

# **Models and analysis of security protocols**

## **1st Semester 2009-2010**

### **Introduction**

### **Lecture 1**

**Pascal Lafourcade**

*Université Joseph Fourier, Verimag*

Master 2 Pro: September 24th 2009

# Instructors

## Instructors

- ▶ Pascal Lafourcade: [pascal.lafourcade@imag.fr](mailto:pascal.lafourcade@imag.fr)
- ▶ Jean-Louis Roch: [jean-louis.roch@imag.fr](mailto:jean-louis.roch@imag.fr)
- ▶ Florent Autreau: [florent.autreau@imag.fr](mailto:florent.autreau@imag.fr)

# Administrative Informations

## Where & When

- ▶ Module of 24 sessions + 16
  - ▶ 12 sessions + 3 with JL Roch
  - ▶ 11 sessions + 3 with P Lafourcade
  - ▶ 1 session + 10 with F Autreau

## Draft of Instructors Schedule

Thursday		Monday	
24/09/09	PL 1,2 JL 1	28/09/09	PL 3,4 JL 2
Tuesday	PL 5 JL 3	29/09/09	
01/10/09	FA c1, 1 PL T1	05/10/09	PL 6 JL 4,5
08/10/09	JL 6 FA 2	12/10/09	PL 7,8 JL 7
		19/10/09	JL 8, T2 FA 3
22/10/09	Partial Exam	02/11/09	PL 9 JL 9,10
05/11/09	PL T2 FA 4	09/11/09	FA 5,6 JL T3
12/11/09	PL 10, 11 JLR 11	16/11/09	PL 12, T3 FA 6,7
19/12/09	FA 8, 9, 10	23/11/09	FA 8,9,10
03/12/09	EXAM		

### Holidays

- ▶ Week 44 (26/10 - 30/10) Toussaint
- ▶ Weeks 52,53 (21/12 - 01/01) Christmas
- ▶ Weeks 7 (15/02 - 19/02) Winter

# Evaluation Informations

## Exams

- ▶ Middel Exam: 22/10/09
- ▶ Exam: 03/12/09
- ▶ Evaluation:
  - ▶ ET: Final examination: 1 written exam (3h)
  - ▶ TP: Practical work: 1
  - ▶ CC: Continuous controls: written control and one lecture report (Friday 27 November 2009).
- ▶ Final mark:  $20\% * TP + 65\% * ET + 15\% * \text{MAX}(ET, CC)$

## Possible Articles

- ▶ 1. Relations Among Notions of Security for Public-Key Encryption Schemes Crypto'98 by Mihir Bellare, Anand Desai, David Pointcheval and Phillip Rogaway
- ▶ 2. Reconciling Two Views of Cryptography (The Computational Soundness of Formal Encryption) 2000, by Martn Abadi, Phillip Rogaway in IFIP International Conference on Theoretical Computer Science (IFIP TCS2000)
- ▶ 3. Mathieu Turuani. The CL-Atse Protocol Analyser. In Term Rewriting and Applications - Proc. of RTA, volume 4098 of Lecture Notes in Computer Science, pages 277-286, Seattle, WA, USA, 2006.
- ▶ 4. Cas Cremers Unbounded Verification, Falsification, and Characterization of Security Protocols by Pattern Refinement In CCS '08: Proceedings of the 15th ACM conference on Computer and communications security, Alexandria, Virginia, USA.  
The Scyther Tool: Verification, Falsification, and Analysis of Security Protocols In Computer Aided Verification: Proceedings of the 20th International Conference on Computer Aided Verification (CAV 2008), Princeton, USA, 2008.
- ▶ 5. Alessandro Armando, and Luca Compagna. SAT-based Model-Checking for Security Protocols Analysis. International Journal of Information Security. ISSN 1615-5262. 21 September 2007.
- ▶ 6. Bruno Blanchet. Automatic Verification of Correspondences for Security Protocols. Journal of Computer Security, 17(4):363-434, July 2009.
- ▶ 7. Bruno Blanchet. A Computationally Sound Mechanized Prover for Security Protocols. IEEE Transactions on Dependable and Secure Computing, 5(4):193-207, October-December 2008. Special issue IEEE Symposium on

## Possible Articles

- ▶ 8. OAEP Reconsidered (2000) by Victor Shoup
- ▶ 9. Y. Boichut, Th. Genet, Feasible Trace Reconstruction for Rewriting Approximations. Proceedings of RTA'06, 2006.
- ▶ 10. David Basin, Sebastian Moedersheim, and Luca Vigan. OFMC: A symbolic model checker for security protocols. In International Journal of Information Security, 4 (3), pages 181-208, 2005.  
Sebastian Moedersheim and Luca Vigan. The Open-source Fixed-point Model Checker for Symbolic Analysis of Security Protocols. Fosad 2007-2008-2009, LNCS 5705, 166-194. Springer-Verlag, 2009.
- ▶ 11. A Sound Type System For Secure Flow Analysis (1996) by Dennis Volpano , Geoffrey Smith , Cynthia Irvine
- ▶ 12. A Computational Analysis of the Needham Schroder Lowe Protocol Bogdan Warinschi Journal of Computer Security;
- ▶ 13. M. Bellare, A. Desai, E. Jorjipii and P. Rogaway. A Concrete Security Treatment of Symmetric Encryption: Analysis of the DES Modes of Operation. The preliminary version of this paper was entitled A Concrete Security Treatment of Symmetric Encryption and appeared in the Proceedings of 38th Annual Symposium on Foundations of Computer Science, IEEE, 1997.

# Organization

## Timetable and Rooms

Mondya or Thursday

- ▶ 8h-9h30
- ▶ 9h45-11h15
- ▶ 11h15-12h45h

During each lecture, some students will take some notes, and produce a latex file for the next week. **REGISTRATION NOW.**

## Registration ASAP

- ▶ Lecture 1, 24th September 2009 scribes are:
- ▶ Lecture 2, 24th September 2009 scribes are:
- ▶ Lecture 3, 28th September 2009 scribes are:
- ▶ Lecture 4, 28th September 2009 scribes are:
- ▶ Lecture 5, 29th September 2009 scribes are:
- ▶ Lecture 6, 5th October 2009 scribes are:
- ▶ Lecture 7, 12th October 2009 scribes are:
- ▶ Lecture 8, 12th October 2009 scribes are:
- ▶ Lecture 9, 2nd November 2008 scribes are:
- ▶ Lecture 10, 12th November 2008 scribes are:
- ▶ Lecture 11, 12th November 2008 scribes are:
- ▶ Lecture 12, 16th November 2008 scribes are:

# Instructor Information (I)

## Address

- ▶ Instructor: Pascal Lafourcade
- ▶ Address:

VERIMAG, team DCS  
Center Equation CTL  
2, avenue de Vignate  
38610 Gières

- ▶ Office: B4D CTL 1st floor
- ▶ Email: [pascal.lafourcade@imag.fr](mailto:pascal.lafourcade@imag.fr)
- ▶ Web: <http://www-verimag.imag.fr/~plafourc/>
- ▶ Phone: +33 (0) 4 56 52 04 21 (but email is better)
- ▶ Available most of the time in my office by appointment

## Instructor Information (II)

### Research in:

Information Security, Formal Verification, Cryptographic Protocols, Rewriting, Unification, Equational Theories, Constraints:

- ▶ e-voting
- ▶ e-auction
- ▶ Group protocols
- ▶ Wireless communications
- ▶ Tools
- ▶ ...

# Web Pages

## Courses Web Pages:

- ▶ Practical Informations online.
- ▶ Slides, homeworks, references, articles...

## Unit part:

[http://www-id.imag.fr/Laboratoire/Membres/Roch\\_Jean-Louis/perso\\_html/COURS/KIOSK-SCCI-SecurityModels//](http://www-id.imag.fr/Laboratoire/Membres/Roch_Jean-Louis/perso_html/COURS/KIOSK-SCCI-SecurityModels//)

## My part:

[www-verimag.imag.fr/~plafourc/teaching/Master\\_Pro\\_2\\_2009\\_2010.php](http://www-verimag.imag.fr/~plafourc/teaching/Master_Pro_2_2009_2010.php)

What about YOU?

Please fill the form.

## Prerequisites

Some familiarity with computer system:

- ▶ operating systems,
- ▶ networks,
- ▶ programming languages.

Some mathematical notions:

- ▶ a little number theory,
- ▶ some probability notions,
- ▶ ability to follow and do proofs,  
e.g., proof by induction, contradiction...
- ▶ acquaintance with logic,
- ▶ ease with formal notation and manipulation,

but no advanced mathematics required.

Please see me if you have any doubt or question.

# What is this course about?

A Master 2nd year course.

A presentation to basics and essential notions, techniques, models used in security and cryptography.

A look at some more advanced topics in cryptography, in particular topics related to the link between mathematical point of view and computer science one.

# Contents (I)

Security touches many domains:

- ▶ cryptography,
- ▶ mathematics,
- ▶ operating system,
- ▶ networking,
- ▶ economics,
- ▶ policy and law ...

We should at least touch most of these topics, but we will not try to cover all aspects of security.

## Contents (II)

- ▶ Not a course on cryptography,
- ▶ Not a complete course on security.

## Course topics, in details

1. Introduction & Probability
2. Indistinguishability and Security Notions OW, IND, NM, CPA, CCA, Hard Problem DDH, CDH ....
3. Reduction Proofs
4. Hybrid Argument
5. Security of Symmetric encryption and OAEP
6. Security of protocols in the Symbolic Model
7. Passive Intruder
8. Active Intruder
9. Tools and Applications
10. Access Control and Security Policies
11. Non-interference, and others
12. Link between Computational and Symbolic

And a little more, if possible...

**Today: Introduction + Indistinguishability**

# Reading

## Required reading:

- ▶ No textbook!
- ▶ Many papers, indicated during the course.

## Some recommended book:

- ▶ Two volumes of:  
“The Foundations of Cryptography” by Oded Goldreich
- ▶ For background on cryptography, online book:  
“The handbook of applied cryptography” by Alfred J. Menezes, Paul C. van Oorschot and Scott A. Vanstone.  
[www.cacr.math.uwaterloo.ca/hac/index.html](http://www.cacr.math.uwaterloo.ca/hac/index.html)
- ▶ Jonathan Katz and Yehuda Lindell “Introduction to modern cryptography”

## More books

- ▶ Bruce Schneier “Applied cryptography”,
- ▶ Matt Bishop “Computer Security: Art and Science”,
- ▶ Douglas Stinson “Cryptography: Theory and Practice”,
- ▶ Simon Singh “The Code Book: The Secret History of Codes and Code Breaking” .

## Course work

- ▶ Reading.
- ▶ Class participation.
- ▶ Homework:
  - ▶ Given and explained in class,
  - ▶ Given in the slides,
  - ▶ Usually due at the start of class one week later.
  - ▶ Presentation of exercise on blackboard
- ▶ Latex file summarizing previous lecture.
- ▶ Exam.

# Cheating (I)

- ▶ All work you make must be your own!
- ▶ If you do not know if something is allowed, please ask.
- ▶ Cheating will result in failure of the course and other standard measures.

Regular class attendance is required!

## Cheating (II)

- ▶ You are encouraged to discuss the course material and assignments with others.
- ▶ You are not allowed to do assignments with others
- ▶ You may use any conversations, texts, or other material, as long as you cite your sources.

# Outline

Presentation

Motivations

Probabilities

- Mathematics Recalls

- Birthday Paradox

- Negligible Functions

Conclusion

## Typical security-critical problems

- ▶ **Secure communication**, e.g., via telephone, email, fax.  
**Objective:** confidentiality and integrity of transmitted information.
- ▶ **Internet banking.** **Objectives:** confidentiality of transactions and account information, prevention of false transactions, impossibility of repudiating (denying) a transaction by a user,  
...
- ▶ **Digital payment systems.**
- ▶ **E-voting systems, ...**

N.B.: specifying objectives (security properties) is not always easy.  
Neither is building systems that satisfy these objectives!

## Traditional security properties

- ▶ Common security properties are:
  - **Confidentiality or Secrecy**: No improper disclosure of information
  - **Authentication**: To be sure to talk with the right person.  
disclosure of information
  - **Integrity**: No improper modification of information
  - **Availability**: No improper impairment of functionality/service

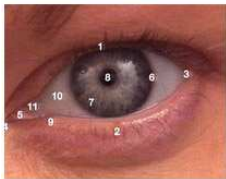
# Authentication



*"On the Internet, nobody knows you're a dog."*

## Mechanisms for Authentication

1. Something that you know  
E.g. a PIN or a password
2. Something that you have  
E.g. a smart-card
3. Something that you are  
Biometric characteristics like voice, fingerprints, eyes, ...
4. Where you are located  
E.g. in a secure building



Strong authentication combines multiple factors:  
E.g., Smart-Card + PIN

## Other security properties

- ▶ **Non-repudiation** (also called **accountability**) is where one can establish responsibility for actions.
- ▶ **Fairness** is the fact there is no advantage to play one role in a protocol comparing with the other ones.
- ▶ **Privacy**
  - Anonymity**: secrecy of principal identities or communication relationships.
  - Pseudonymity**: anonymity plus link-ability.
  - Data protection**: personal data is only used in certain ways.

## Example: banking

- ▶ A bank may require
  - ▶ authenticity of clients (at teller, ATMs, or on the Internet),
  - ▶ non-repudiation of transactions,
  - ▶ integrity of accounts and other customer data,
  - ▶ secrecy of customer data, and
  - ▶ availability of logging.
- ▶ The conjunction of these properties might constitute the bank's (high-level) security policy.

## Another example: e-voting

- ▶ An e-voting system should ensure that
  - ▶ only registered voters vote,
  - ▶ each voter can only vote once,
  - ▶ integrity of votes,
  - ▶ privacy of voting information (only used for tallying), and
  - ▶ availability of system during voting period
- ▶ In practice, many policy aspects are difficult to formulate precisely.

**Exercise:** Give the security properties that an international airport should guarantee.

More details with Florent Autreau later.

# Outline

Presentation

Motivations

Probabilities

- Mathematics Recalls

- Birthday Paradox

- Negligible Functions

Conclusion

# Outline

Presentation

Motivations

Probabilities

Mathematics Recalls

Birthday Paradox

Negligible Functions

Conclusion

# Probability Distribution (I)



A **finite probability distribution**  $D = (U, P)$  is:  
a finite, non-empty set  $U$ , together with a function  $P$  that maps  
 $u \in U$  to  $P[u] \in [0, 1]$ , such that

$$\sum_{u \in U} P[u] = 1$$

The set  $U$  is called the **sample space** and the function  $P$  is called the **probability function**.

## Example

If we think of rolling a fair die, then  $U := \{1, 2, 3, 4, 5, 6\}$ , and  $P[u] := 1/6$  for all  $u \in U$  gives a probability distribution describing the possible outcomes of the experiment.

## Probability Distribution (II)

An event is a subset  $A$  of  $U$ , and the probability of  $A$  is:

$$P[A] := \sum_{u \in A} P[u]$$

### Properties

For an event  $A \subseteq U$ , let  $\bar{A}$  denote the **complement** of  $A$  in  $U$ .

$$P[\emptyset] = 0, P[U] = 1, P[\bar{A}] = 1 - P[A]$$

For any events  $A, B \subseteq U$ , if  $A \subseteq B$ , then  $P[A] \leq P[B]$ . Also, for any events  $A, B \subseteq U$ , we have

$$P[A \cup B] = P[A] + P[B] - P[A \cap B] \leq P[A] + P[B]$$

in particular, if  $A$  and  $B$  are disjoint, then  $P[A \cup B] = P[A] + P[B]$ .

## Probability Distribution (III)

More generally, for any events  $A_1, \dots, A_n \subseteq U$  we have

$$P[A_1 \cup \dots \cup A_n] \leq P[A_1] + \dots + P[A_n]$$

and if the  $A_i$  are pairwise disjoint, then

$$P[A_1 \cup \dots \cup A_n] = P[A_1] + \dots + P[A_n]$$

### DeMorgan's law

Let  $A$  and  $B$  two events, we have:

$$\blacktriangleright \overline{A \cup B} = \bar{A} \cap \bar{B}$$

$$\blacktriangleright \overline{A \cap B} = \bar{A} \cup \bar{B}$$

### Distributive law

For events  $A, B, C$ , we have:

$$\blacktriangleright A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$\blacktriangleright A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

## Conditional distribution

Let  $D = (U, P)$  be a probability distribution. For any event  $B \subseteq U$  with  $P[B] \neq 0$  and any  $u \in U$ , let us define:

$$P[u|B] := \begin{cases} \frac{P[u]}{P[B]} & \text{if } u \in B, \\ 0 & \text{otherwise} \end{cases}$$

For any event  $A \subseteq U$ , we have

$$P[A|B] = \sum_{u \in A} P[u|B] = \frac{P[A \cap B]}{P[B]}$$

value  $P[A|B]$  is called the **conditional probability** of  $A$  given  $B$ .

## Theorem: Bayes



Suppose we have a collection  $B_1, \dots, B_n$  of events that partitions  $U$ , such that each event  $B_i$  occurs with non-zero probability. Then it is easy to see that for any event  $A$ ,

$$P[A] = \sum_{i=1}^n P[A \cap B_i] = \sum_{i=1}^n P[A|B_i]P[B_i]$$

Furthermore, if  $P[A] \neq 0$ , then for any  $j = 1, \dots, n$ , we have:

$$P[B_j|A] = \frac{P[A \cap B_j]}{P[A]} = \frac{P[A|B_j]P[B_j]}{\sum_{i=1}^n P[A|B_i]P[B_i]}$$

This equality, known as Bayes' theorem, lets us compute the conditional probability  $P[B_j|A]$  in terms of the conditional probabilities  $P[A|B_i]$ .

# Bayes Example: Cookies



	Bowl 1	Bowl 2	Totals
Chocolate Chip	10	20	30
Plain	30	20	50
Total	40	40	80

Figure: Number of cookies in each bowl by type of cookie

Fred picks a bowl at random, and then picks a cookie at random.

What is the probability that Fred has a plain cookie, given that he has picked bowl 1?

The cookie turns out to be a plain one. How probable is it that Fred picked it out of bowl 1?

# Cookies Answer (I)



	Bowl 1	Bowl 2	Totals
Chocolate Chip	1/8	1/4	3/8
Plain	3/8	1/4	5/8
Total	1/2	1/2	1

Figure: Relative frequency of cookies in each bowl by type of cookie

- ▶  $Pr(A)$  = the probability that Fred picked bowl 1 regardless of any other information. Hence  $Pr(A) = 1/2$
- ▶  $Pr(B)$  = the probability of getting a plain cookie regardless of any information on the bowls.
  - ▶  $Pr(B) = 3/8 + 1/4 = 0.625$
  - ▶  $Pr(B) = 50/80 = 0.625$
- ▶  $Pr(B|A)$  = the probability of getting a plain cookie given that Fred has selected bowl 1.  $Pr(B|A) = 30/40 = 0.75$

## Cookies Answer (II)



Given all this information, we can compute the probability of Fred having selected bowl 1 given that he got a plain cookie, as such:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.75 \times 0.5}{0.625} = 0.6$$

As we expected, it is more than half.

## Exercise: Drug Test

Suppose a certain drug test is 99% accurate, that is, the test will correctly identify a drug user as testing positive 99% of the time, and will correctly identify a non-user as testing negative 99% of the time. Let's assume a corporation decides to test its employees for opium use, and 0.5% of the employees use the drug.

We want to know the probability that, given a positive drug test, an employee is actually a drug user.

## Random variables

Let  $D = (U, P)$  be a probability distribution. A **random variable**  $X$  is a function from  $U$  into a set  $\mathcal{X}$ . If  $\mathcal{X}$  is a subset of the real numbers, then  $X$  is called a **real random variable**:

$$X(U) = \{X(u) : u \in U\}.$$

### Properties

If  $X : U \rightarrow \mathcal{X}$  is a random variable, and  $f : \mathcal{X} \rightarrow Y$  is a function, then  $f(X) := f \circ X$  is also a random variable.

Let  $X : U \rightarrow X$  be a random variable. For  $x \in \mathcal{X}$ , we write " $X = x$ " as shorthand for the event  $\{u \in U : X(u) = x\}$ .

# Independent Random Variables

## Definition

Two random variables  $X$ ,  $Y$  are **independent** if for all  $x$  in the image of  $X$  and all  $y$  in the image of  $Y$ , the events  $X = x$  and  $Y = y$  are independent:

$$P[X = x \wedge Y = y] = P[X = x]P[Y = y]$$

Equivalently,  $X$  and  $Y$  are independent if and only if their joint distribution is equal to the product of their individual distributions.

**Exercise:** Prove that  $X$  and  $Y$  are independent if and only if for all values  $x$  taken by  $X$  with non-zero probability, the conditional distribution of  $Y$  given the event  $X = x$  is the same as the distribution of  $Y$ .

# Pairwise Independent Random Variables

Let  $X_1, \dots, X_n$  be a collection of random variables, and let  $X_i$  be the image of  $X_i$  for  $i = 1, \dots, n$ . We say  $X_1, \dots, X_n$  are **pairwise independent** if for all  $i, j = 1, \dots, n$  with  $i \neq j$ , the variables  $X_i$  and  $X_j$  are independent.

# Mutually Independent Random Variables

We say that  $X_1, \dots, X_n$  are **mutually independent** if for all  $x_1 = X_1, \dots, x_n = X_n$ , we have

$$P[X_1 = x_1 \wedge \dots \wedge X_n = x_n] = \prod_{i=1}^n P[X_i = x_i]$$

More generally, for  $k = 2, \dots, n$ , we say that  $X_1, \dots, X_n$  are  $k$ -wise independent if any  $k$  of them are mutually independent.

## Example

We toss three coins, and set  $X_i := 0$  if the  $i$ th coin is "tails," and  $X_i := 1$  otherwise.

Show that the variables  $X_1, X_2, X_3$  are mutually independent.

Let us set  $Y_{12} := X_1 \oplus X_2$ ,  $Y_{13} := X_1 \oplus X_3$ , and  $Y_{23} := X_2 \oplus X_3$ , where " $\oplus$ " denotes "exclusive or," that is, addition modulo 2.

Show that the variables  $Y_{12}, Y_{13}, Y_{23}$  are pairwise independent, but not mutually independent.

# Probability Notation

$$Pr[A(X_n) = 1] = \sum_x Pr[X_n = x].Pr[A(x) = 1]$$

# Expectation

Let  $D = (U, P)$  be a probability distribution. If  $X$  is a real random variable, then its **expected value** is:

$$E[X] := \sum_{u \in U} X(u)P[u]$$

## Properties

If  $\mathcal{X}$  is the image of  $X$ , we have:

$$E[X] = \sum_{x \in \mathcal{X}} \sum_{u \in X^{-1}(x)} xP[u] = \sum_{x \in \mathcal{X}} xP[X = x]$$

More generally,

$$E[f(X)] = \sum_{x \in \mathcal{X}} f(x)P[X = x]$$

## Examples

Let  $X$  be uniformly distributed over  $\{1, \dots, n\}$ .

$$E[X] = \sum_{x=1}^n x \cdot \frac{1}{n} = \frac{n(n+1)}{2} \cdot \frac{1}{n} = \frac{n+1}{2}$$

Let  $X$  denote the value of a die toss. Let  $A$  be the event that  $X$  is even.  $X$  is uniformly distributed over  $\{2, 4, 6\}$ , and hence

$$E[X|A] = \frac{2+4+6}{3} = 4$$

Similarly, in the conditional probability space given  $\bar{A}$ , we see that  $X$  is uniformly distributed over  $\{1, 3, 5\}$ , and hence

$$E[X|\bar{A}] = \frac{1+3+5}{3} = 3$$

Hence

$$E[X] = E[X|A]P[A] + E[X|\bar{A}]P[\bar{A}] = 4 \cdot \frac{1}{2} + 3 \cdot \frac{1}{2} = \frac{7}{2}$$

## Properties

First, if  $X$  is equal to a constant  $c$  (i.e.,  $X(u) = c$  for all  $u \in U$ ), then  $E[X] = E[c] = c$ .

Second, if  $X$  takes only non-negative values (i.e.,  $X(u) \geq 0$  all  $u \in U$ ), then  $E[X] \geq 0$ . Similarly, if  $X$  takes only positive values, then  $E[X] > 0$ .

### Linearity of expectation

For real random variables  $X$  and  $Y$ , and real number  $a$ , we have

$$E[X + Y] = E[X] + E[Y] \text{ and } E[aX] = aE[X]$$

If  $X$  and  $Y$  are independent real random variables, then

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

Exercise: Proofs

# Variance

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

$$\text{Var}[X] := E[(X - E[X])^2]$$

It is a measure of the spread or dispersion of the distribution of  $X$  around its expected value  $E[X]$ . Variance is always non-negative.

## Properties

Let  $X$  be a real random variable, and let  $a$  and  $b$  be real numbers. Then we have:

- ▶  $\text{Var}[X] = E[X^2] - (E[X])^2$
- ▶  $\text{Var}[aX] = a^2 \text{Var}[X]$
- ▶  $\text{Var}[X + b] = \text{Var}[X]$

Exercise: Proofs

## Examples

Let  $X$  be uniformly distributed over  $\{1, \dots, n\}$ .

$$E[X] = \sum_{x=1}^n x \cdot \frac{1}{n} = \frac{n(n+1)}{2} \cdot \frac{1}{n} = \frac{n+1}{2}$$

We also have

$$E[X^2] = \sum_{x=1}^n x^2 \cdot \frac{1}{n} = \frac{n(n+1)(2n+1)}{6} \cdot \frac{1}{n} = \frac{(n+1)(2n+1)}{6}$$

Therefore,

$$\text{Var}[X] = E[X^2] - (E[X])^2 = \frac{n^2 - 1}{12}$$

## Theorem: Markov's inequality



Let  $X$  be a random variable that takes only non-negative real values. Then for any  $t > 0$ , we have

$$P[X \geq t] \leq \frac{E[X]}{t}$$

Exercise: Proof

# Theorem Chebyshev's inequality



Let  $X$  be a real random variable.

Then for any  $t > 0$ , we have:

$$P[|X - E[X]| \geq t] \leq \frac{\text{Var}[X]}{t^2}$$

Exercise: Proof

## Theorem: Chernoff bound



Let  $X_1, \dots, X_n$  be mutually independent random variables, such that each  $X_i$  is 1 with probability  $p$  and 0 with probability  $q := 1 - p$ . Assume that  $0 < p < 1$ . Also, let  $\bar{X}$  be the sample mean of  $X_1, \dots, X_n$ . Then for any  $\epsilon > 0$ , we have:

$$P[\bar{X} - p \geq \epsilon] \leq e^{-n\epsilon^2/2q}$$

$$P[\bar{X} - p \leq -\epsilon] \leq e^{-n\epsilon^2/2p}$$

$$P[|\bar{X} - p| \geq \epsilon] \leq 2 \cdot e^{-n\epsilon^2/2}$$

Exercise: Proof

# Outline

Presentation

Motivations

Probabilities

Mathematics Recalls

Birthday Paradox

Negligible Functions

Conclusion

# Birthday Paradox: My bet

When are you born ?



# Birthday Paradox & Probabilities (I)

- ▶ How many people must be in a room such that the probability  $p$  that one has your birthday is  $p > .5$ ?

# Birthday Paradox & Probabilities (I)

- ▶ How many people must be in a room such that the probability  $p$  that one has your birthday is  $p > .5$ ?
- ▶ In a room of  $n$  people the probability is  $1 - \left(\frac{365-1}{365}\right)^n$ .

# Birthday Paradox & Probabilities (I)

- ▶ How many people must be in a room such that the probability  $p$  that one has your birthday is  $p > .5$ ?
- ▶ In a room of  $n$  people the probability is  $1 - \left(\frac{365-1}{365}\right)^n$ .
- ▶ So we have:

$$n \geq \frac{\ln\left(\frac{1}{2}\right)}{\ln\left(\frac{365-1}{365}\right)} \approx 252,651 \approx 253$$

## Birthday Paradox & Probabilities (II)

- ▶ How many people must be in a room such that the probability  $p$  that any two share the same birthday is  $p > .5$ ?

## Birthday Paradox & Probabilities (II)

- ▶ How many people must be in a room such that the probability  $p$  that any two share the same birthday is  $p > .5$ ?
- ▶ As stated above, the probability that no two birthdays coincide is

$$1 - p(n) = \bar{p}(n) = \prod_{k=1}^{n-1} \left(1 - \frac{k}{365}\right)$$

## Birthday Paradox & Probabilities (II)

- ▶ How many people must be in a room such that the probability  $p$  that any two share the same birthday is  $p > .5$ ?
- ▶ As stated above, the probability that no two birthdays coincide is

$$1 - p(n) = \bar{p}(n) = \prod_{k=1}^{n-1} \left(1 - \frac{k}{365}\right)$$

- ▶ using the inequality  $1 + x < e^x$

$$\bar{p}(n) = \prod_{k=1}^{n-1} \left(1 - \frac{k}{365}\right) < \prod_{k=1}^{n-1} \left(e^{-k/365}\right) = e^{-(n(n-1))/(2 \cdot 365)}$$

$$e^{-(n(n-1))/(2 \cdot 365)} < \frac{1}{2}, \text{ we get } n^2 - n > 2 \cdot 365 \ln 2$$

## Birthday Paradox & Probabilities (II)

- ▶ How many people must be in a room such that the probability  $p$  that any two share the same birthday is  $p > .5$ ?
- ▶ As stated above, the probability that no two birthdays coincide is

$$1 - p(n) = \bar{p}(n) = \prod_{k=1}^{n-1} \left(1 - \frac{k}{365}\right)$$

- ▶ using the inequality  $1 + x < e^x$

$$\bar{p}(n) = \prod_{k=1}^{n-1} \left(1 - \frac{k}{365}\right) < \prod_{k=1}^{n-1} \left(e^{-k/365}\right) = e^{-(n(n-1))/(2 \cdot 365)}$$

$$e^{-(n(n-1))/(2 \cdot 365)} < \frac{1}{2}, \text{ we get } n^2 - n > 2 \cdot 365 \ln 2$$

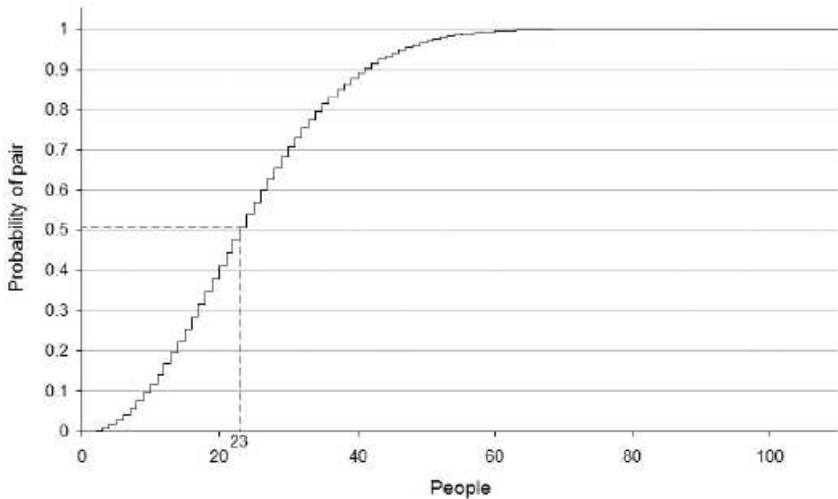
- ▶ Hence  $n \geq 23$  persons!

$A_n = \{\text{At least two of the } n \text{ people share a birthday}\}$

n	$A_n$	n	$A_n$
1	0.00000000	16	0.28360400
2	0.00273972	17	0.31500766
3	0.00820416	18	0.34691141
4	0.01635591	19	0.37911852
5	0.02713557	20	0.41143838
6	0.04046248	21	0.44368833
7	0.05623570	22	0.47569530
8	0.07433529	23	0.50729723
9	0.09462383	24	0.53834425
10	0.11694817	25	0.56869970
11	0.14114137	26	0.59824082
12	0.16702478	27	0.62685928
13	0.19441027	28	0.65446147
14	0.22310251	29	0.68096853
15	0.25290131	30	0.70631624

## Probabilities

## Birthday Paradox



## Birthday Paradox Generalization (I)

**Exercise:** The setting is that we have  $q$  balls. View them as numbered,  $1, \dots, q$ . We also have  $N$  bins, where  $N \geq q$ . We throw the balls at random into the bins, one by one, beginning with ball 1. At random means that each ball is equally likely to land in any of the  $N$  bins, and the probabilities for all the balls are independent. A collision is said to occur if some bin ends up containing at least two balls. We are interested in  $C(N, q)$ , the probability of a collision. The birthday paradox is the case where  $N = 365$ . We are asking what is the chance that, in a group of  $q$  people, there are two people with the same birthday, assuming birthdays are randomly and independently distributed over the days of the year.

## Birthday Paradox Generalization (II)

Let  $C(N, q)$  denote the probability of at least one collision when we throw  $q \geq 1$  balls at random into  $N \geq q$  buckets. Then

$$C(N, q) \leq \frac{q(q-1)}{2N}$$

$$C(N, q) \geq 1 - e^{q(q-1)/2N}$$

Also if  $1 \leq q \leq \sqrt{2N}$  then  $C(N, q) \geq 0.3 \cdot \frac{q(q-1)}{N}$

Hint: first prove the inequality  $(1 - 1/e) \cdot x \leq 1 - e^{-x} \leq x$

# Outline

Presentation

Motivations

Probabilities

Mathematics Recalls

Birthday Paradox

Negligible Functions

Conclusion

## Negligible functions

We call a function  $\mu : \mathbb{N} \rightarrow \mathbb{R}^+$  negligible if for every positive polynomial  $p$  there exists an  $N$  such that for all  $n > N$

$$\mu(n) < \frac{1}{p(n)}$$

### Properties

Let  $f$  and  $g$  be two negligible functions, then

1.  $f.g$  is negligible.
2. For any  $k > 0$ ,  $f^k$  is negligible.
3. For any  $\lambda, \mu$  in  $\mathbb{R}$ ,  $\lambda.f + \mu.g$  is negligible.

Exercise: Proofs

# Negligible Functions

Exercise: Prove or disprove:

- ▶ The function  $f(n) := \left(\frac{1}{2}\right)^n$  is negligible.
- ▶ The function  $f(n) := 2^{-\sqrt{n}}$  is negligible.
- ▶ The function  $f(n) := n^{-\log n}$  is negligible.

# Noticeable Functions

Instead of "there exists an  $N$  such that for all  $n > N$ " we will in the following often say "for all sufficiently large  $n$ ".

We call a function  $\nu : \mathbb{N} \rightarrow \mathbb{R}$  noticeable if there exists a positive polynomial  $p$  such that for all sufficiently large  $n$ , we have:

$$\nu(n) > \frac{1}{p(n)}$$

Note: A function can be neither noticeable nor negligible.

## Exercises

Prove or disprove the following statements:

1. If both  $f, g \geq 0$  are noticeable, then  $f - g$  and  $f + g$  are noticeable.
2. If both  $f, g \geq 0$  are not noticeable, then  $f - g$  is not noticeable.
3. If both  $f, g \geq 0$  are not noticeable, then  $f + g$  is not noticeable.
4. If  $f \geq 0$  is noticeable, and  $g \geq 0$  is negligible, then  $f.g$  is negligible.
5. If both  $f, g > 0$  are negligible, then  $f/g$  is noticeable.

# Outline

Presentation

Motivations

Probabilities

- Mathematics Recalls

- Birthday Paradox

- Negligible Functions

Conclusion

# Today

1. Motivation
2. Probabilities (Random Variables)
3. Birthday Paradox
4. Negligeable Functions

# Next Time

## 1. Indistinguishability

**Thank you for your attention.**

**Questions ?**