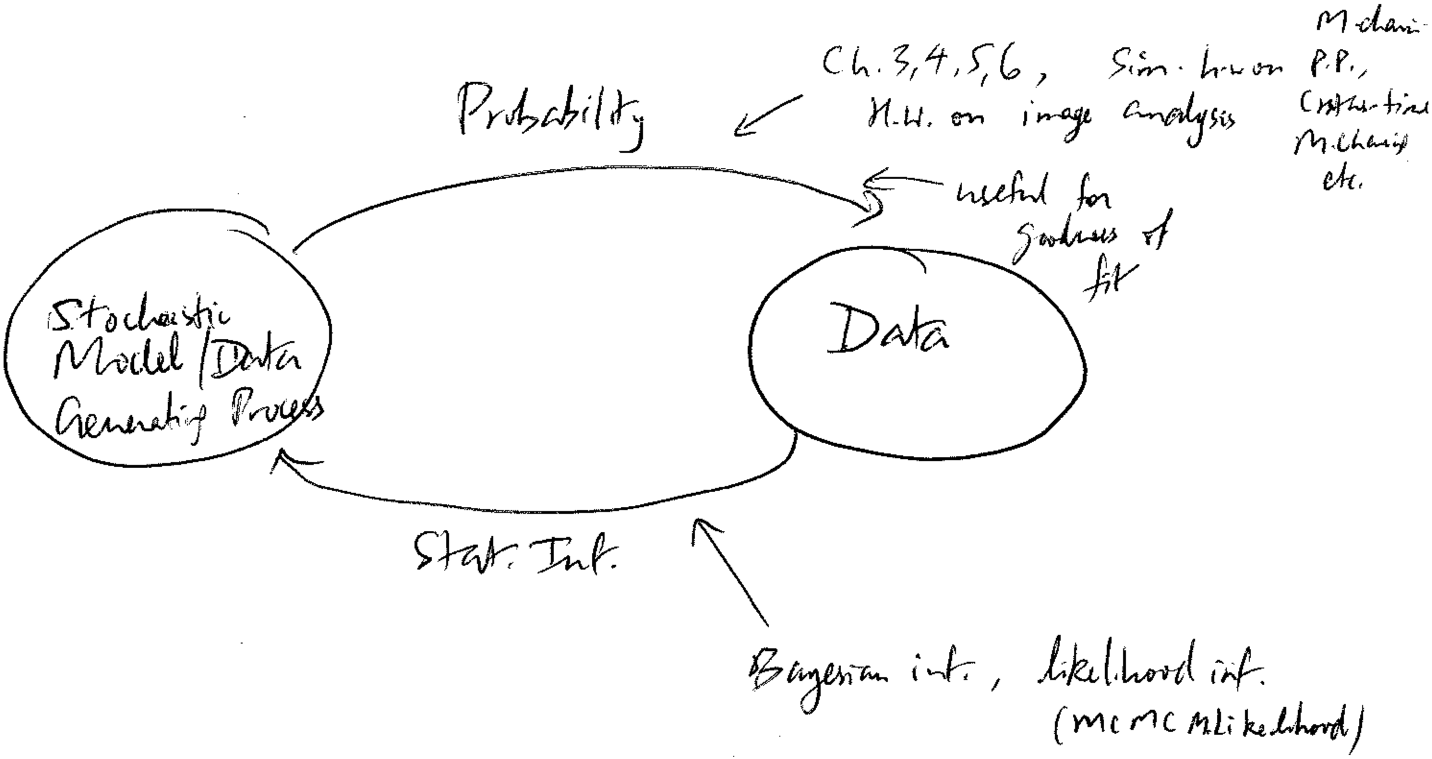
Overview of Prob., Stats., Stoch. Proc. & Monte Carlo



An overview of STAT 515

Ch. 3 Basic condtl prob., expectation : review
Using conditioning as a tool to simplify calculations : law of iterated expectations, iterated variance.
e.g. conditionally specified models, Bayesian models, calculating prob. in M. chains
proof for rejection sampling algorithm,
Gambler's ruin, coin tossing problems (toss until K heads in a row).

Ch. 4 Markov chains : discrete space, discrete time.
Classification of states / properties of states, chain.
Stationarity, ergodic thm / SLLN (immediate result of ergodic theory) ($X_1, X_2, \ldots$ stationary and ergodic w/ $E|X_1| < \infty$ then $\bar{X}_n \to E(X_1)$.)
Reversibility, Branching processes skip

Later: in context of MCMC : continuous space, discrete time
~~slightly~~ different but closely related classification of M chains,
Harris ergodicity, reversibility, ergodic thm / SLLN (crucial for Monte Carlo).

Ch. 5 Poisson processes : basics, connection to exponential r.v.s

Ch. 6 Continuous-time discrete space M. chains : generators,
transition probability ; birth-death processes.

Markov chains: useful for modeling physical processes.
Theory: help study behavior of processes; can help w/ inference (we did not talk about latter here).

Monte Carlo: general approach for estimating expectations
- can often convert problems into one involving expectations and obtain Monte Carlo solns.
  Rejection sampling: simple algorithm to generate samples
  Metropolis-Hastings: very general algorithm to generate samples $\approx$ from complicated / multivariate distribution using Markov chain theory.

Importance sampling: given samples (from another algorithm) from one distribution calculate expectations w.r.t. another distr. Very useful: tail probabilities, MCML, multiple expectations efficiently

Monte Carlo: allows for inference using very complicated models. eg. glms, normalizing function - exponential families etc.

Monte Carlo Max. Likelihood Example

Toy e.g.

$X_1, \ldots, X_n \overset{iid}{\sim} \text{Expon}(\beta = 2)$ $\quad (E(X_i) = 2)$ $\qquad f_\beta(x) = \frac{1}{\beta} e^{-x/\beta}$

Goal: find MLE of β.

Easy: maximize $\cancel{\ell(\beta)=}$ $\log \mathcal{L}(\beta ; \underset{\sim}{x}) = -n \log \beta - \frac{\sum_{i=1}^n x_i}{\beta}$

$$\frac{d}{d\beta} \log \mathcal{L}(\beta) = -\frac{n}{\beta} + \frac{\sum x_i}{\beta^2} = 0$$

$$\Rightarrow \hat{\beta} = \frac{\sum x_i}{n} = \bar{x}$$

(easy to see that $\frac{d^2}{d\beta^2} \ell(\hat\beta) < 0$)

Pretend normalizing constant is unknown.

Only know $h_\beta(x) = e^{-x/\beta}$

Can still ~~find~~ estimate MLE. Maximize $\ell(\beta) = \log \left\{ \frac{\mathcal{L}_\beta(\underset{\sim}{x})}{\mathcal{L}_\psi(\underset{\sim}{x})} \right\}$ for some constant ψ.

① Fix ψ and simulate $Y_1, \ldots, Y_M \sim f_\psi$

Can do this w/o knowing normalizing constants

(rej. sampling, MCMC etc.)

② $$\ell(\beta) = \log \left\{ \frac{h_\beta(\underset{\sim}{x})}{h_\psi(\underset{\sim}{x})} \right\} - n \log \left\{ E_{f_\psi} \left\{ \frac{h_\beta(\cdot)}{h_\psi(\cdot)} \right\} \right\}$$

So $$\hat{\ell}(\beta) = \underbrace{\log \left\{ \frac{h_\beta(\underset{\sim}{x})}{h_\psi(\underset{\sim}{x})} \right\}}_{\text{fully known}} - n \log \underbrace{\left\{ \frac{1}{M} \sum_{j=1}^{M} \frac{h_\beta(Y_j)}{h_\psi(Y_j)} \right\}}_{\text{MC estimate of ratio of normalizing constants}}$$

$\hat{\beta}$ that maximizes simulated likelihood $\hat{\ell}(\beta)$ is an estimate of MLE.

Non-toy example:

Image analysis : $X = (x_1, x_2, \ldots, x_N)$, $x_i \in \{-1, 1\}$
$N = \#$ pixels B/W image

Prob. distr. : $P(X = \underset{\sim}{x}) = \frac{1}{Z} \exp\{-2\phi\, U(\underset{\sim}{x})\}$, $\phi > 0$.

where $U(\underset{\sim}{x}) = \sum_{i \sim j} I(x_i \neq x_j)$ = # 'unlike' neighbors

$i \sim j$ mean i, j are neighbors

$Z = \sum_{\underset{\sim}{x}} \exp\{-2\phi\, U(\underset{\sim}{x})\}$, a summation over 2^N values

Hence, have $P(\underset{\sim}{X} = \underset{\sim}{x}) \propto \exp\{-2\phi\, U(\underset{\sim}{x})\}$

$\mathcal{L}(\underset{\sim}{X}; \phi) \propto \exp\{-2\phi\, U(\underset{\sim}{X})\}$ Want MLE for ϕ.

Other e.g.: network models, models for genetics problems, random field models for spatial data etc.

We can now fit models we like rather than models that are chosen for computational simplicity!

MCML for generalized linear mixed models
(and models w/ "missing data")

follow McCulloch (JASA 1997) and Geyer et al (2002).

$\underset{\sim}{Y}$: vector of observed data

$\underset{\sim}{u}$: vector of missing data or "random effects"

$\underset{\sim}{\theta}$: vector of parameters

Jt. model : $f_{\underset{\sim}{Y},\underset{\sim}{u}}(\underset{\sim}{y}, \underset{\sim}{u} ; \underset{\sim}{\theta})$

$$\mathcal{L}(\theta; \underset{\sim}{Y}) = \int f_{Y,u}(\underset{\sim}{y}, \underset{\sim}{u}; \underset{\sim}{\theta})\, d\underset{\sim}{u}$$

Integrate out or average over missing data / random effects.

$$\text{MLE for } \theta = \underset{\theta \in \Theta}{\operatorname{argmax}} \; \mathcal{L}(\theta; Y)$$

Approaches: Numerical methods, EM / Monte Carlo EM, MC Newton-Raphson.
, MCML ("Simulated max. likelihood")

Simple e.g. Logit-normal model

$$Y_{ij} \mid \underset{\sim}{u} \overset{\text{indep}}{\sim} \text{Ber}(p_{ij}) \qquad \begin{array}{ll} i = 1, \dots, n & \text{(individual within gp)} \\ j = 1, \dots q & \text{(group ids)} \end{array}$$

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta X_{ij} + u_j$$

X_{ij} : fixed, known covariates (could also generalize and have these be unknown/missing)

$$u_j \overset{\text{indep}}{\sim} N(0, \sigma^2).$$

$$\mathcal{L}(\beta, \sigma^2 ; \underset{\sim}{Y}) = \prod_{j=1}^{q} \int_{-\infty}^{\infty} \prod_{i=1}^{n} \left\{ \frac{\exp(Y_{ij}(\beta X_{ij} + u_j))}{1+\exp(\quad (\beta X_{ij} + u_j))} \cdot \frac{\exp(-u_j^2/2\sigma^2)}{(2\pi\sigma^2)^{1/2}} \right\} d u_j$$

Note: $\{ - \} = f(Y_{ij} \mid u_j ; \beta) \, f(u_j ; \sigma^2)$
$= f(Y_{ij}, u_j ; \beta, \sigma^2)$,

$$\prod_{i=1}^{n} f(Y_{ij}, u_j ; \beta, \sigma^2) = f(\underset{\sim}{Y}_{\cdot j}, u_j ; \beta, \sigma^2),$$

and $\displaystyle\int_{-\infty}^{\infty} f(\underset{\sim}{Y}_{\cdot j}, u_j ; \beta, \sigma^2) \, du_j = f(\underset{\sim}{Y}_{\cdot j} ; \beta, \sigma^2)$

Want MLEs $\hat{\beta}$, $\hat{\sigma}^2$.

For this simple e.g. possible to use numerical methods.

More general cases, may not be possible.

MC ML approach: (follow McCulloch '97 JASA)

$$\mathcal{L}(\underset{\sim}{\theta};\underset{\sim}{Y}) = \int f_{Y|u}(Y|u;\underset{\sim}{\theta}_1)\, f(u;\underset{\sim}{\theta}_2)\, d\underset{\sim}{u}$$

with $\underset{\sim}{\theta} = (\underset{\sim}{\theta}_1, \underset{\sim}{\theta}_2)$

$$= \int \frac{f_{Y|u}(Y|u;\underset{\sim}{\theta}_1) f(u;\underset{\sim}{\theta}_2)}{g_u(u)}\, g_u(u)\, d\underset{\sim}{u}$$

(where $g_u(u)$ is an importance function s.t.
$g_u(u) = 0 \Rightarrow f_{Y,u}(Y,u;\underset{\sim}{\theta}) = 0$.)

$$= E_{g_u}\left\{ \frac{f_{Y,u}(Y,u;\underset{\sim}{\theta})}{g_u(u)} \right\}$$

Estimate by Monte Carlo!

Generate $u^{(1)}, \ldots, u^{(M)} \sim g_u(u)$ by iid Monte Carlo or Markov chain (M-H alg.).

Then, $\mathcal{L}(\underset{\sim}{\theta};Y) \approx \frac{1}{M}\sum_{i=1}^{M} \frac{f_{Y,u}(Y,u^{(i)};\underset{\sim}{\theta})}{g_u(u^{(i)})} = \hat{\mathcal{L}}(\underset{\sim}{\theta};Y)$

$\hat{\ell}(\theta;\underset{\sim}{Y}) = \log \hat{\mathcal{L}}(\theta;Y)$

$\underset{\theta\in\Theta}{\arg\max}\ \hat{\ell}(\theta;Y)$ is approximation to MLE.

Can get $\nabla\hat{\ell}(\theta;Y)$, $\nabla^2\hat{\ell}(\theta;Y)$ by taking derivatives of above.

Choice of importance function g_u is critical.

Poor choice will result in very poor estimates even for large M.

Good idea to iterate: get MLE est., then find new importance fn, get new MLE estimate etc.

Final exam

2007: May 7 10:10am - 12pm 323E HH DEVE

2008: May 5

Covering all material from entire class

Calculator allowed.

2 Sheets of paper

1st half : like midterm

2nd " : theory, ideas related to Monte Carlo, MCMC.

Length : 1 hr 50 mins but exam ≤ 1.5 × midterm
1 hr. 15 mins

Overview of course

Overview of research problems