# Statistics 36-462
# Chaos, Complexity, and Inference

Prof. Cosma Shalizi

Spring 2009

## Prerequisites

**course in mathematical statistics** E.g., 36-310, 36-401, or 36-625/626

***and* course in probability, *including random processes*** E.g. 36-217, 36-225/226, 36-410, or 36-625/626

***or* consent of instructor**

The last page of this handout lists the concepts you should be familiar with. Some programming experience will be *extremely* helpful.

## Textbooks and Readings

Three books are **required**: Flake's *The Computational Beauty of Nature*; Miller and Page's *Complex Adaptive Systems*; Smith's *Chaos: A Very Short Introduction*. Four books are **optional but recommended**: Guttorp's *Stochastic Modeling of Scientific Data*; Krugman's *The Self-Organizing Economy*; Fraser's *Hidden Markov Models and Dynamical Systems*; and Braun and Murdoch's *A First Course in Statistical Programming with R*. Various papers are also going to be **required**. The detailed reading schedule, including links to download the papers, is online at

`http://www.stat.cmu.edu/~cshalizi/462/syllabus.html`

Check it often, since it will almost certainly be updated as we go along.

## Logistics

**Class** Tuesdays and Thursdays 12:00 to 1:20 in 208 Scaife Hall

**Office Hours** Wednesdays 10:00 to 11:00 and Thursdays 4:00 to 5:00 in 229C Baker Hall, or by appointment

**Homeworks due** Fridays at 5 pm *by e-mail*

## Contact

**Office** 229C Baker Hall

**Phone** 412-268-7826

**E-Mail** cshalizi [at] cmu.edu
This is by far the best way to contact me.

## Grading

**Problem sets** There will be one problem set every week (approximately). Problems will be mostly simulations and calculations, with only a few proofs. Submit your homework electronically as PDFs (no Word files). For computational questions, do *not* submit raw sessions or code, but describe methods and results in English; include code, when appropriate, as a separate attachment (so I can run it). This will be $1/3$ of your grade.

**Writing** In addition to the problems, you will write one page per week (at least) on that week's readings, to be e-mailed at the same time as the homework. (You can include it in the same PDF if you like.) This will be $1/6$ of your grade. (Do the readings!)

**Class participation** will be $1/6$ of your grade. (Come to class!)

**Final exam** There will be a take-home final, which will consist of about half-dozen long problems requiring you to set up models and/or evaluate existing models using the techniques taught in class. You will be expected to do *one* of the problems. The exam will be made available on 28 April, and your write-ups will be due (by e-mail) at 5 pm on 9 May. This will be $1/3$ of grade of your grade.

## Things You Should Already Know

If more than a handful of these concepts are unfamiliar to you — not just rusty, "I used to know that" things, but new or "I never got that at all" things — then see me at once.

**From Probability** : event, random variable, indicator variable; probability mass function, probability density function, cumulative distribution function; joint and marginal distributions; transformation of distributions; conditional probability; independence, conditional independence; independent and identically distributed (IID); expectation, variance; Markov and Chebyshev inequalities; binomial, multinomial, geometric and Poisson distributions; Bernoulli sequences; exponential and Gaussian distributions; law of large numbers; central limit theorem; random or stochastic process; stationary and non-stationary processes; random walk; Markov chain, state, transition matrix; transient and recurrent state; invariant or stationary distribution.

**From Statistics** : sampling from a population; sample mean, variance, standard deviation, median; covariance and correlation; histogram; likelihood, maximum likelihood estimation; point estimates, accuracy, precision, bias, standard errors, consistency, efficiency; confidence intervals; hypothesis testing, error rates; contingency table, chi-squared ($\chi^2$) test; goodness-of-fit, $p$-value; mean-squared error, bias-variance decomposition; linear regression, coefficients, residuals; time series.