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

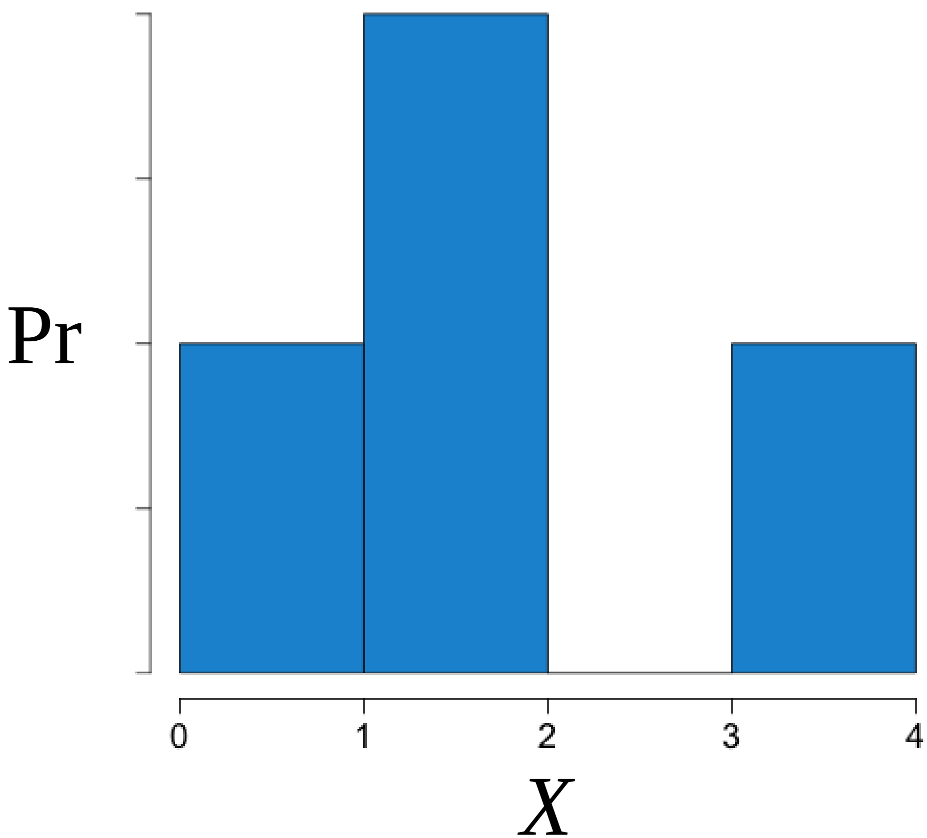
Thursday

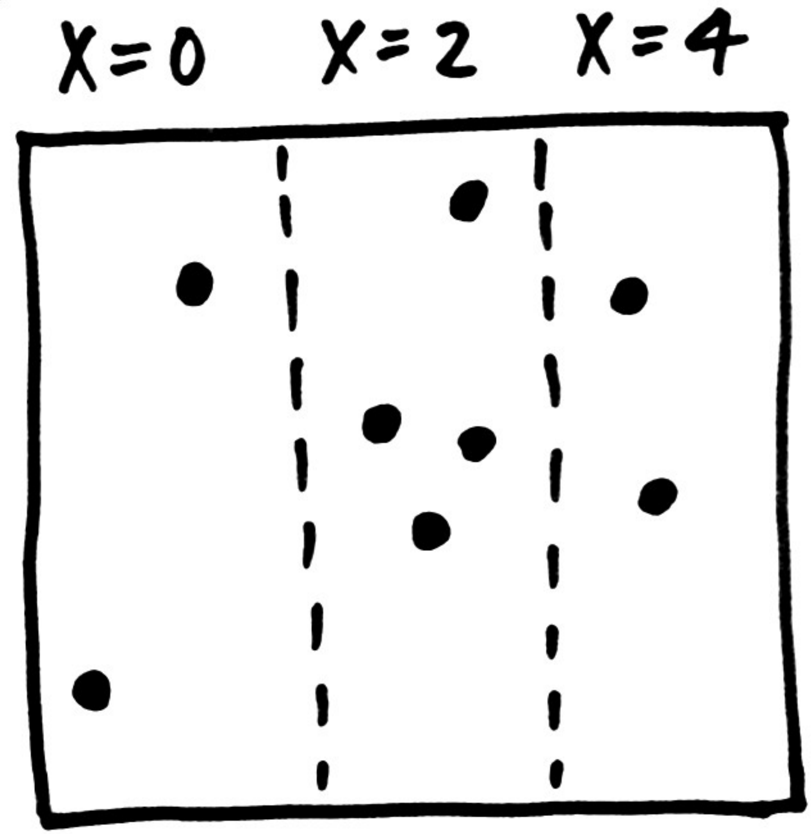
Divergence
Kelly Gambling

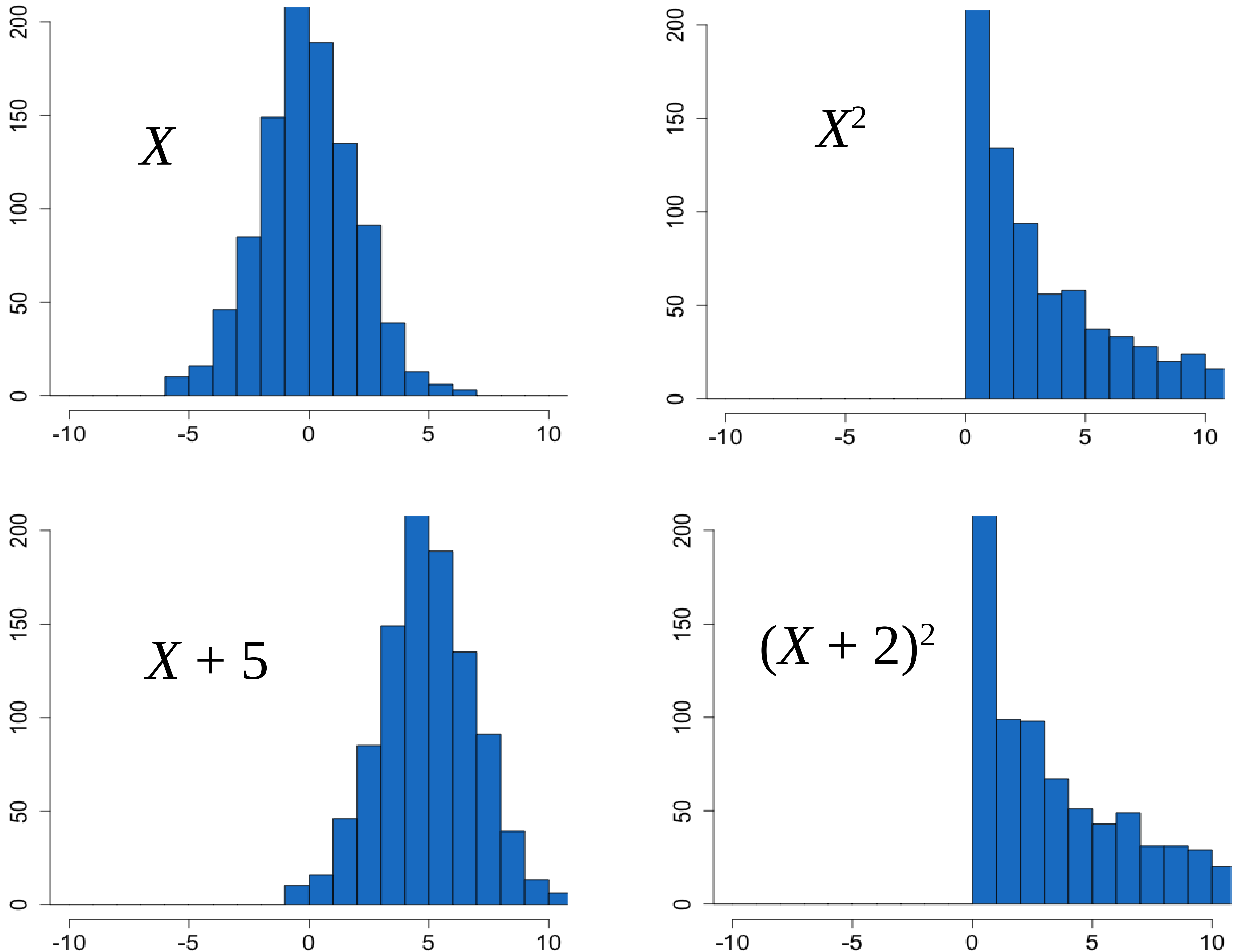
Friday

Kolmogorov Complexity
The limits of statistics

Random Variables





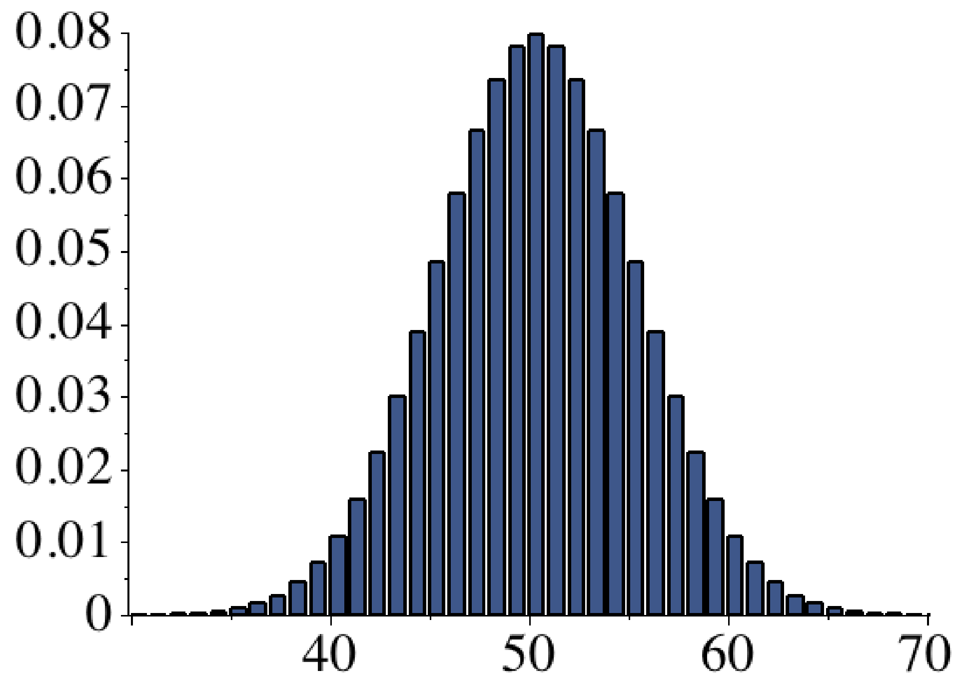

X
X^2
$X + 5$
$(X + 2)^2$
0
50
100
150
200
-10
-5
0
5
10

Variance

Two events:

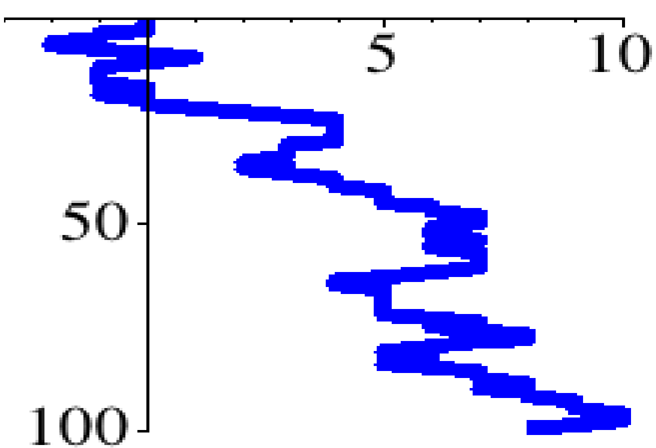
- **I flip a fair coin 100 times, and k = 60.**
- **I flip it 1,000,000 times, and k = 600,000.**

Are these equally probable?



A dust particle takes one step left (1/4), one step right (1/4), or no step (1/2).

After 100 time units, how far away from the starting point would you guess the particle is?



Expected value

The **expected value** of a random variable is the weighted sum (or integral) of its values:

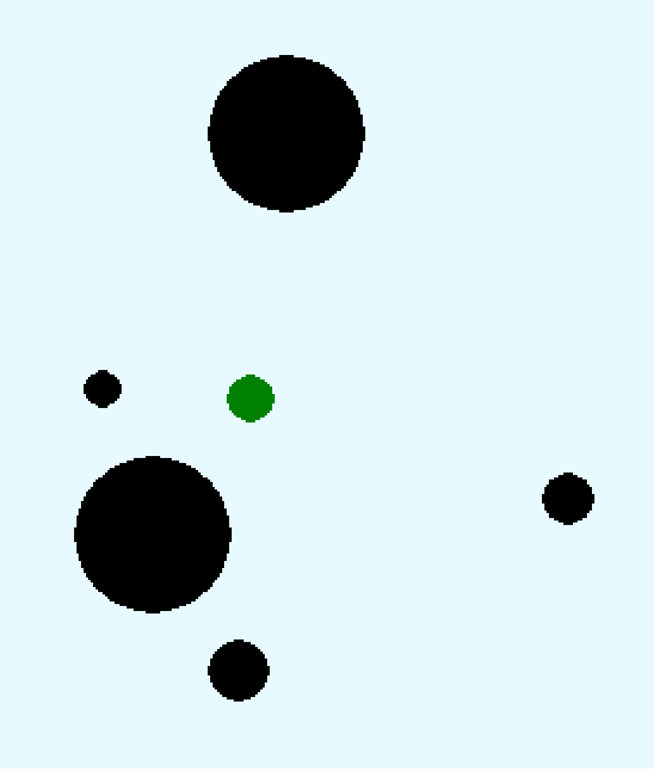
$$E[X] = p(x_1)\, x_1 + p(x_2)\, x_2 + \ldots + p(x_n)\, x_n$$

$$E[cX] = c\,E[X]$$

$$E[X + Y] = E[X] + E[Y]$$

For **independent** X, Y:

$$E[XY] = E[X]E[Y]$$



Variance

w	w_1	w_2	E
X	0	9	?
$\Pr(X)$	2/3	1/3	

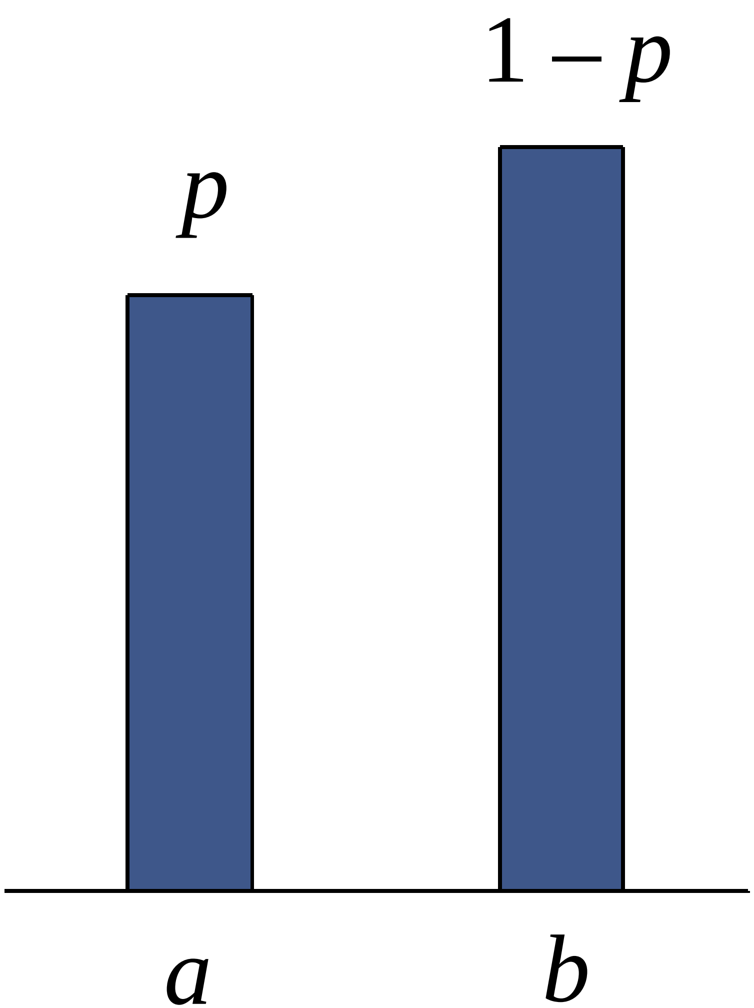
Variance

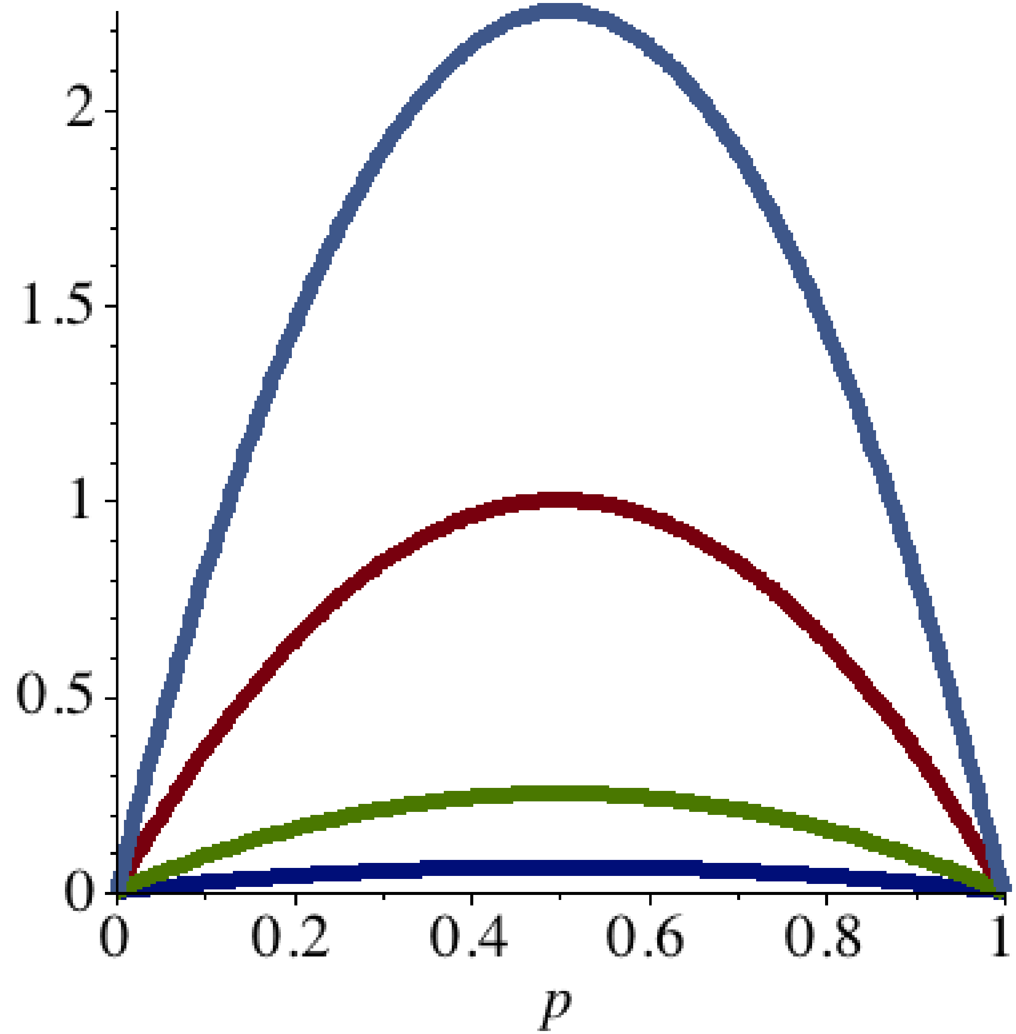
w	w_1	w_2	E
X	0	9	3
$X-3$	?	?	?
$\lvert X-3 \rvert$	?	?	?
$(X-3)^2$	?	?	?
$\Pr(X)$	2/3	1/3	

Variance

w	w_1	w_2	E
X	0	9	3
$X - 3$	–3	6	0
$\lvert X - 3 \rvert$	3	6	4
$(X - 3)^2$	9	36	18
$\Pr(X)$	2/3	1/3	

Variance





$$\mathrm{VAR}[X] == p\,(1-p)\,(b-a)^2$$

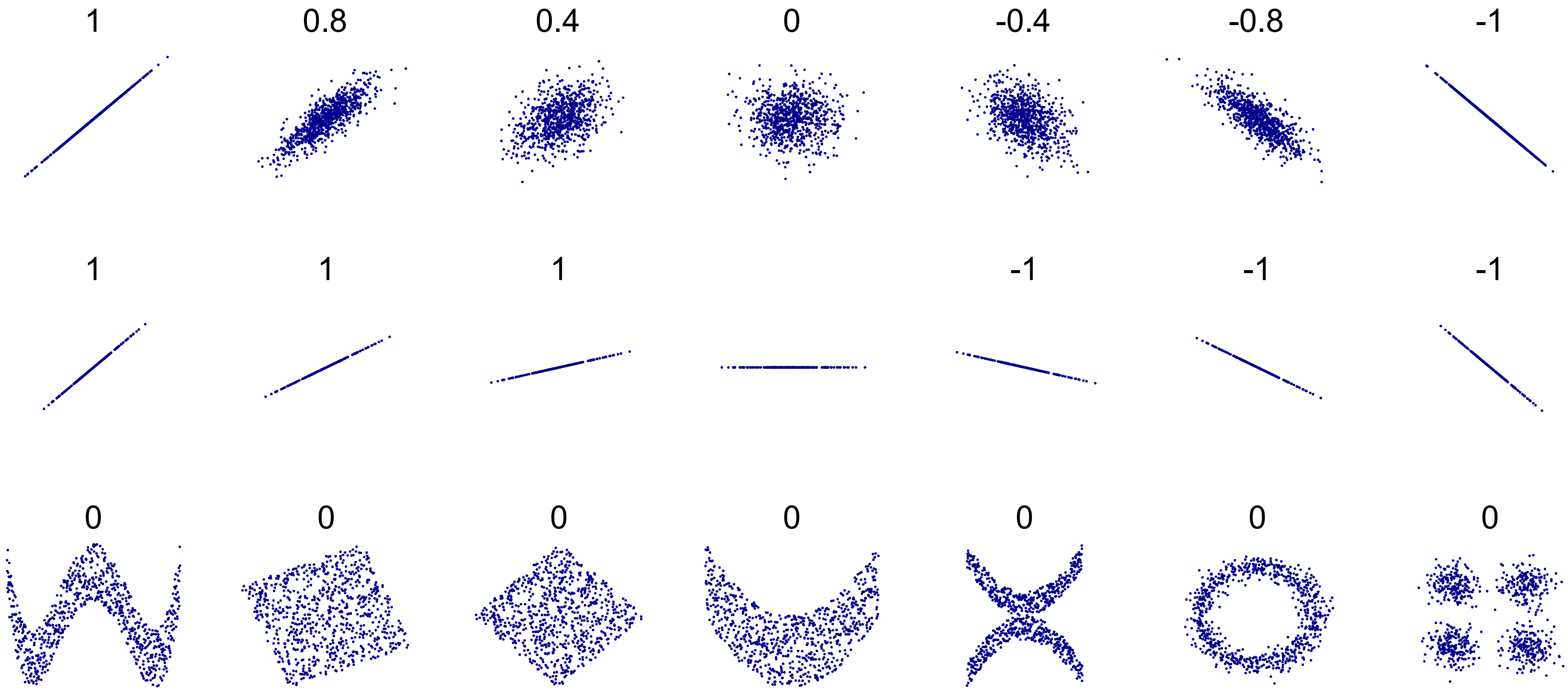
Variance

For any X,

$$\mathrm{VAR}[cX] = c^2\,\mathrm{VAR}[X]$$

For **independent** X and Y,

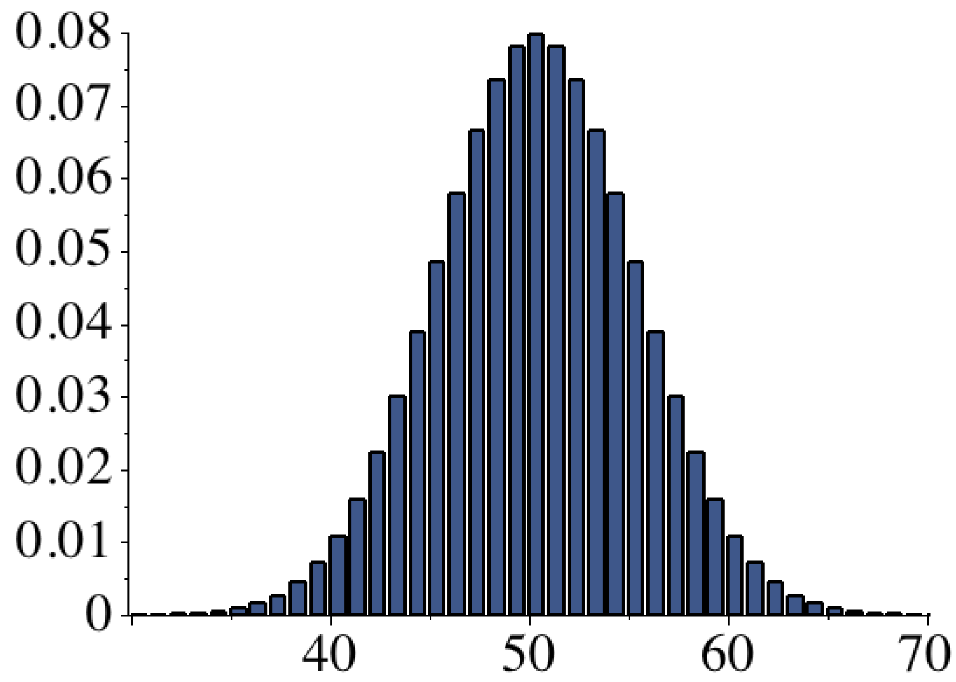
$$\mathrm{VAR}[X + Y] = \mathrm{VAR}[X] + \mathrm{VAR}[Y]$$



Normalized covariances (correlations)

en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient

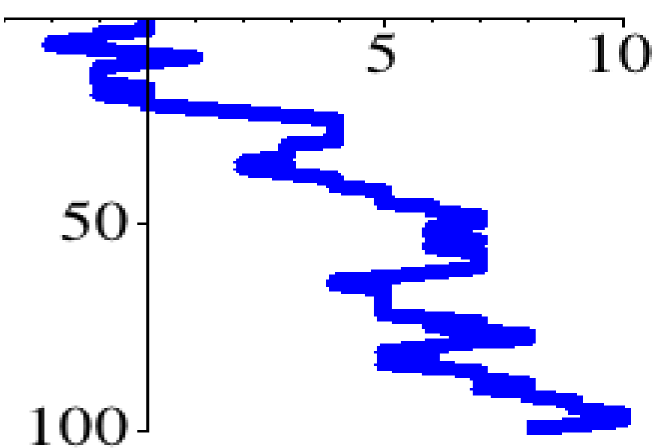
Variance



Two events:

- I flip a fair coin 100 times, and k = 60.
- I flip it 1,000,000 times, and k = 600,000.

Are these equally probable?



A dust particle takes one step left (1/4), one step right (1/4), or no step (1/2).

After 100 time units, how many steps away from the starting point would you guess the particle is?

The Weak Law of Large Numbers

Let X_1, X_2, X_3, ... be a series of **independent** and **identically distributed** random variables with a (common) expected value μ. Let further

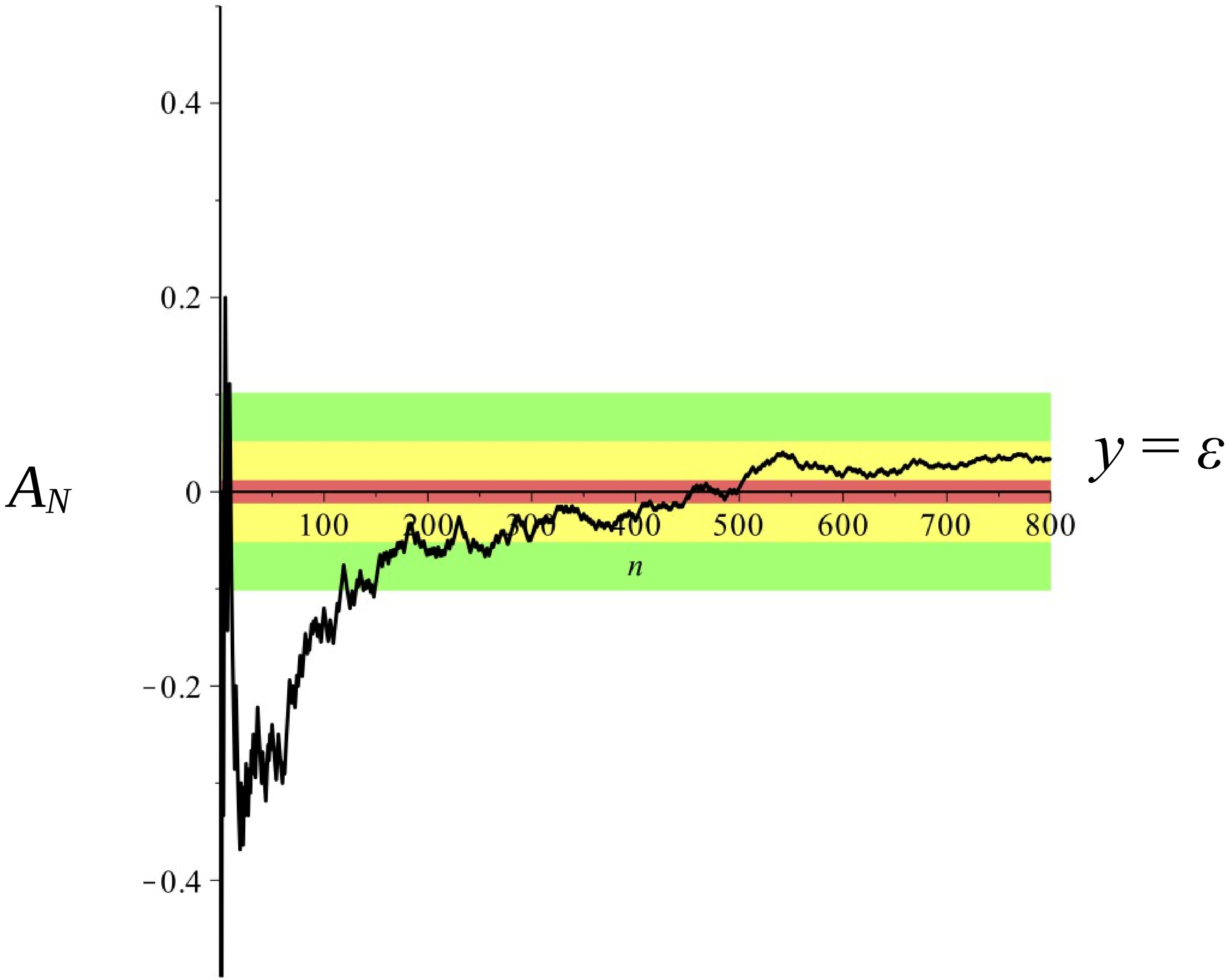
$$A_k = \frac{X_1 + X_2 + X_3 + \ldots + X_k}{k}$$

be the **average** of the first k variables.

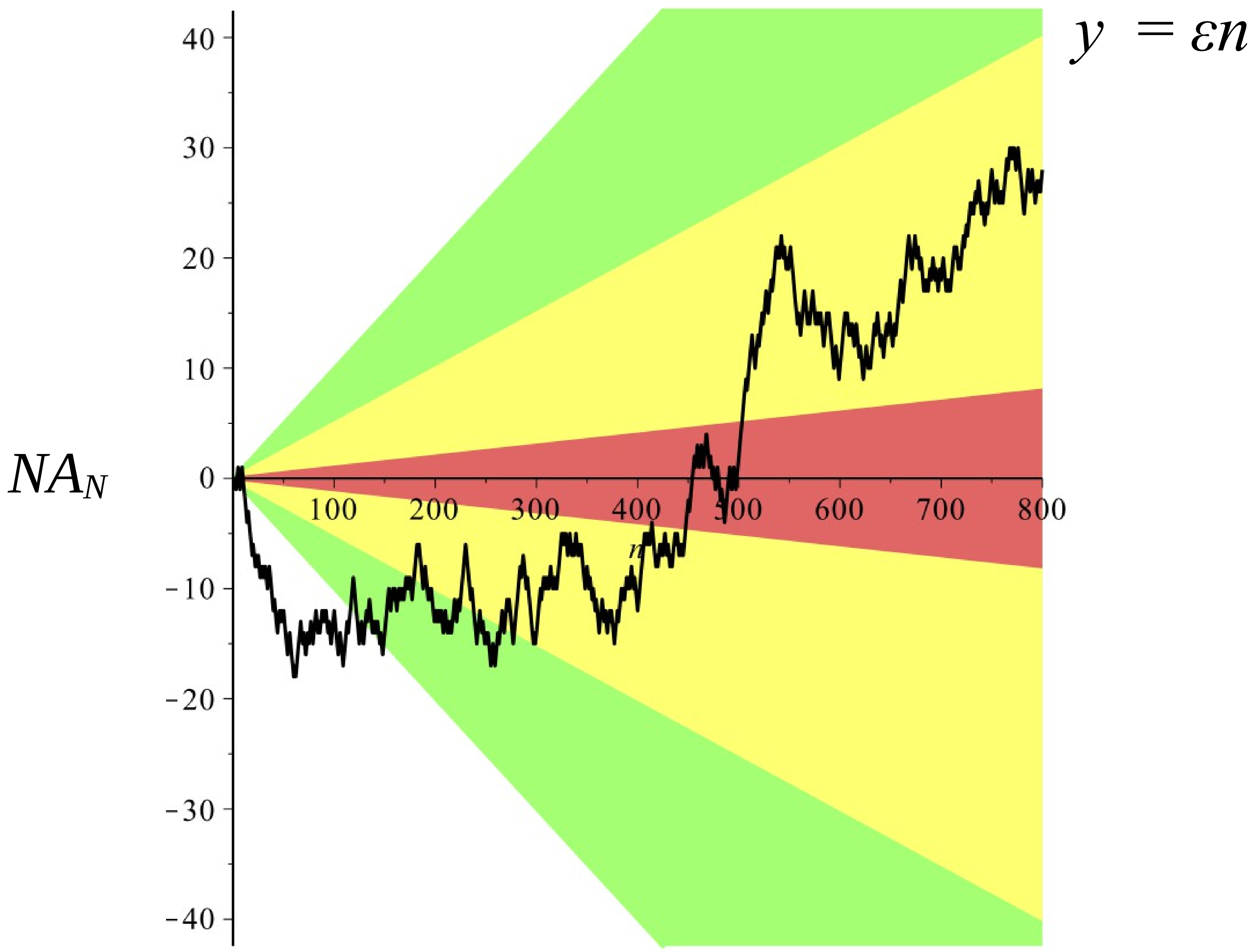
Then for every ε, $\tau > 0$ there exists an N such that

$$\left| A_N - \mu \right| < \varepsilon$$

with probability at least $1 - \tau$.



A_N
0.4
0.2
0
-0.2
-0.4
100
200
300
400
500
600
700
800
n
$y = \varepsilon$



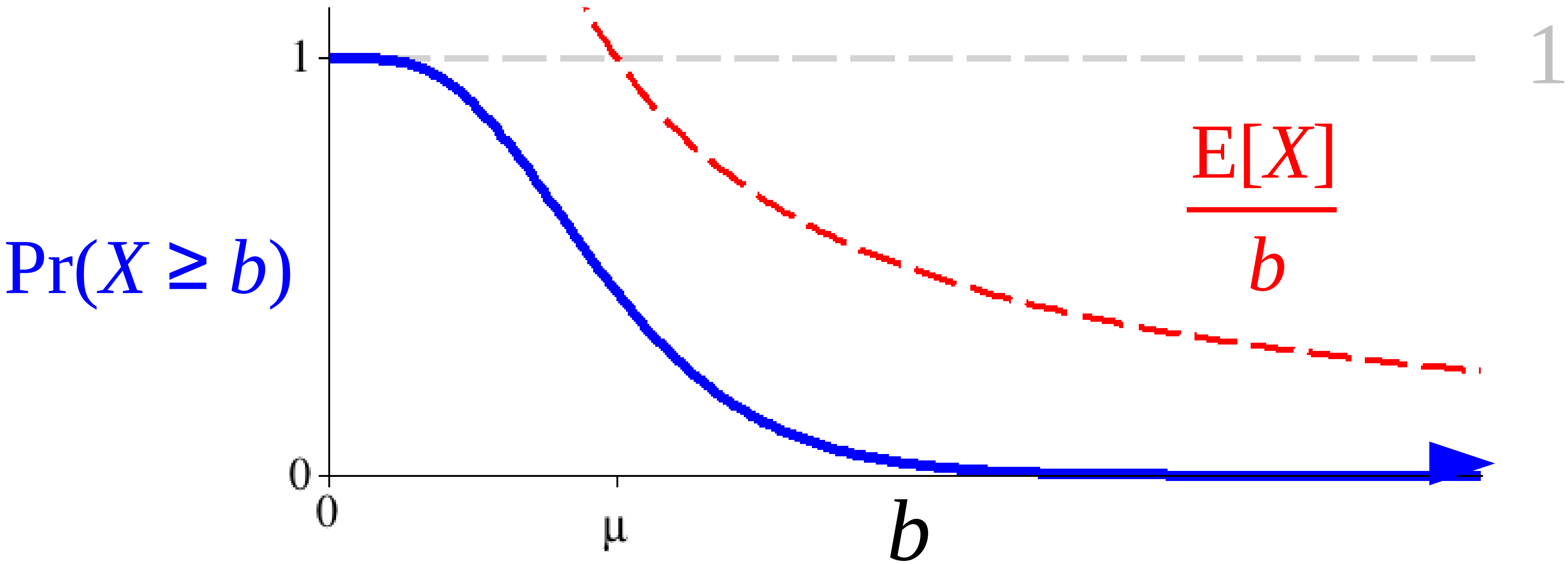
$y = \varepsilon n$
NA_N
40
30
20
10
0
-10
-20
-30
-40
100
200
300
400
500
600
700
800
n

The Markov bound

Suppose X is a **nonnegative** random variable (i.e., Pr(X < 0) = 0) with expected value E[X].

Then for any real number b,

$$\Pr(X \geq b) \leq \frac{\mathrm{E}[X]}{b}$$

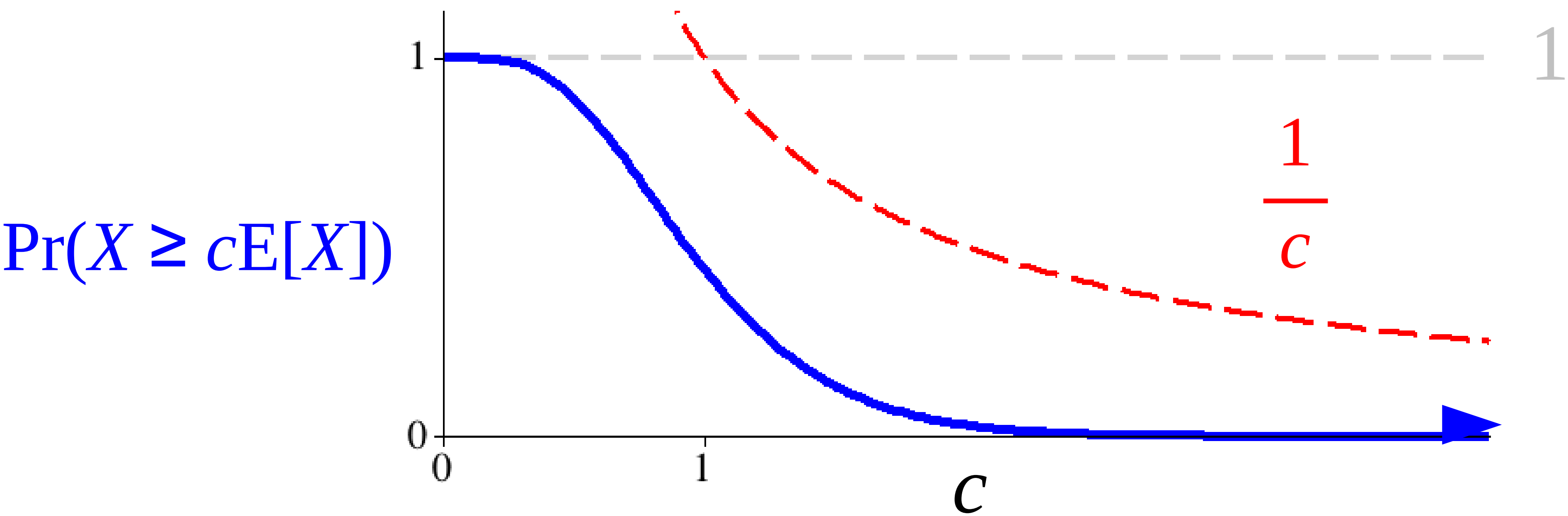


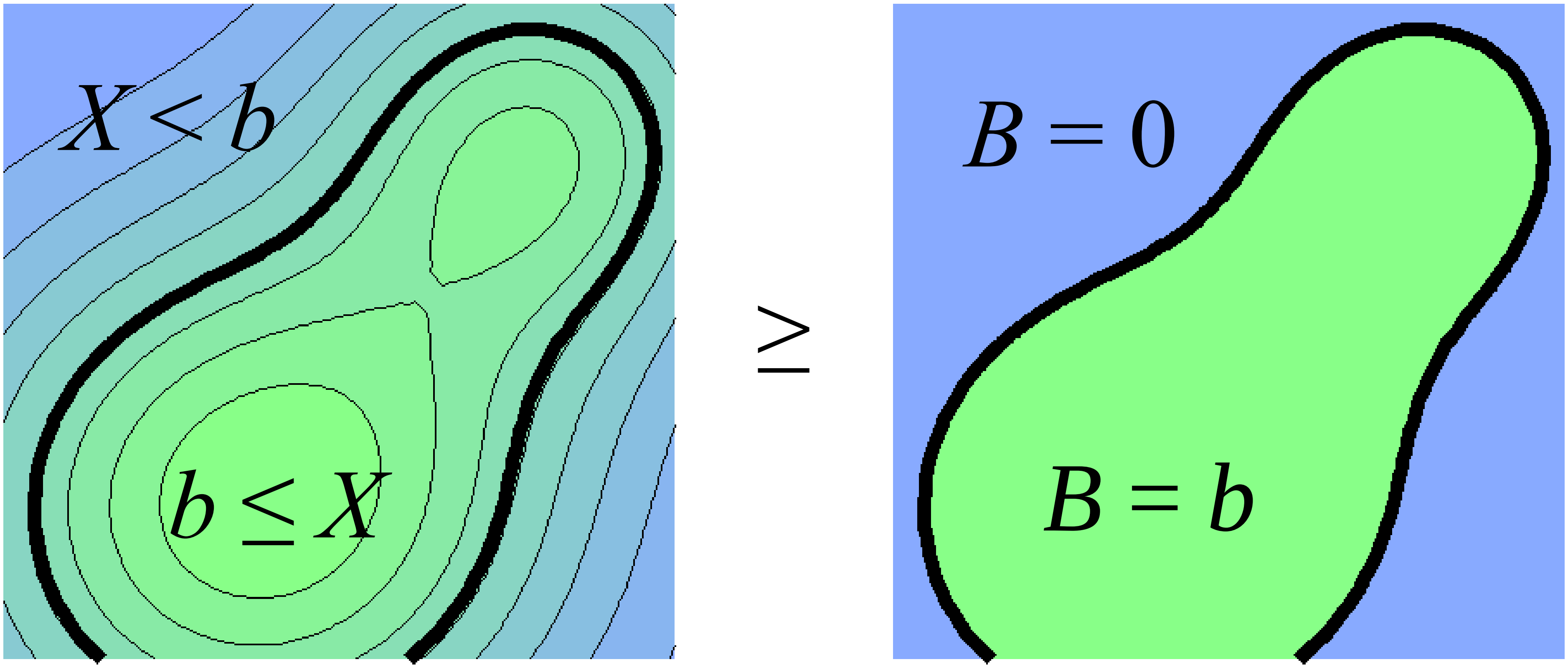
The Markov bound

Suppose X is a **nonnegative** random variable (i.e., Pr(X < 0) = 0) with expected value E[X].

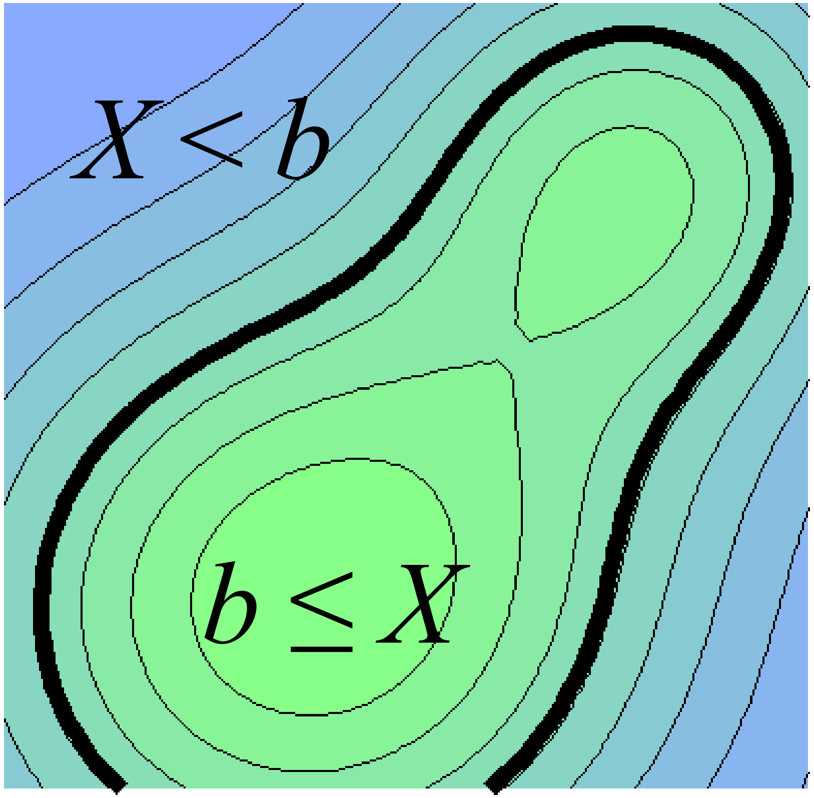
Then for any real number c,

$$\Pr(X \geq c\mathrm{E}[X]) \leq \frac{1}{c}$$

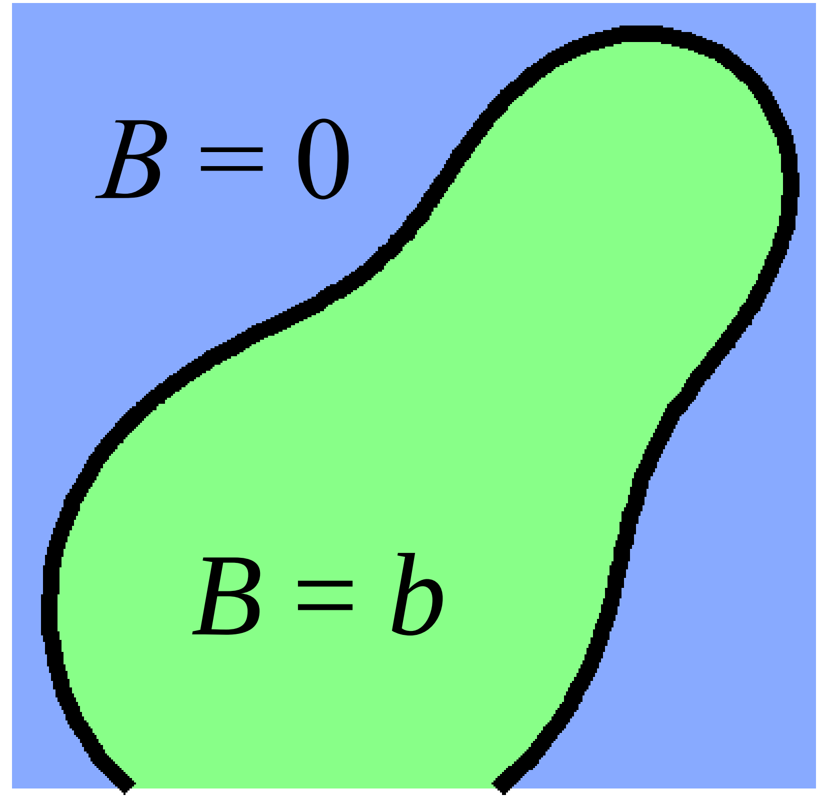




w	w_1	w_2	w_3	w_4	w_5	w_6
X	3	0	7	9	3	8
B	0	0	7	7	0	7



$\geq$



$$\begin{tikzcd}
 & \mathrm{E}[X \mid X \geq b] & (\geq b) \\
\mathrm{E}[X] \arrow[ur, "\Pr(X \geq b)"] \arrow[dr, "\Pr(X < b)"'] & & \\
 & \mathrm{E}[X \mid X < b] & (\geq 0)
\end{tikzcd}$$

The Weak Law of Large Numbers

Let X_1, X_2, X_3, ... be a series of **independent** and **identically distributed** random variables with a (common) expected value μ. Let further

$$A_k = \frac{X_1 + X_2 + X_3 + \ldots + X_k}{k}$$

be the **average** of the first k variables.

Then for every ε, $\tau > 0$ there exists an N such that

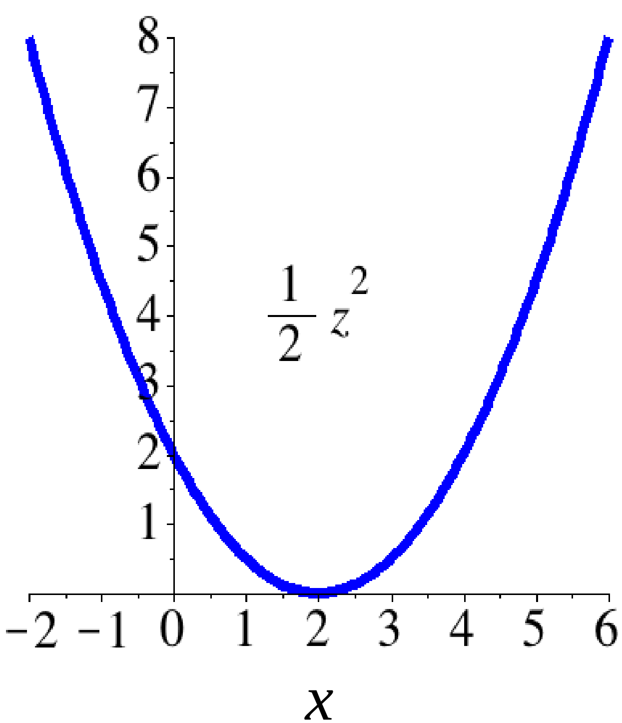
$$\left| A_N - \mu \right| < \varepsilon$$

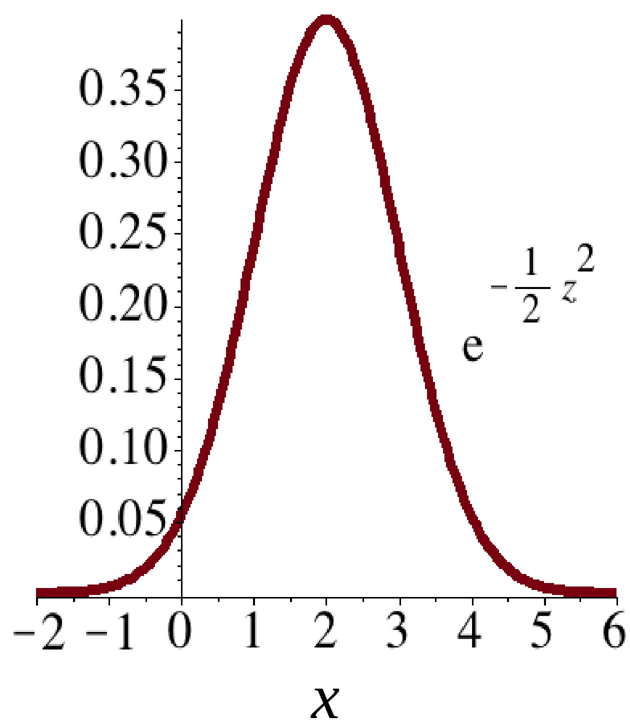
with probability at least $1 - \tau$.

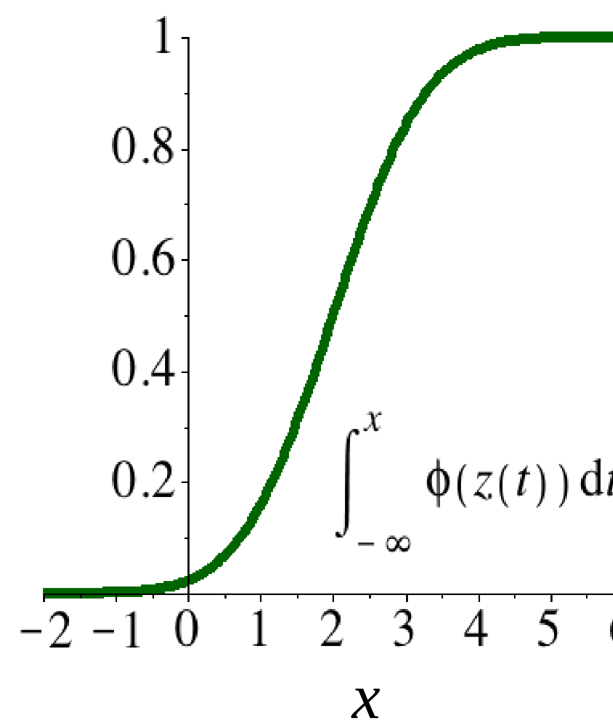
The central limit theorem

The **normal distribution** with variance 1 and mean
is the distribution for which the probability density of
$Z = (X - \quad)^2$ is proportional to $\exp(-z^2/2)$.

The normal distribution with variance $\ ^2$ and mean   is
the distribution for which this density is proportional to
$\exp(-z^2/2\ \ ^2)$







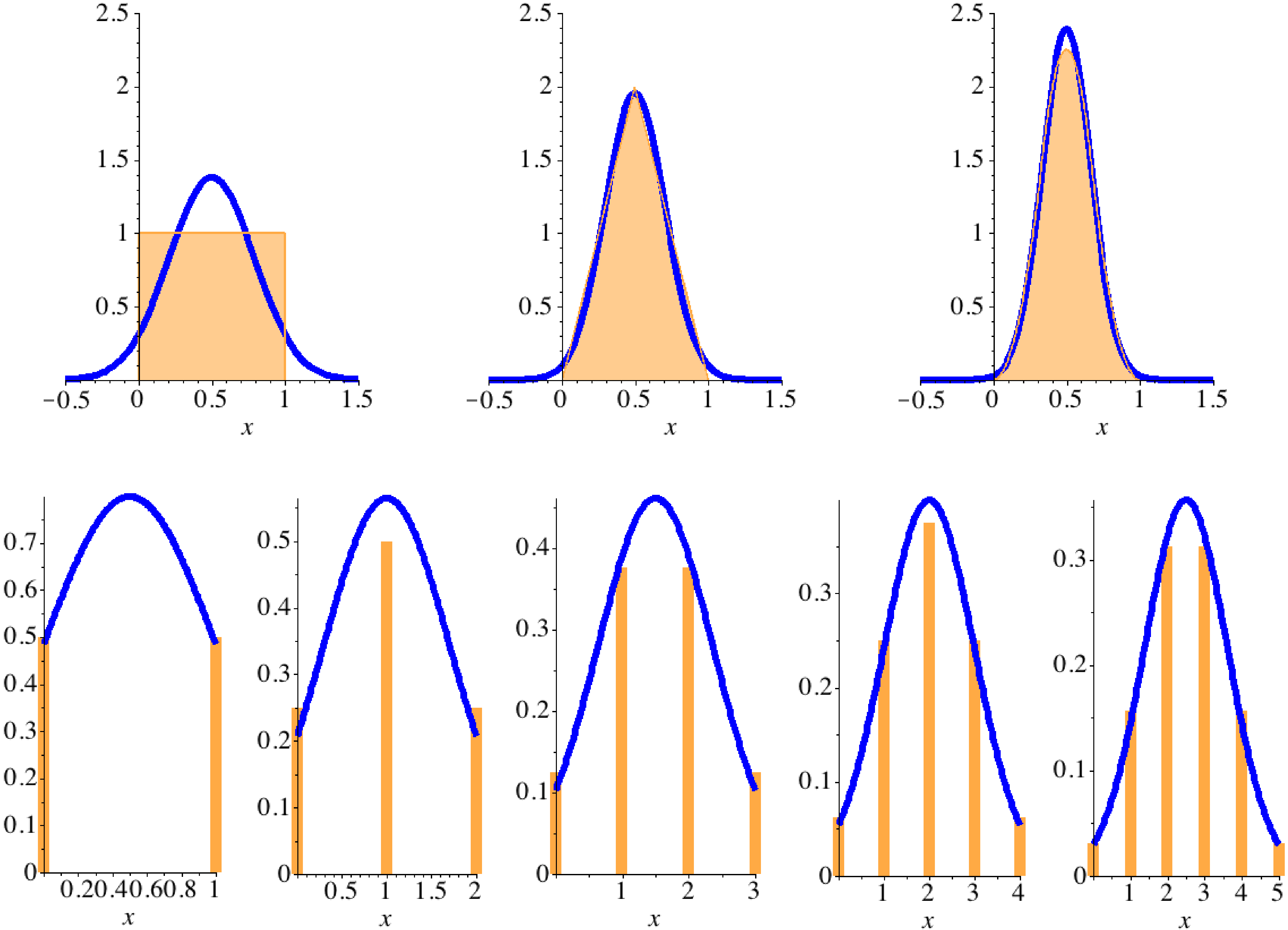
The central limit theorem

Let X_1, X_2, X_3, ... be a series of **independent** and **identically distributed** random variables with mean 0 and variance 1. Let further

$$A_k = \frac{X_1 + X_2 + X_3 + \ldots + X_1}{k}$$

be the average of the first k variables.

Then the probability density function of A_k converges (non-uniformly) to the probability density function of the normal distribution with mean 0 and variance 1.



2.5
2
1.5
1
0.5
−0.5
0
0.5
1
1.5
x
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.20.40.60.8
1
1.5
2
3
4
5

- Expected value
- Variance
- The weak law of large numbers
- The Markov bound
- The central limit theorem

Exercises for Tuesday, first hour

Mathias Winther Madsen
mathias.winther@gmail.com

January 12, 2014

Chebyshev's ineqality Suppose X is random variable with expected value $\mathsf{E}[X] = \mu$ and variance $\mathsf{VAR}[X] = \sigma^2$. Prove that

$$\Pr\left(|X-\mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2}.$$

This theorem is called **Chebyshev's inequality**.

Confidence interval Suppose a random variable X has an expected value of $\mathsf{E}[X] = \mu \geq 0$, and that most of its probability mass is located in the interval

$$[\,\mu - \sqrt{\mu},\, \mu + \sqrt{\mu}\,].$$

For which values of μ does the observation $X = 30$ lie outside this interval?

Frequentist inference A coin with an unknown bias p is flipped 100 times, and it comes up heads $S_{100} = 65$ times.

1. Suppose that $p = 0.5$. Compute the variance of each individual coin flip and the variance of S_{100}.
2. Still Assuming that $p = 0.5$, find an upper bound on the likelihood
$$\Pr\left((S_{100} - 50)^2 \geq (65 - 50)^2 \,\middle|\, p = 0.5 \right).$$
3. For which p does S_{100} have the greatest variance? What is the value of $\mathsf{VAR}[S_{100}]$ in this "worst" case scenario?
4. Find a reasonable upper bound on the likelihood
$$\Pr\left((S_{100} - \mu)^2 \geq (65 - \mu)^2 \,\middle|\, p \right)$$
for an arbitrary p. ($\mu = \mu(p)$ is the expected value of S_{100}.)
5. Exhibit a set of p for which
$$\Pr\left((S_{100} - \mu)^2 \geq (65 - \mu)^2 \,\middle|\, p \right) \leq 0.05.$$