STT 465 (Fall, 2015): Bayesian Statistical Methods

⇒ Instructor: Gustavo de los Campos (Asoc. Prof., EPI-Biostat & Statistics)
Contact: gustavoc@msu.edu
Primary office: 909 Fee Rd. Room B637
Websites: http://quantgen.github.io

=> TA: Scott Manski (Email: manskisc@stt.msu.edu)

⇒ Office Hours: MW 9:00-10:00 AM (Wells, office TBA)

⇒ Course website: https://github.com/gdlc/stt465

⇒ Syllabus: see course website for info about the course, pre-requisites, grading, etc.

⇒ Pre-requisites, rules, exams & grading policy (see syllabus)

Tentative Schedule

Week	Day	Date	Chapter	Events
Week 1	W	2-Sep	Ch. 1 (Intro)	1st class
	M	7-Sep		No class
Week 2	W	9-Sep	Ch. 2 (Belief, Prob. & Exchangeability)	HW 1 posted
	M	14-Sep	Ch. 3 (One parameter models) Beta-Binomial & Poisson Models	
	W	16-Sep		HW1 due
	M	21-Sep	Ch. 4 (Monte Carlo Approximations)	
Week 3	W	23-Sep		HW 2 posted
	M	28-Sep	Ch. 5 (Normal Model)	
Week 4	W	30-Sep		HW2 due
	M	5-Oct	Ch. 6 (Gibbs Sampler & multiple linear regression)	
Week 5	W	7-Oct		HW 3 posted
	M	12-Oct	Ch. 7 (Multivariate Normal Model)	
Week 6	W	14-Oct		HW3 due
	M	19-Oct	Ch. 9 (Linear Regression, will touch on Ch. 8 as special case, students must read both Ch. 8 and Ch. 9)	
Week 7	W	21-Oct		
	M	26-Oct		
Week 8	W	28-Oct		Midterm
	M	2-Nov	Ch. 10 (Metropolis-Hastings algorithm)	
Week 9	W	4-Nov		Proposal due
	M	9-Nov	Ch. 12 (Threshold model)	
Week 10	W	11-Nov		
	M	16-Nov	Ch. 1 (Mixed models)	
Week 11	W	18-Nov		
	M	23-Nov		
Week 12	W	25-Nov		Final project due
	M	30-Nov		
Week 13	W	2-Dec		
	M	7-Dec		
Week 14	W	9-Dec		Final project presentations
	M	14-Dec		Final project presentations
Week 15	W	16-Dec		

Statistical Inference

⇒ Review of basic concepts:

- Population $Y=\{y_1,\ldots\}$ (may be finite or infinitely large).
- Sample We collect a sample of size n from the population $Y_s=\{y_{1,\ldots,}y_n)$

- Estimator: $\theta(Y_s)$

⇒ Inference: we make statements about population parameters based on the sampled data.

⇒ Two approaches:

- Classical (frequentist) inference: statements are based on the sampling distribution of the estimator over conceptual repeated sampling

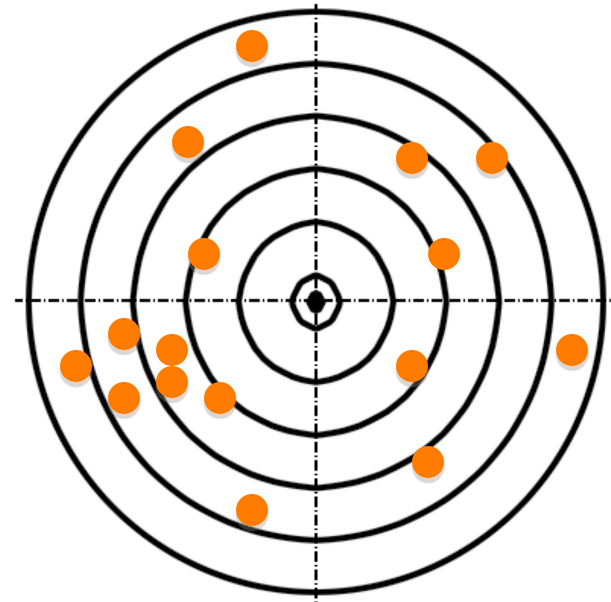
- Bayesian: statements are conditional on the observed data (the only sample we have drawn from the population).

Frequentist approach

$$\textit{Sampling Model} \quad \mathrm{p}\left(Y_S \mid \theta\right)$$

$$\textit{Sample } Y_s = \left\{y_i\right\} \quad \textit{Estimator } \hat{\theta}\left(Y_s\right)$$

Sampling Distribution of Estimates



$$Bias\left(\hat{\theta}\right) = E\left(\hat{\theta}\right) - \theta$$

$$Var\left(\hat{\theta}\right) = E\left(\hat{\theta} - E\left(\hat{\theta}\right)\right)^2$$

$$MSE\left(\hat{\theta}\right) = \left[E\left(\hat{\theta}\right) - \theta\right]^2 + E\left[\hat{\theta} - E\left(\hat{\theta}\right)\right]^2$$

Squared-Bias

Variance

Let's look at an example:

(Binomial Sampling)

Sampling from binomial

- Code: example_1.md

- Discuss:
  - Binomial model
  - Maximum Likelihood estimator.
  - Expected value and variance of the estimator.
  - Compare the bias and variance with Monte Carlo estimates of those quantities.

Bayesian approach

Elements of a Bayesian Model

(1) Sampling model: $\mathrm{p}\left(Y|\theta\right)$

- Describes the probability of the data
- Usually indexed by a set of parameters

(2) Prior distribution of unknowns (typically model parameters) $\mathrm{p}\left(\theta\right)$

(3) Posterior distribution: the probability of the parameters given the sample. $\mathrm{p}\left(\theta|Y\right)$

Bayes Theorem

Preliminaries

- The joint distribution is the product of the marginal times the conditional distribution

$$p(A,B) = p(A)P(B|A)$$
$$= p(B)p(A|B)$$

- Law of total probability

$$p(B) = \sum_j p(A_j)P(B|A_j)$$

Bayes Theorem

$$p(\theta,Y) = p(Y)p(\theta|Y) = p(\theta)p(Y|\theta)$$

$$p(\theta|Y) = \frac{p(\theta)p(Y|\theta)}{p(Y)}$$

Beta-Binomial Model

- Sampling model (iid Bernoulli)

$$\mathrm{p}\left(y_1, y_2, \ldots, y_n | \theta\right) = \prod_i \theta^{y_i} \left(1-\theta\right)^{1-y_i} = \theta^{n_1} \left(1-\theta\right)^{n_0}$$

- Prior (Beta)

$$\mathrm{p}\left(\theta\right) = \frac{\theta^{\alpha-1}\left(1-\theta\right)^{\beta-1}}{B(\alpha,\beta)} \propto \theta^{\alpha-1}\left(1-\theta\right)^{\beta-1}$$

- Posterior Distribution (will discuss later on this result in detail)

$$\mathrm{p}\left(\theta | y\right) = \frac{\mathrm{p}\left(y|\theta\right)\mathrm{p}\left(\theta\right)}{\mathrm{p}\left(y\right)} \propto \mathrm{p}\left(y|\theta\right)\mathrm{p}\left(\theta\right) \propto \prod_i \theta^{\alpha-1}\left(1-\theta\right)^{\beta-1} \prod_i \theta^{y_i}\left(1-\theta\right)^{1-y_i}$$

$$= \theta^{[n_1+\alpha-1]}\left(1-\theta\right)^{[n_0+\beta-1]}$$

- Maximum a-posteriori

$$\tilde{\theta} = \frac{[n_1+\alpha-1]}{[n_1+\alpha-1]+[n_0+\beta-1]} = \frac{[n_1+\alpha-1]}{n+\alpha+\beta-2}$$

Beta-Binomial Model

- Prior Mean

$$E(\theta) = \frac{\alpha}{\alpha + \beta}$$

- MLE

$$\hat{\theta}_{MLE} = \frac{n_1}{n_1 + n_2}$$

- Posterior Mean

$$\hat{\theta} = \frac{n_1 + \alpha}{n + \alpha + \beta} = \left[\frac{n}{n + n + \alpha}\right]\frac{n_1}{n} + \left[\frac{\alpha + \beta}{n + \alpha + \beta}\right]\frac{\alpha}{\alpha + \beta}$$

- Bayesian estimates are ‘shrunken’ estimates. You can view them as MLE estimates shrunken towards the prior mode/mean. Shrinkage reduces the variance of the estimator, at the expense of potentially introducing more bias. The extent of shrinkage depends on how informative the prior is and how informative is the likelihood (sample size). If parameters are identified at the likelihood, as sample size increases Bayesian estimates converge to MLE.