Lec10 3/7/16 Math 390.03-02

1

Concepts Covered for Midterm I (i.e. thus far in this class)

General

Each building block in detail

MLEs, $P(\theta|x) = \frac{P(x|\theta)\,P(\theta)}{P(x)}$, $\underbrace{\frac{P(\theta_1|x)}{P(\theta_2|x)}}_{\text{a posteriori odds}} = \underbrace{\frac{P(x|\theta_1)}{P(x|\theta_2)}}_{\text{Bayes Factor}}\ \underbrace{\frac{P(\theta_1)}{P(\theta_2)}}_{\text{a priori odds}}$, exchangeability vs iid

Bayesian ~~[illegible]~~ CLT,
$Var[\theta|x]$ vs. $Var[\theta]$

$\hat{\theta}_{MMSE}$, $\hat{\theta}_{MAE}$, $\hat{\theta}_{MAP}$, principle of indifference → troubles...

posterior predictive dist $P(X^*|x)$ X^* can be 1-dim. or m-dim.

compare to frequentist stuff

credible intervals, hypothesis tests (1-sided), 2-sided, 2 comps,

Bayes Factor ~~comparisons~~, shrinkage estimation, conjugacy,
uninformative priors, improper priors, Jeffrey's priors

Specific

$P(X|\theta) = Bin(n,\theta)$, $P(\theta) = Beta(\alpha,\beta)$, What is the beta distr? Shapes?
$P(\theta|X) = Beta$ ... Law of succession
$U(0,1) = Beta(1,1)$, $\alpha < 1$ or > 1, $\beta < 1$ or > 1, pseudo counts, prior exp, variance

$P(X^*|X) = $ BetaBin, $P(X) = $ BetaBin

→ mixture models, kernel functions, model checks

MIDTERM 2

⇓

Estimating Batting Avgs in pro baseball

$$BA: \hat{\theta} = \frac{\#\text{ HITS}}{\#\text{ At BATS}} = \frac{X}{n} = \hat{\theta}_{MLE} \text{ for a}$$

$X_1, \ldots, X_n \overset{iid}{\sim} \text{Bern}(\theta)$

model $\Leftrightarrow X \sim \text{Bin}(n, \theta)$

Who are the best BA belong to? $\geq .300$ is considered great

$$\frac{1}{1} = 1.000$$
$$\frac{2}{2} = 1.000$$
$$\frac{1}{2} = 0.500$$

they can't be the best! Not enough data...

Who are the worst?

$$\frac{0}{1} = 0$$
$$\frac{0}{2} = 0$$

they can't be the worst... Not enough data!!

Solution: SHRINK

$X \sim \text{Bin}(n, \theta)$, $\theta \sim \text{Beta}(\alpha, \beta)$

$\theta \mid X \sim \text{Beta}(x+\alpha, n-x+\beta)$

$$\hat{\theta}_{MMSE} := E(\theta \mid x) = \frac{x+\alpha}{n+\alpha+\beta} = \underbrace{\rho}_{\frac{\alpha+\beta}{\alpha+\beta+n}} E(\theta) + (1-\rho)\,\hat{\theta}_{MLE}$$

If I don't have a lot of data... I get a "better" estimate since my prior is smart!

But how to pick prior????

Beta$(0,0)$, Beta$(\frac{1}{2},\frac{1}{2})$, Beta$(1,1)$

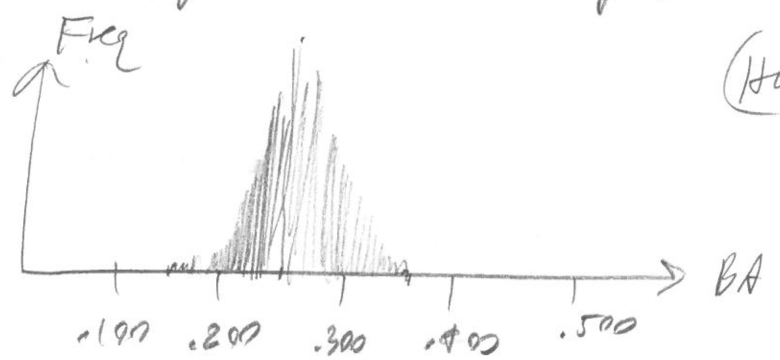
not "smart"... they're "objective"... goal is to **not** be subjective and let only the data speak for itself.

How to design a subjective prior?

Not smart here...

Let's look at all 7,256 baseball players... career BA's

And let's ignore all career players with < 500 at-bats:



(Histogram)

let's let these $\hat{\theta}_{MLE}$'s inform our Beta(α,β)

Fit a Beta(α,β) to this. $\hat{\alpha}_{MLE} = 78.7$, $\hat{\beta}_{MLE} = 224.8$

Let $\theta \sim$ Beta$(78.7, 224.8)$ be your prior with $E(\theta) = \frac{78.7}{303.5} = .259$

Equivalent to seeing 77.7 Hits in 302.5 at bats $\Longrightarrow$ **STRONG** not objective

$\Rightarrow \theta|X \sim \text{Beta}(x+78.7,\ n-x+303.5) \Rightarrow \hat{\theta}_{MMSE} = \dfrac{x+78.7}{n+303.5}$

$= \dfrac{303.5}{303.5+n}(.259) + \dfrac{n}{303.5+n}\bar{X}$

So for a $\frac{1}{1}$ BA...

$$\hat{\theta} = \underbrace{\frac{303.5}{303.5+1}}_{99.7\%}(.259) + \underbrace{\frac{1}{303.5+1}}_{.3\%}(1)$$

$$= .258 + .003 \qquad = .262$$

But for $\frac{167}{1000}$

$$\hat{\theta} = \underbrace{\frac{303.5}{303.5+1000}}_{23.3\%}(.259) + \underbrace{\frac{1000}{303.5+1000}}_{76.7\%} .167$$

$$= .060 + .128 = .188$$

"extraordinary outliers require extraordinary evidence"

But... we used data to create a prior

"Empirical Bayesian Model"
or "Empirical Bayes"

Use data to estimate hyperparams then estimate params...

~~We will create ... first...~~

New Model: $X_1, \dots X_n \overset{iid}{\sim} \text{Geom}(\theta)$

Recall: 0 - - - - - 0 1 $\quad$ $\text{Supp}[X] = \mathbb{N}_0$, Param Space: $\theta \in (0,1)$

Run iid Bernoullis until you achieve a success.

The # of failures is the realization.

$$p(x;\theta) = \underbrace{(1-\theta)^{x}}_{\substack{x \text{ failed} \\ \text{experiments}}} \quad \underbrace{\theta}_{\substack{\text{one success} \\ \text{on the } x\text{th} \\ \text{experiment}}}$$

$x^{a-b} = x^a x^{-b} = \frac{x^a}{x^b}$

$$P(X_1, \dots, X_n \mid \theta) = \prod_{i=1}^{n} (1-\theta)^{x_i} \theta = \theta^n (1-\theta)^{\Sigma x_i} = \theta^n (1-\theta)^{\Sigma x_i}$$

But

$$P(\theta \mid X_1, \dots, X_n) \propto P(X_1, \dots, X_n \mid \theta) \, P(\theta)$$

$$\underbrace{\theta^n (1-\theta)^{\Sigma x_i}}_{\substack{\text{seen this} \\ \text{form before?}}}$$

Kernel is Beta $\Rightarrow$ Conjugate is Beta! AGAIN!!!

$X_1,\dots,X_n \overset{\text{each}}{\sim} \text{Geom}(\theta), \quad \theta \sim \text{Beta}(\alpha,\beta)$

all data

$$p(\theta \mid X) \propto p(X \mid \theta)\, p(\theta)$$

$$= \theta^n (1-\theta)^{\sum x_i} \, \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$\propto \theta^{n+\alpha-1}(1-\theta)^{\sum x_i + \beta - 1}$$

$$\propto \text{Beta}\left(n+\alpha, \sum x_i + \beta\right) \quad \text{(always proper)}$$

$\hat{\theta}_{MMSE} = \frac{n+\alpha}{n+\alpha+\sum x_i+\beta}$ $\quad \hat{\theta}_{MAE}$ need computer $\quad \hat{\theta}_{MAP} = \frac{n+\alpha-1}{n+\alpha+\sum x_i+\beta-2}$

MLE

$$\ell(\theta;x) = n\ln\theta + \sum x_i \ln(1-\theta)$$

$$\ell'(\theta;x) = \frac{n}{\theta} - \frac{\sum x_i}{1-\theta} \overset{\text{set}}{=} 0 \Rightarrow \frac{n}{\theta} = \frac{\sum x_i}{1-\theta} \Rightarrow n - n\theta = \theta \sum x_i$$
$$\Rightarrow n = \theta\left(\sum x_i + n\right) \Rightarrow \hat{\theta}_{MLE} = \frac{n}{n+\sum x_i} = \frac{1}{1+\bar{x}}$$

If $\alpha=\beta=0$ (Haldane prior)

$$\hat{\theta}_{MMSE} = \hat{\theta}_{MLE}$$

If $\alpha=\beta=1$ (Laplace / Uniform)

$$\hat{\theta}_{MAP} = \hat{\theta}_{MLE}$$

Jeffreys prior:

$$\ell''(\theta;x) = -\frac{n}{\theta^2} - \frac{\sum x_i}{(1-\theta)^2}$$

$$I(\theta) = E\left[-\ell''(\theta;x)\right] = E\left[\frac{n}{\theta^2} + \frac{\sum x_i}{(1-\theta)^2}\right]$$

$$\to \frac{n}{\theta^2} + \frac{n E[X_i]}{(1-\theta)^2}$$

$X \sim \text{Geom}(p)$, $\quad E(X) = \frac{1-p}{p}$

$$= \frac{n}{\theta^2} + \frac{n\left(\frac{1-\theta}{\theta}\right)}{(1-\theta)^2}$$

$$= n\left(\frac{1}{\theta^2} + \frac{1}{\theta(1-\theta)}\right)$$

$$= n\left(\frac{(1-\theta) + (\theta)}{\theta^2(1-\theta)}\right) = \frac{n}{\theta^2(1-\theta)}$$

$$\Rightarrow j(\theta) \propto \sqrt{\frac{1}{\theta^2(1-\theta)}} = \theta^{-1}(1-\theta)^{-\frac{1}{2}}$$

$$\Rightarrow j(\theta) = \text{Beta}\left(0, \tfrac{1}{2}\right)$$ Improper!

Shrinkage

$$\hat{\theta}_{MMSE} = \frac{n+\alpha}{n+\alpha+\sum x_i+\beta} \neq \rho E(\theta) + (1-\rho)\hat{\theta}_{MLE}$$ Beta...

NOT COVERED

$$\hat{\theta}_{MMSE}^{-1} = \frac{n+\alpha+\sum x_i+\beta}{n+\alpha} = \frac{n+\sum x_i}{n+\alpha}\,\frac{n}{n} + \frac{\alpha+\beta}{n+\alpha}\,\frac{\alpha}{\alpha}$$

$$= \underbrace{\frac{n+\sum x_i}{n}}_{\hat{\theta}_{MLE}^{-1}}\underbrace{\frac{n}{n+\alpha}}_{1-\rho} + \underbrace{\frac{\alpha+\beta}{\alpha}}_{E(\theta)^{-1}}\underbrace{\frac{\alpha}{n+\alpha}}_{\rho}$$

$$= \rho E(\theta)^{-1} + (1-\rho)\hat{\theta}_{MLE}^{-1}$$

$$\frac{1}{\hat{\theta}_{MMSE}} = \rho\frac{1}{E(\theta)} + (1-\rho)\frac{1}{\hat{\theta}_{MLE}}$$

weighted harmonic mean shrinkage ... weird!!!!

---

pr/post. pred. distr? Wait...

$$X_1, \ldots, X_n \overset{iid}{\sim} \text{NegBin}(r, \theta)$$ Same as Bern exp

$Supp[X] = \mathbb{N}_0$ param space $r \in \mathbb{N}$, $\theta \in (0,1)$

wait until you get r successes. r is known, x is # failures

$$p(\theta \mid X) \propto p(X \mid \theta)\, p(\theta)$$

$$\prod_{i=1}^n \binom{x_i+r-1}{x_i} \theta^r (1-\theta)^{x_i} \quad p(\theta)$$

$$\left(\prod_{i=1}^n \binom{x_i+r-1}{x_i}\right) \theta^{rn} (1-\theta)^{\sum x_i} \left(p(\theta)\right)$$

$$\propto \theta^{rn}(1-\theta)^{\sum x_i}\left(p(\theta)\right)$$ Beta Kernel again!!!

$\theta \sim \text{Beta}(\alpha, \beta)$

$\theta \mid X \sim \text{Beta}(rn + \alpha, \ \sum x_i + \beta)$

Always proper!

Estimators ... HW ...

Jeffreys Prior ... HW ... [MA]

Shrinkage ... HW ... [MA]

prior/post pred distr ... the next obs, x^* ...

$$P(X^* \mid X) = \int_\Theta P(X^* \mid \theta)\, P(\theta \mid X)\, d\theta = \int_0^1 \binom{x+r-1}{x} \theta^r (1-\theta)^x \cdot \frac{1}{B(rn+\alpha, \sum x_i + \beta)} \theta^{rn+\alpha} (1-\theta)^{\sum x_i + \beta} \, d\theta$$

MATH ... LOTS ...

$$= \text{BetaNegBin}(r, \ rn+\alpha, \ \sum x_i + \beta)$$

$$Y \sim \text{BetaNegBin}(r, \alpha, \beta) := \frac{\Gamma(r+x)}{x!\,\Gamma(r)} \frac{B(\alpha + r, \beta + x)}{B(\alpha, \beta)}$$

$\text{BetaGeom}(\alpha, \beta) :=$

(special case $r=1$)

HW exercise: find PDF [MA]

post pred. distr for $X_1, \ldots, X_n \overset{iid}{\sim} \text{Geom}(\theta)$

$\theta \sim \text{Beta}(\alpha, \beta)$

model