Final Review

Coverage–comprehensive, with some emphasis post-midterm

pre-mid: ~B&T ch 1-2

post-mid: ~B&T ch 3,5,9, continuous, limits, mle, em, hypothesis testing.

all slides, hw, sols, non-supl reading on “Schedule & Reading” web page

Mechanics

closed book, aside from one page of notes (8.5 x 11, both sides, handwritten)

I’m more interested in setup and method than in numerical answers, so concentrate on giving a clear approach, perhaps including a terse English outline of your reasoning.

Corollary: calculators are probably irrelevant, but bring one to the exam if you want, just in case.

Format–similar to midterm:

T/F, multiple choice, problem-solving, explain, …

Story problems

General groaning, tooth-gnashing and head-banging

b&t chapters 1-2

see midterm review slides

especially 3.1–3.3; light coverage: 3.4–3.6

probability density function (pdf)

cdf as integral of pdf from $-\infty$; pdf as derivative thereof

expectation and variance

why does variance matter? a simple example: a random X arrives at a server, and chews up f(X) seconds of CPU time. If f(x) is a quadratic or cubic or exponential function, then randomly sampled X's in the right tail of the distribution can greatly inflate average CPU demand even if rare, so variance (and, more generally, the shape of the distribution) matters a lot, even for a fixed mean. Recall, in general, $E[f(X)] \neq f(E[X])$!

important examples know pdf and/or cdf, mean, variance of these

uniform, normal (incl Φ, "standardization"), exponential

tail bounds

- Markov
- Chebyshev
- Chernoff (lightly)

limit theorems

- weak/strong laws of large numbers
- central limit theorem

moment generating functions

- lightly - see ~2-3 slides in “limits” section; skim B&T 4.4 for more

likelihood, parameter estimation, MLE (b&t 9.1)

likelihood

- "likelihood" of observed data given a model
- usually just a product of probabilities (or densities: "$\lim_{\delta \to 0}$…"), by independence assumption
- a function of (unknown?) parameters of the model

parameter estimation

- if you know/assume the *form* of the model (e.g. normal, poisson,...), can you estimate the *parameters* based on observed data
- many ways, e.g.:

maximum likelihood estimators

- one way to do it–choose values of the parameters that maximize likelihood of observed data
- usual method – solve "∂/∂ param of (log) likelihood = 0" (and check for max not min, boundaries…)

confidence intervals

EM

- iterative algorithm trying to find MLE in situations that are analytically intractable
- usual framework: there are 0/1 *hidden variables* (e.g., from which component was this datum sampled) & problem would be much easier if they were known
- E-step: given rough parameter estimates, find expected values of hidden variables
- M-step: given rough expected values of hidden variables, find (updated) parameter estimates to maximize (expected) likelihood
- Algorithm: iterate above alternately until convergence

I have data, and 2 hypotheses about the process generating it. Which hypothesis is (more likely to be) correct?

Again, a very rich literature on this. Here consider the case of 2 *simple* hypotheses, e.g. $p = \frac{1}{2}$ vs $p = \frac{2}{3}$

One of the many approaches: the "Likelihood Ratio Test"

calculate: $$\frac{\text{likelihood of data under } \textit{alternate} \text{ hypothesis } H_1}{\text{likelihood of data under } \textit{null} \text{ hypothesis } H_0}$$

ratio > 1 favors alternate, < 1 favors null, etc.

type 1, type 2 error, α, β, etc. Of special interest: α = "significance" - prob of falsely rejecting null when it's true.

Neyman/Pearson: given these assumptions, LRT is *optimal*

significance testing (b&t 9.4)

As above:

- I have data, and 2 hypotheses about the process generating it. Which hypothesis is (more likely to be) correct?

But, consider *composite* hypotheses, e.g., $p = ½$ vs $p \neq ½$.

- Can't do likelihood for composite, so no easy LRT

But can often still evaluate *significance*: what is prob "*q*" of seeing data that cause you to falsely reject the null when it's true? Devise a summary statistic whose distribution you can calculate *under the null,* so you can estimate *q*. [Very often the stat. approx. follows normal- or t-dist. Thank you, CLT!]

p-values: smallest α allowing rejection; probability of generating this (or even less plausible) data *assuming the null is true*, **not** the probability that the null is false. [Note that "Null=T/F" is usually not a probabilistic question, so "prob that null is F" is a nonsensical question.]

Noise, uncertainty & variability are pervasive

Learning to model it, derive knowledge, and compute despite it are critical

E.g., knowing the mean is valuable, but two scenarios with the same mean and different variances can behave very differently in practice.

Stat 390/1	probability & statistics
CSE 427/8	computational biology
CSE 440/1	human/computer interaction
CSE 446	machine learning
CSE 472	computational linguistics
CSE 473	artificial intelligence

and others!

Please fill out the online course eval form:

https://uw.iasystem.org/survey/146020

Tell us what was useful, what was hard, what was fun, what we should do more/less of. Tell us the instructor was tall, handsome, witty, charming. Tell us it was the best course you have ever taken, beyond your wildest imaginings. Or not.

BY SUNDAY–last chance!

(Thanks!)

Thanks and Good Luck!

Hell's library

what to expect
on the final in
more detail

...