STAT 238 01 / STAT 538 01 (F15)

Stat 238a/538a Syllabus Fall, 2015

Remember to reload this page to be sure you are seeing the most recent version.

Announcements:

- Sample final exams are available from 2013 and 2014, with solutions. I sent you an email about this (and you can find them in the Resources folder of classesv2).
- TA sessions/Office hours for this week include:
  - Dana Yang: Wed 6:30 in room 107 of Statistics Dept, Friday 2:00 (location?)
  - Bre Chryst will have hours as usual: Tuesdays at 6:00 in 17 Hillhouse room 07, Wed 2:00-3:00, 3rd floor of the Statistics Dept
  - Joe Chang: Tuesday 4:00-6:00 and Thursday 3:00-5:00 in 24 Hillhouse room 211.

Instructor: Joe Chang
TAs: Bre Chryst, Xiaoqian (Dana) Yang, Vivek Ashok
TA sessions: These are recommended but optional. Obviously, if you are having any difficulties or feel that you could use any extra help, they are *highly* recommended! (Not that you need to be having difficulties to attend; you may just feel like talking about the class or socializing with cool people.)

- Dana Yang: Mondays at 6:30 in 17 Hillhouse room 07
- Bre Chryst: Tuesdays at 6:00 in 17 Hillhouse room 07

Office hours:

- J. Chang Tuesdays 4:00-6:00 in 24 Hillhouse room 211.
- Bre: Wed 2:00-3:00, 3rd floor of the Statistics Dept (24 Hillhouse Ave.)
- Dana: Wed 6:30-7:30 in room 107 of the Statistics Dept (24 Hillhouse Ave.)

Our Piazza site. (Explanation of what this is about.)

Some materials:

- Handout from the first day of class: General information about the class.
- Further information about class policies, including homework.
- Lecture notes for this class from 2009. This file was from a previous incarnation of this class; I am posting this right away because I expect things to run similarly this year and it could be useful to get an idea where we are going. Further material will be added and updated over the semester and posted to this syllabus.

- R stuff: R project home page Download R, packages, etc. A brief R "reference card": brief list of useful commands.
- Other references of possible interest:
    - Grinstead and Snell's book *Introduction to Probability.* This is a good reference for the probability theory we will cover as part of the course; it has also been the textbook for Stat 241a/541a.
    - Bolstad's book, *Introduction to Bayesian Statistics.* Contains material on Bayesian statistics, including some probability, and even a little appendix called "Introduction to Calculus."
    - Cowles' book, *Applied Bayesian Statistics: With R and OpenBUGS Examples.* This is a very useful reference, particularly for topics later on in the course.
    - John Kruschke's book *Doing Bayesian Data Analysis.* This is becoming popular; it is very nice and friendly, particularly for JAGS. You can see it online by using this Orbis link and then click "Online book".

Other references may be added over time.

Week 1.

- Thurs 9/3/15. We talked a lot about the class, including going over the information sheet for the class and previewing what we will be covering. Notes from class. We also started the class material, including the fundamental counting principle and its application to the birthday problem. For next time, please install R and R-Studio on your computers and then please bring your laptops to class if possible -- we'll be introducing them in class.

Week 2.

- Tues 9/8/15. We started by solving the birthday problem mathematically. Then we introduced R in class (you brought your computers) and went through a lot of basic functions and stressed how to work with R and RStudio. Here are links for the script we actually wrote while in class (you can open this in R, run the commands, etc.) and a web page that shows the R output along with the commands (you can view this in a browser).
- Thurs 9/10/15. We continued working with R, including writing a function that calculates birthday probabilities, learning about "for" loops, and writing a Monte Carlo simulation to approximate birthday probabilities. The R script from class, and a web page with commands and output. We talked about counting and combinations a bit more, did an example of the probability of a "full house" in poker, and some of the definitions in the mathematical theory of probability, including sample space, events, probability measures, and random variables. Here is a record of the "white board" from class.
- Problem set 1 is assigned; due in a week.
- Problem set 1 solutions.

Week 3.

- 9/15/15. I showed you how to make a word (or pdf or html) document out of an R script by pushing the "Compile Notebook" button in RStudio. From the class notes, we talked about the "windshield example," axioms of probability and definition of a probability measure, conditional probability and independence, the product rule, binomial distributions, and the Newton-Pepys correspondence as an example. Here is a short R script illustrating some binomial calculations to solve the Newton-Pepys problem.
- 9/17/15. We reviewed the definition of conditional probability and then introduced the law of total probability and Bayes' rule using a little problem about a frog taking two hops. Here are extra notes on the frog problem, and here is the "white board" from class. We also discussed the "Monty Hall" problem, which will be on the next problem set; here is the video we watched in class.
- Problem set 2 assigned.
- Problem set 2 solution.

Week 4.

- 9/22/15. Using Bayes' rule to do the "clinical trial" statistical problem by brute force discretization. The "white board" from class. For the R script, I'll put put a combined version in for 9/24/15 below.
- 9/24/15. Finished the clinical trial problem. We then went though pp. 26-29 in the notes, covering probability density functions, cumulative distribution functions, how they are related, uniform distributions, exponential distributions, and the memoryless property. R script (this is the one that does a discrete Bayesian treatment of the statistical problem where X ~ Bin(21, theta) and we observe X=21).
- Problem set 3 assigned.
- Problem set 3 solution.

Week 5.

- 9/29/15. First we went over a couple of miscellaneous things about R; the first was the use of the "apply" function to apply a function (for example, max or median) to the rows or columns of a matrix. [Note: This could be used in the simulation in problem 2 of ps3, although it is not at all necessary to use this; you could do the problem just fine the way we have been doing our other simulations by first doing the experiment once and then putting that inside a loop to repeat it, where one repetition of the experiment is simply generating 3 uniform rv's and taking the maximum (or median).] The second thing was about density histograms and overlaying a theoretical density. Next we went back to talking about cumulative distribution functions and probability density functions, and did some examples, including the laser pole problem. R script. White board from class.
- 10/1/15. Finished discussion of the laser pole problem, including prior, likelihood, and posterior for a small data set, and trying it out for a larger simulated data set to see how the effect of the prior became very minor. We went through some definitions and concepts from sections 1.14 and 1.15, including joint, marginal, and conditional

distributions. Then we started talking about expectation (section 1.16), giving the definition and motivating the definition as a limiting long-run average. We stated the linearity of expectations (for example, E(X+Y)=E(X)+E(Y)) and showed how this could be applied to give a nice simple derivation of *np* as the expected value of the Binomial(n,p) distribution. R script. White board from class. Extra notes on why multiply densities to get a likeihood? (in response to a Piazza question) and extra examples related to sections 1.14 and 1.15 (which might also be helpful for ps4, particularly #2).
- Problem set 4 assigned.
- Problem set 4 solution.

Week 6.

- 10/6/15. More on expectation. Interpretations, examples including Uniform distributions, "LOTUS" (Law of the Unconscious Statistician), some properties including linearity of expectation, expectation of a product, started discussing variance. White board from class.
- 10/8/15. What does it mean to say that two random variables are independent? Variance: definition, examples, properties. Variance of binomial distributions. We also distinguished the variance of a random variable from the "sample variance" of a sample of values (what R would calculate with its "var" command). White board from class. R script.

Week 7.

- 10/13/15. Variance and polls. Law of Large Numbers: the idea, precise statement, simulation showing it in action "as usual" and the scary behavior that can happen for distributions without a mean, and a mathematical proof. Started discussing Normal distributions. White board from class. R script.
- 10/15/15. Midterm exam, in class.

Week 8.

- 10/20/15. Normal distributions, normal densities, "68-95-99.7 rule", use of R and Normal tables, Normal approximations (see sections 1.21 and 1.24 in the notes). White board from class. R script.
- Problem set 5 assigned.
- Problem set 5 solution.
- 10/22/15. [10/24/15. October recess, no class.]

Week 9.

- 10/27/15. More on the idea of convolutions and the Central Limit Theorem. Started Markov chains. White board from class. R script. A quick video about matrix multiplication. A matrix multiplication song. Another multiplication song (well, it involves vectors if not matrices... we didn't get to the actual singing in class).
- 10/29/15. Markov chains, more on examples and simulation, including random walks on the integers, Wright-Fisher chain, random walks on graphs. Markov property. Distribution of Xt, and how matrix multiplication describes this. Stationary distributions. White board from class. R script.
- Problem set 6 assigned.
- Problem set 6 solution.
- 10/30/15 is the last day to withdraw from a course without the course appearing on the transcript.

Week 10.

- 11/3/15. More about Markov chains, long-run behavior, Ergodic theorem, symmetry, "time-reversibility" (or "detailed balance"), random walks on graphs, Monte Carlo. White board from class. R script.
- 11/5/15. Interpretation of stationarity and time reversibility (or "global balance" and "detailed balance") in terms of monetary flows. Started talking about MCMC: random walk Metropolis method. White board from class. R script.
- Problem set 7 assigned.
- Problem set 7 solution.

Week 11.

- 11/10/15. General idea of Metropolis-Hastings sampler, and how you could have thought of it yourself. Probability versus Statistics. MCMC via Metropolis sampling applied to the "subliminal" problem. White board from class. R script.
- Materials uploaded before class about JAGS: quick start instructions and JAGS "hello world" script. Here is a JAGS 4.4.0 user manual; it is not the friendliest of manuals, but at least it has useful summaries of functions and distributions included in JAGS, and I've added bookmarks for those things in the pdf file.
- 11/12/15. Reviewed MCMC "from scratch" applied to the subliminal problem. An animation you can page through to get more of a feeling for what the MCMC is doing. We learned how to use JAGS through the rjags library. White board from class. R script.
- Problem set 8 assigned.
- Problem set 8 solution.

Week 12.

- 11/17/15. More miscellanoues issues about JAGS including the meaning of "<-" and transforming and monitoring variables, and error messages. The idea of conjugate priors. Beta distributions and Bayesian updating for beta priors and binomial likelihoods. Gamma distributions. White board from class. R script.
- 11/19/15. The idea of scale-free priors and the use of a prior like Gamma(.01,.01) for scale parameters (e.g. precision parameters in Normal distributions). Regression: the IQ-vs-crying example, and the salary-vs-male-and-experience example. White board from class. R script. [And here is the heartwarming and inspiring article on crying and IQ.]
- Problem set 9 assigned.
- Problem set 9 solution.

No class 11/24, 11/26: Fall Recess.

Week 13.

- 12/1/15. No class.
- 12/3/15. Some miscellaneous things: correlation, ideas about the "regression effect" and "regression fallacy", conjugate prior and Bayesian updating results for a Normal mean (including the nice interpretation of the formulas), and started talking about the Gibbs sampler. White board from class. R script.

Week 14.

- 12/8/15. We talked more about the regression model -- in particular, the conditional variance of Y given X=x (the variance in "vertical strips" of the data). Then we talked more about the Gibbs sampler. This enabled us to understand how centering the "'x" variable in a regression can help the Gibbs sampler work a lot better. Then we talked about hidden Markov models. White board from class. R script.
- 12/10/15. (Computational takehome exam due, or tomorrow by 1:00.) More on hidden Markov models. Discussion of examples of making up our own models, and some concepts of model selection, including example of application to fusion theory of cancer metastasis, idea of penalizing for complexity illustrated by polynomial fits to the crying-IQ data, deviance information criterion (article from Journal of the Royal Statistical Society) and how to calculate the DIC very conveniently in JAGS and R, and the Galileo data (nonsensical pages from *The Statistical Sleuth* and a physical derivation that hopefully makes more sense). R script.
- "Problem set 10" (actually just two extra problems), with solutions.

Our final exam time slot is Sun 12/20/15, 2:00 PM.