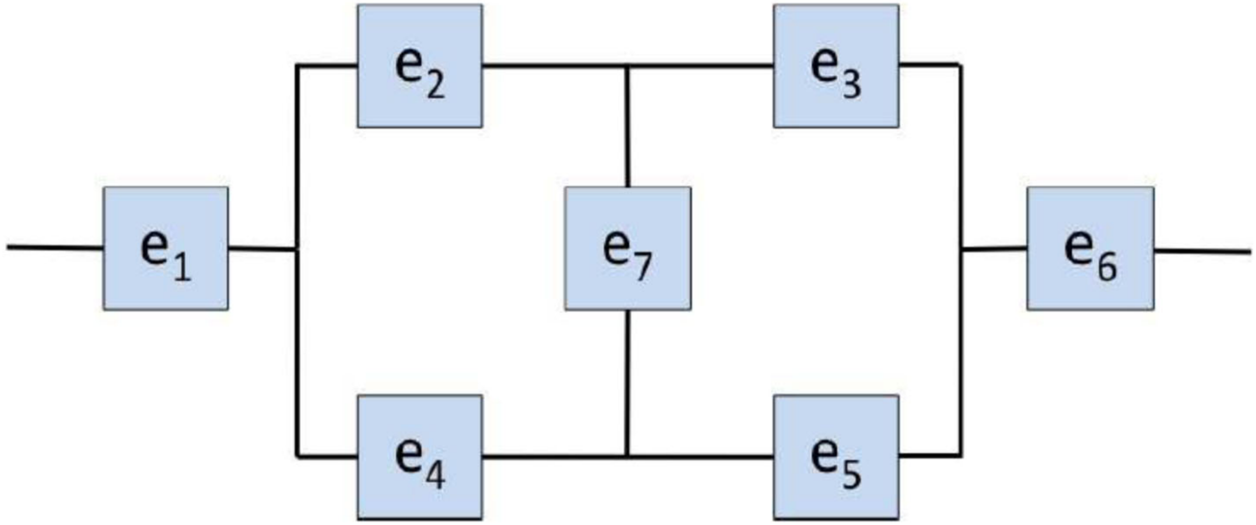
The contribution to total probability:
P(circuit works |e7 is broken)*P(e7 is broken)=0.6*0.0439=0.0264



Component	e_1	e_2	e_3	e_4	e_5	e_6	e_7
Probability of component working	0.3	0.8	0.2	0.2	0.5	0.6	0.4

P(circuit works |e7 works)=P(e1 works)*
[1-(1-P(e2 works))*(1-P(e4 works))]
[1-(1-P(e3 works))(1-P(e5 works))]*
P(e6 works)=0.3*(1-(1-0.8)*(1-0.2))*(1-(1-0.2)*(1-0.5)))*0.6=0.0907

The contribution to total probability:
P(circuit works |e7 works)*P(e7 works)=0.4*0.0907=0.0363

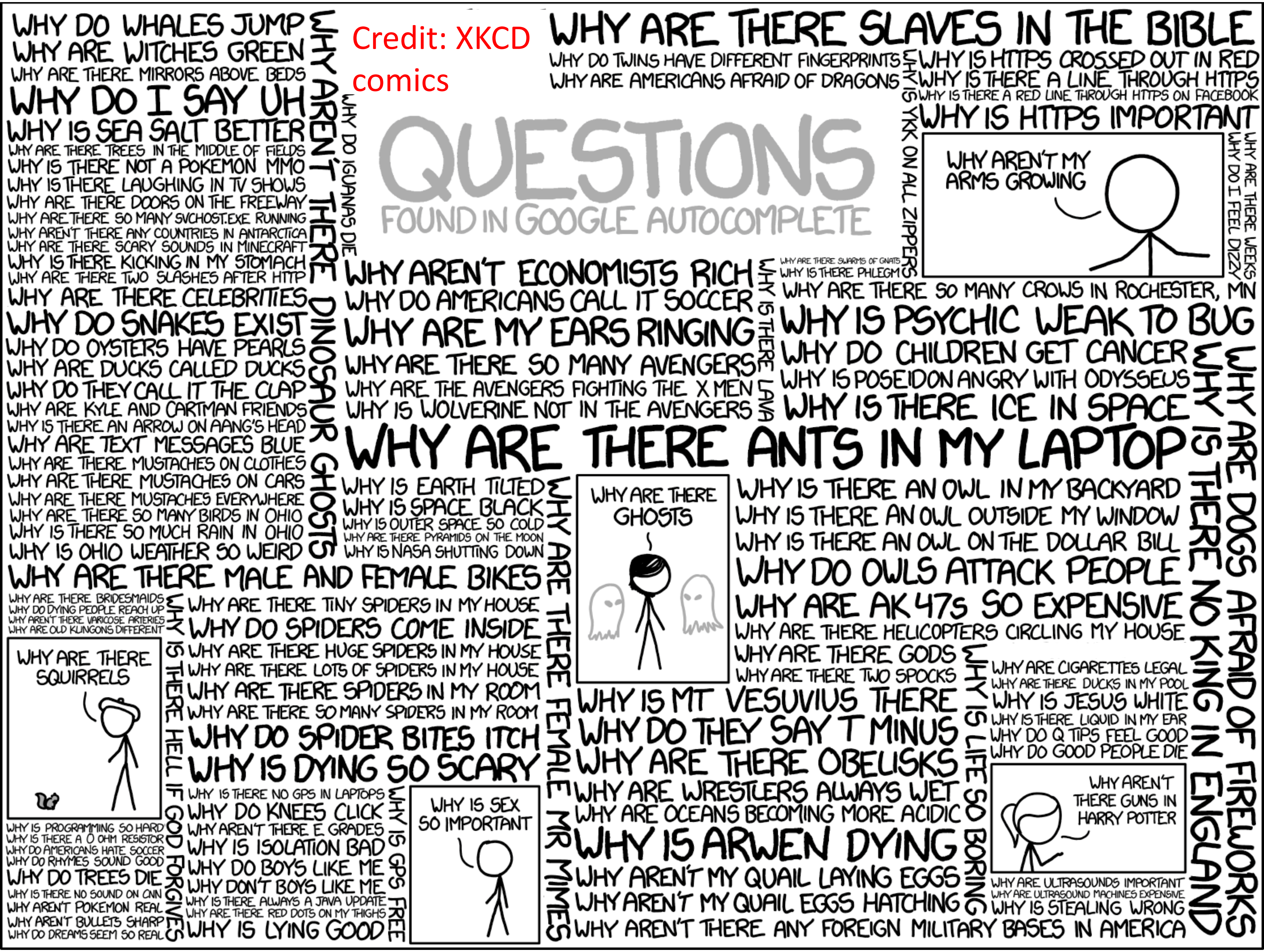


Component	e_1	e_2	e_3	e_4	e_5	e_6	e_7
Probability of component working	0.3	0.8	0.2	0.2	0.5	0.6	0.4

P(circuit works)=
P(circuit works |e7 works)*P(e7 works)+
P(circuit works |e7 is broken)*P(e7 is broken)=
=0.0264+0.0363=0.0627

Answer: 6.27%

Credit: XKCD comics



Discrete Probability Distributions

Random Variables

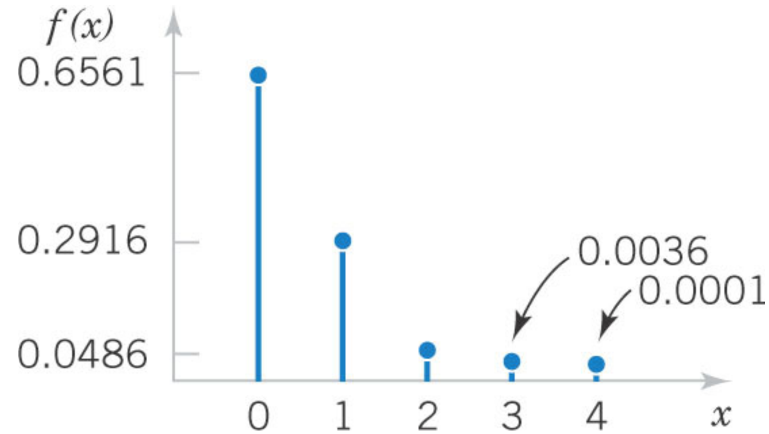
- A variable that associates a number with the outcome of a random experiment is called a random variable.
- Notation: random variable is denoted by an uppercase letter, such as X. After the experiment is conducted, the measured value is denoted by a lowercase letter, such a x. Both X and x are shown in italics, e.g., $P(X=x)$.

Continuous & Discrete Random Variables

- A discrete random variable is usually an integer number
  - N - the number of p53 proteins in a cell
  - D - the number of nucleotides different between two gene sequences of length L
- A continuous random variable is a real number
  - C=N/V – the concentration of p53 protein in a cell of volume V
  - Percentage (D/L)*100% of different nucleotides in gene sequences of different lengths L
    (depending on the set of L's it may be discrete but dense)

Probability Mass Function (PMF)

- I want to compare all <u>4-mers</u> in a pair of human genomes
- X – random variable: the number of nucleotide differences in a given 4-mer
- Probability Mass Function: f(x) or P(X=x) – the probability that the # of SNPs is exactly equal to x



Probability Mass Function for the # of mismatches in 4-mers

$P(X=0) =$	0.6561
$P(X=1) =$	0.2916
$P(X=2) =$	0.0486
$P(X=3) =$	0.0036
$P(X=4) =$	0.0001
$\sum_x P(X=x) =$	1.0000

Cumulative Distribution Function (CDF)

x	$P(X=x)$	CDF: $P(X\leq x)$	CCDF: $P(X>x)$
-1	0.0000	0.0000	1.0000
0	0.6561	0.6561	0.3439
1	0.2916	0.9477	0.0523
2	0.0486	0.9963	0.0037
3	0.0036	0.9999	0.0001
4	0.0001	1.0000	0.0000

Cumulative Distribution Function CDF: $F(x)=P(X\leq x)$

Example:

$F(3)=P(X \leq 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3) = 0.9999$

Complementary Cumulative Distribution Function (tail distribution) or CCDF: $F_{>}(x)=P(X>x)$

Example: $F_{>}(0)= P(X>0) = 1- P(X \leq 0) = 1-0.6561=0.3439$

Mean or Expected Value of X

The mean or expected value of the discrete random variable X, denoted as μ or $E(X)$, is

$$\mu = E(X) = \sum_x x \cdot P(X = x) = \sum_x x \cdot f(x)$$

- The mean = the weighted average of all possible values of *X*. It represents its "center of mass"

- The mean may, or may not, be an allowed value of X

- It is also called the arithmetic mean (to distinguish from e.g. the geometric mean discussed later)

- Mean may be infinite if X any integer and tail $P(X=x)>c/x^2$

Outcomes of 6 random experiments

0, 1, 0, 0, 2, 1

$$\text{Mean} = \frac{0+1+0+0+2+1}{6} =$$

$$= \frac{3\times 0 + 2\times 1 + 1\times 2}{6} =$$

$$= 0 \times \frac{3}{6} + 1 \times \frac{2}{6} + 2 \times \frac{1}{6} = \sum_{x=0}^{2} x P(X=x)$$

- $$E[X] = \sum_{x} x \cdot P(X=x)$$
- $$E[X^2] = \sum x^2 \cdot P(X=x)$$
- $$E[a \cdot X + b \cdot X^2] = \sum (ax + bx^2) \times P(X=x) = a \cdot \sum x P(X=x) + b \sum x^2 P(X=x)$$
- $$E[e^X] = \sum e^x P(X=x)$$

Variance $V(X)$: square of a typical deviation from the mean $\mu = E(X)$

$V(X) = \sigma^2$, where σ is called standard deviation

$$\sigma^2 = V(X) = E((X-\mu)^2) =$$

$$= E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2 =$$

$$= E(X^2) - \mu^2 = E(X^2) - (E(X))^2$$

Variance of a Random Variable

If X is a discrete random variable with probability mass function $f(x)$,

$$E\left[h(X)\right]=\sum_x h(x)\cdot P(X=x)=\sum_x h(x)f(x) \qquad (3\text{-}4)$$

If $h(x)=(X-\mu)^2$, then its expectation, $V(x)$, is the variance of X.

$\sigma=\sqrt{V(x)}$, is called standard deviation of X

$$\sigma^2=V(X)=\sum_x (x-\mu)^2 f(x) \text{ is the definitional formula}$$

$$=\sum_x\left(x^2-2\mu x+\mu^2\right)f(x)$$

$$=\sum_x x^2 f(x)-2\mu\sum xf(x)+\mu^2\sum_x f(x)$$

$$=\sum_x x^2 f(x)-2\mu^2+\mu^2$$

$$=\sum_x x^2 f(x)-\mu^2 \text{ is the computational formula}$$

Variance can be infinite
if X can be any integer
and tail of P(X=x) ≥c/x³

Skewness of a random variable

- Want to quantify how asymmetric is the distribution around the mean?
- Need any odd moment: $E[(X-\mu)^{2n+1}]$
- Cannot do it with the first moment: $E[X-\mu]=0$
- Normalized 3-rd moment is skewness: $\gamma_1=E[(X-\mu)^3/\sigma^3]$
- Skewness can be infinite if X takes unbounded positive integer values and the tail $P(X=x) \geq c/x^4$ for large x

Geometric mean of a random variable

- Useful for very broad distributions (many orders of magnitude)?
- Mean may be dominated by very unlikely but very large events. Think of a lottery
- Exponent of the mean of *log X*: *Geometric mean=exp(E[log X])*
- Geometric mean usually is not infinite

Summary: Parameters of a Probability Distribution

- Probability Mass Function (PMF): $f(x)=\text{Prob}(X=x)$
- Cumulative Distribution Function (CDF): $F(x)=\text{Prob}(X \leq x)$
- Complementary Cumulative Distribution Function (CCDF): $F_{>}(x)=\text{Prob}(X>x)$
- The mean, $\mu=E[X]$, is a measure of the center of mass of a random variable
- The variance, $V(X)=E[(X-\mu)^2]$, is a measure of the dispersion of a random variable around its mean
- The standard deviation, $\sigma=[V(X)]^{1/2}$, is another measure of the dispersion around mean. Has the same units as X
- The skewness, $\gamma_1=E[(X-\mu)^3/\sigma^3]$, a measure of asymmetry around mean
- The geometric mean, $\exp(E[\log X])$ is useful for very broad distributions