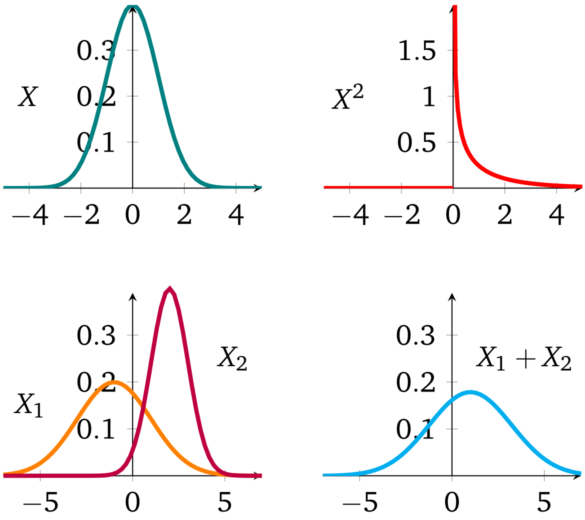
ILLC Project Course in Statistical Learning Theory

Mathias Winther Madsen
mathias.winther@gmail.com

Institute for Logic, Language, and Computation
University of Amsterdam

January 2015

Random Variables



Expected Values

Definition

The **mean** or **expected value** of a discrete variable X is

$$E[X] \;=\; p(x_1)\,x_1 + p(x_2)\,x_2 + p(x_3)\,x_3 + \cdots$$

The mean of a continuous variable is the corresponding integral.

Theorem

1. $E[X+Y] = E[X] + E[Y]$;
2. $E[c \cdot X] = c \cdot E[X]$ *when* c *is a constant;*
3. $E[X \cdot Y] = E[X] \cdot E[Y]$ *when* X *and* Y *are independent.*

Variance

Problem

Which of the following is more probable?

1. 600 tails in 1,000 coin flips;
2. 600,000 tails in 1,000,000 coin flips.

Problem

A dust particle starts at $X_1 = 0$ and walks one step left or one step right in each time period. Approximately how far away from the origin is the particle going to be after $t = 100$ time periods?

Variance

Z		z		$E[Z]$
X	-4	0	12	
$X - 2$	-6	-2	10	
$\|X - 2\|$	6	2	10	
$(X - 2)^2$	36	4	100	
$\Pr(Z = z)$	$1/4$	$1/2$	$1/4$	

Variance

Z	z			$E[Z]$
X	-4	0	12	2
$X-2$	-6	-2	10	0
$\lvert X-2\rvert$	6	2	10	5
$(X-2)^2$	36	4	100	36
$\Pr(Z=z)$	$1/4$	$1/2$	$1/4$	

Variance

Definition

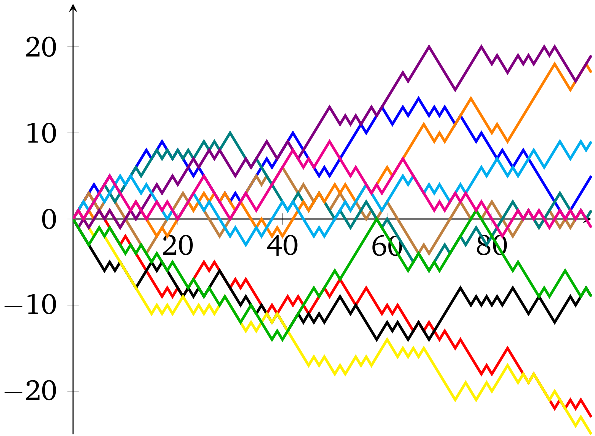
The **variance** of a random variable X is

$$\mathsf{Var}[X] \;=\; E\left[(X - E[X])^2\right].$$

Theorem

1. $\mathsf{Var}[X] = E[X^2] - E[X]^2$;
2. $\mathsf{Var}[c \cdot X] = c^2 \cdot \mathsf{Var}[X]$ *when c is a constant;*
3. $\mathsf{Var}[X + Y] = \mathsf{Var}[X] + \mathsf{Var}[Y]$ *when X and Y are independent.*

Variance



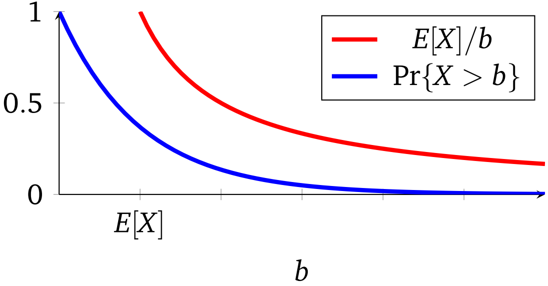
$$\mathrm{Var}[X_1 + \cdots + X_t] \; = \; \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_t]$$

The Markov Bound

Theorem

Suppose that $\Pr\{X \geq 0\} = 1$; *then*

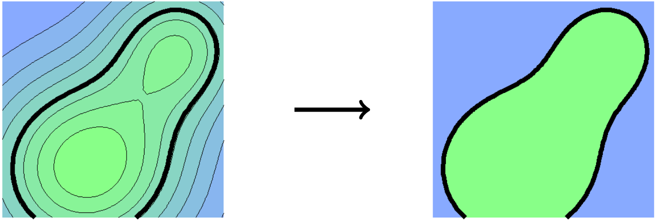
$$\Pr(X > b) \ \leq \ \frac{E[X]}{b}$$



The Markov Bound

$$B = \begin{cases} b & (X > b) \\ 0 & (X \le b) \end{cases}$$

X	0	1	2	3	4
B	0	0	0	2	2
Pr	.25	.10	.25	.30	.10



$$E[X] \ge E[B] = b \Pr\{X > b\}.$$

The Weak Law of Large Numbers

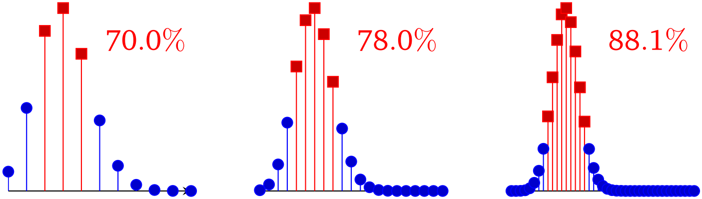
Theorem (Tchebichef, 1867)

Let

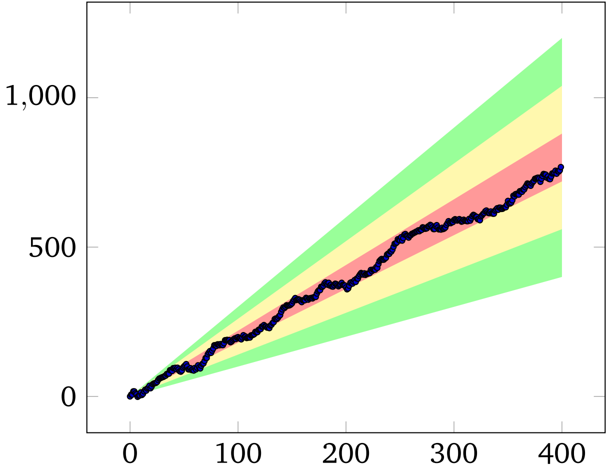
$$S_t = X_1 + X_2 + \cdots + X_t$$

be a sum of independent and identically distributed variables with a shared mean $E[X]$ and variance $\mathsf{Var}[X]$. Then

$$\Pr\left\{\left|\frac{S_t}{t} - E[X]\right| > \varepsilon\right\} \leq \frac{\mathsf{Var}[X]}{t\varepsilon^2}.$$



The Weak Law of Large Numbers



$$(E[X] - \varepsilon)t \ \leq \ S_t \ \leq \ (E[X] + \varepsilon)t$$

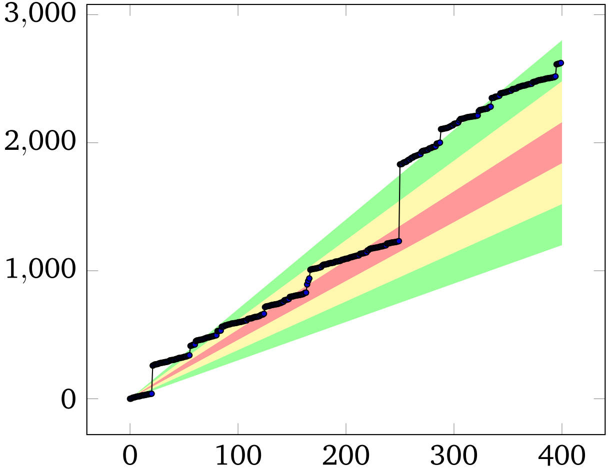
The Weak Law of Large Numbers

Problem

How many times do you have to throw a loaded die in order for the average result to deviate from the expected result by less that $\varepsilon = 1.50$ with a probability of less than $\alpha = .05$?

$$\Pr\left\{\left|\frac{S_t}{t} - E[X]\right| > \varepsilon\right\} \ \leq \ \frac{\mathsf{Var}[X]}{t\varepsilon^2}.$$

The Weak Law of Large Numbers



$$(E[X] - \varepsilon)t \ \leq \ S_t \ \leq \ (E[X] + \varepsilon)t$$