Math 1070-2: Spring 2008

Lecture 4

Davar Khoshnevisan

Department of Mathematics
University of Utah
http://www.math.utah.edu/~davar

January 30, 2009

Gathering data

- It is important to gather data well (Representation and accuracy, for example)
- It is important to understand how the data is gathered (Analysis, reliability)
- Perfect data $\longrightarrow$ all experiments agree (physics?)

The cell-phone example

Are cell phones dangerous?

- A German study compared 118 patients with uveal melanoma (an eye cancer) to 475 without. Cell-phone usage was measured via a questionnaire. U.M. was higher for those who used their cell phones more
- A US study compared 469 patients with brain cancer to 422 without. Cell-phone usage was measured via a questionnaire. The two groups had similar cell-phone usage
- An Australian study considered 200 transgenic mice, bred to be susceptible to cancer of the immune system. 100 mice were exposed to two $\frac{1}{2}$-hour periods a day to micorwaves akin to those of cell phones. The other 100 mice weren't exposed. After 18 months the brain-tumor rate for the exposed category was twice the other group
- What is going on? Why are there contradictions?

Controlled experiments vs observational studies

- Controlled experiments:
  - Have/want: explanatory/response variables
  - Goal/hope: remove (as possible) the effect of other variables
  - The experimental conditions that enforce the explanatory variables (and those only) are called treatment
- Observational studies:
  - Observe the response and explanatory variables without affecting the subject(s) [e.g., no treatment]
  - Focus groups
- If possible control (control lurking variables; wins hands down!)
- Can never establish cause and effect using an observational study (Association does not imply causation, and all that!)

Questions on the cell-phone case

- Questionnaires $\longrightarrow$ observational study
  - The German and the US studies
  - Voluntary response
  - Subjects' perception plays a greater role
- In the Australian study the experimenter *controlled* the amount of radiation per subject
- In the Australian study, the subjects' background (here, susceptibility of the transgenic mice to the disease at hand) was *controlled*

Why not always control?

. . . and how much control?

- The perfect experiment: Expose $\frac{1}{2}$ the people to cell-phone microwaves; do not expose $\frac{1}{2}$; see what happens
  - Cannot be done (ethics, law, and all that!)
  - Would have to control for cell-phone use for the duration of the experiment (virtually impossible)
  - Ideally, the subjects have to be similarly susceptible (impossible)
- Use of observational studies:
  - Sometimes we seek answers other than causation
  - Reasonable examples:
    - Matters of opinion/taste such as public opinion, advertising, etc.
    - Sample surveys such as the Census, Consumer Price Index, etc.

The quality of a sample

- A not atypical scenario:
- “NYC is dangerous”
- ... yeah ... and the last time I was there I was robbed ☹
  - Sample could be biased
  - Sample too small in size and breadth
  - Robbery is an atypical event to begin with . . . intentionally sampled one such
- End-result is statistically useless (though perhaps entertaining ☺)
- The other 99 people in the room didn’t say they have been to NYC once a year, bet were OK every time ☯

Random samples

1. Select a sampling frame
   1.1 This is where you will take your sample from
   1.2 Some times the population is not tangible
   1.3 Want it to “look like the population”
2. Decide on a good method to select a sample (sampling design)
   2.1 What if you can sample the entire sampling frame? Don’t need statistics ☺
   2.2 Stand in a street corner and . . . no good ☹
   2.3 “Random” sampling
   - Fair and representative 👍
   - Statistically reproducible 👍
   - Not Census approved ☹
   - Requires understanding the nature of randomness ☹
   - We’ll dig not too deeply ☺

What is a random sample?

- Any individual has the same chance of getting selected
- Random $\neq$ haphazard, or “anything goes”
- Random samples have statistical properties (laws)
- Random samples have patterns, but “not obvious” ones
- How can we take a random sample?
  - Physical machines (coins, dice, etc)
  - Random-number generators (?how?)

Using your in-text random-number generator

Table E, page A6

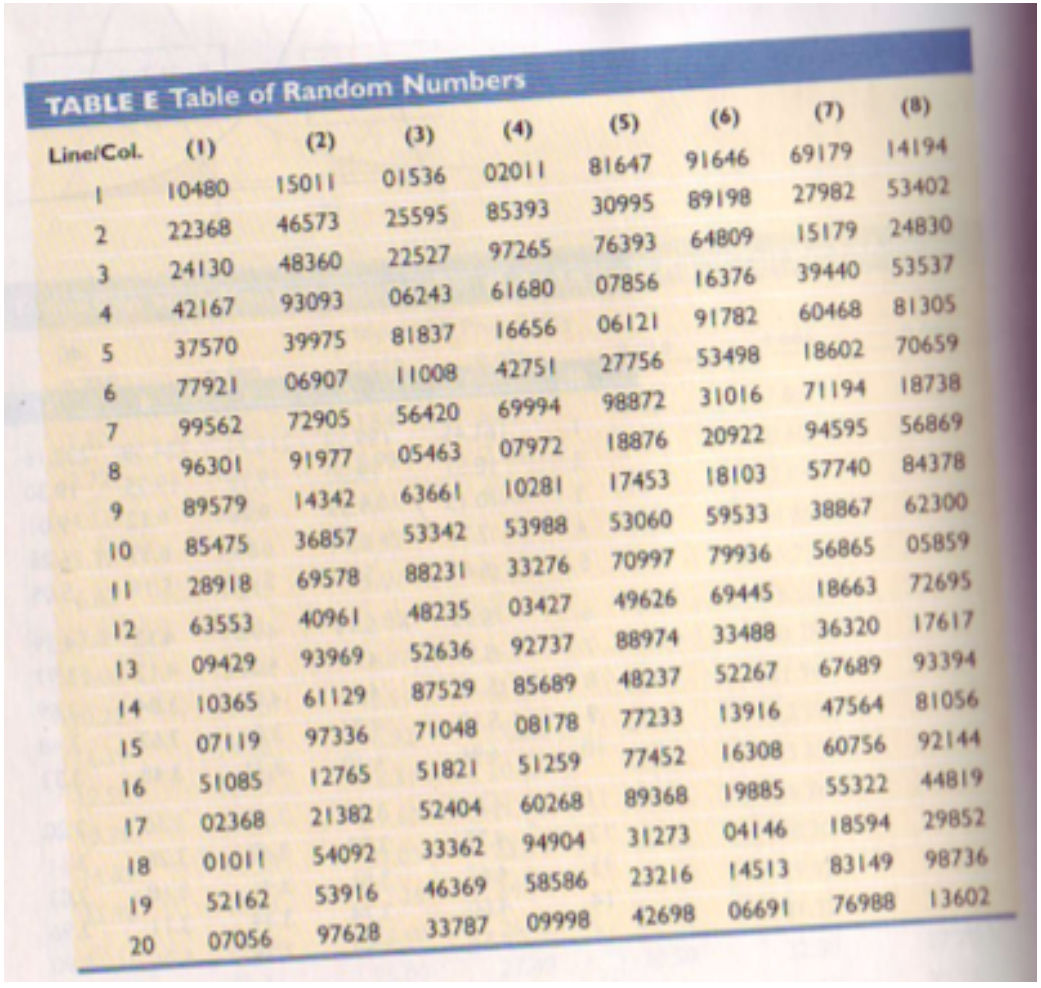

TABLE E Table of Random Numbers

Line/Col.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	10480	15011	01536	02011	81647	91646	69179	14194
2	22368	46573	25595	85393	30995	89198	27982	53402
3	24130	48360	22527	97265	76393	64809	15179	24830
4	42167	93093	06243	61680	07856	16376	39440	53537
5	37570	39975	81837	16656	06121	91782	60468	81305
6	77921	06907	11008	42751	27756	53498	18602	70659
7	99562	72905	56420	69994	98872	31016	71194	18738
8	96301	91977	05463	07972	18876	20922	94595	56869
9	89579	14342	63661	10281	17453	18103	57740	84378
10	85475	36857	53342	53988	53060	59533	38867	62300
11	28918	69578	88231	33276	70997	79936	56865	05859
12	63553	40961	48235	03427	49626	69445	18663	72695
13	09429	93969	52636	92737	88974	33488	36320	17617
14	10365	61129	87529	85689	48237	52267	67689	93394
15	07119	97336	71048	08178	77233	13916	47564	81056
16	51085	12765	51821	51259	77452	16308	60756	92144
17	02368	21382	52404	60268	89368	19885	55322	44819
18	01011	54092	33362	94904	31273	04146	18594	29852
19	52162	53916	46369	58586	23216	14513	83149	98736
20	07056	97628	33787	09998	42698	06691	76988	13602

- There are better RNGs out there ☹
- The principles are the same ☺

Selecting k people from this lecture room

- Compile a sampling frame (here, 1–42)
- If there are n people in the sampling frame, then we want to select n-digit numbers from the table k times
- Here, $n = 2$ and $k = 3$
- All n-digit numbers should have the same "chance" of coming up (do they?)
- Is there true randomness?
- **Nir•va•na:** *a transcendent state in which there is neither suffering, nor sense of self, and the subject is released from the effects of karma and samsara. It represents the final goal of Buddhism*

Sampling for observational studies

- Want a random sample:
  - Bad RNG ☹
  - Sampling bias (stand on the corner of 15th and 15th and . . .) ☹
  - Non-response bias ☹
- Some common methods (obs. studies):
  - Personal interviews
    - Interviewer records the response of the subjects
    - Good response ☺
    - Expensive ☹
  - Telephone interviews
    - Usually random-digit dialing ☺
    - Low cost ☺
    - Inconsistent results ☹
  - Self-administered interviews
    - Ask people to respond via mail, email, phone, . . .
    - Low to low-ish costs ☺
    - Non-response bias ☹
    - Inconsistent results ☹

Sample size

- Does sample-size matter?
- Yes, and no
- A random sample of 50,000 can be better than an opinion-poll of 2.4 million (example next)

Landon vs Roosevelt

Some background

- Pres. Franklin Delano Roosevelt (FDR) running for re-election for the US Presidency (1936)
- Main Rep. candidate = Gov. (Alf) Alfred Landon (KS)
- Recovery from The great Depression (9 million unemployed; real income had dropped by $\frac{1}{3}$ from 1929–1933, and was only just turning around)
- Landon's Platform: Economy in Government ("the spenders must go")
- FDR's platform: Deficit financing ("...balance the budget of the American people before ...the national government")
- Both candidates concerns with domestic affairs (Nazis arming Germany; Civil war in Spain reaching its climax)

Literary Digest's prediction

- Most people thought FDR would win
- *Literary Digest* predicted only 43% for FDR
  - Readership of Lit. Dig. was about 2.4 million!
  - Lit. Dig. had correctly predicted the elections five times!
- In fact: FDR = 62% (landslide!)
- Largest gross prediction error in pres. races 

!
- The Lit. Dig. went bankrupt soon after

The Literary Digest's sample

- Mail questionnaire to 10 million people (2.4 million responded)
- Addresses were chosen from:
  - Telephone books ($\frac{1}{4}$ of US households had ☎)
  - club membership listings ☹, . . .
  - **Bi•as:** A systematic tendency on the part of the sampling procedure to exclude portions of the sampling frame
    - Selection bias (here, rich/educated not poor/undereducated; "undercoverage")
    - Nonresponse bias (here, 2.4 out of 10 million = 24% nonresponse rate)

The Literary Digest's sample

- George Gallup used random sampling (50,000) and predicted FDR to win 56% fact: 62%; error = 6 percentage points
- He also used random sampling to predict the prediction of lit. digest error = 1 percentage point!!
- Gallup did well for himself thereafter ☺
- From 1936-1948, Gallup's predictions were at about 5%; now they are a lot smaller (barring a couple of recent glitches)
- *The Gallup Poll* currently uses sample sizes of $\approx 5,000$ ($\frac{1}{10}$th of 1936! $<$ 5 out of 100,000 persons!)

A little more on nonresponse

- Lit. Digest made a special effort for Chicago voters
- They mailed a questionnaire to every third registered voter in Chicago (1936)
- About 20% responded; of these about $\frac{1}{2}$ favored Landon
- In fact, about $\frac{2}{3}$ of Chicago voters voted for FDR!
- TV studies typically have about 60-80% nonresponse

Random sampling

- A *random sample* of size n: All possible samples of size n have the same chances of being selected
  - Need some probability theory
  - The margin of error (the give or take) is $\approx 100\%/\sqrt{n}$
- For George Gallup's prediction of the FDR/Landon race,

$$\text{margin of error} = \frac{100\%}{\sqrt{50,000}} \approx 0.45\%$$

- Gallup was off by $6\% \approx 13$ times the margin of error
- Gallup's was probably also not a simple random sample (Later: chance variation vs statistically-significant difference)
- $n = 5,000 \rightarrow \text{MOE} \approx 100\%/\sqrt{5,000} \approx 3\%$ (today's standards)

Tidying up some loose ends

- A random sample of 1,000 very often beats a “clever” sample of 10 million
- Watch out for sources of potential bias
  - Sampling bias (nonrandom samples and/or undercoverage)
  - Nonresponse bias (some subjects can’t be reached or don’t respond; TV studies are at about 60-80% nonresponse)
  - Response bias (questions are misleading; subject is lying; subject is giving wrong answers . . .)
- Convenience sampling is **bad** (usually nonrandom *and* biased in various ways)
- We’ll learn more about prediction error later (important!)

Back to controlling experiments

- A typical scenario:
  - Have a new drug
  - Divide subjects into 2 groups
  - One group gets the drug
  - One group (control) gets a fake drug (placebo, a different drug, . . .)
  - Compare results
- Watch out for the placebo effect (Beecher $\approx$ 1955):
  - A certain percentage of patients' condition improves by taking the placebo
  - Counter-studies show that this might not be all that true
  - In fact those studies show only that in specific cases there might be other conditions
  - Some form of placebo effect is really there, though we could argue about the percentage healed