Lecture 11 3/21/17 Math 341

Review from last time ...

$X_1, \ldots, X_n \mid \theta \overset{iid}{\sim} \text{Geom}(\theta) \Rightarrow \hat{\theta}_{MLE} = \ldots = \frac{1}{1+\bar{x}}$ see previous lecture

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

$\theta \mid X_1 \ldots X_n \sim \text{Beta}(n+\alpha, \sum x_i + \beta) \Rightarrow \hat{\theta}_{MMSE} = \frac{n+\alpha}{n+\alpha+\sum x_i + \beta}$

α : # pseudosuccesses

β : # pseudofailures universal

Uninformative priors

Haldane : $\theta \sim \text{Beta}(0,0) \Rightarrow \hat{\theta}_{MMSE} = \frac{n}{n+\sum x_i} = \frac{1}{1+\frac{\sum x_i}{n}} = \frac{1}{1+\bar{x}} = \hat{\theta}_{MLE}$

Laplace : $\theta \sim \text{Beta}(1,1) \Rightarrow \hat{\theta}_{MMSE} = \frac{n+1}{n+1+\sum x_i + 1} = \frac{1}{1+\frac{\sum x_i + 1}{n+1}} \neq 1$

Jeffreys : $\theta \sim \text{Beta}(0, \frac{1}{2}) \Rightarrow \hat{\theta}_{MMSE} = \frac{n}{n+\sum x_i + \frac{1}{2}}$

$\approx \bar{x}$ but $\bar{x}$ shrunk towards $E(\theta) = 0.5$

$= \frac{1}{1+\frac{\sum x_i + \frac{1}{2}}{n}} \neq 1$

same idea

$\theta \sim U(0,1)$

$$\hat{\theta}_{mmse} = \frac{n+1}{n+1+\sum x_i+1} = \frac{1}{1+\frac{\sum x_i+1}{n+1}}$$

$\bar{x}'$ shrunk toward $E(\theta)$

$\theta \sim Beta(0,0)$

$$\hat{\theta}_{mmse} = \frac{n}{n+\sum x} = \frac{1}{1+\bar{x}} = \hat{\theta}_{MLE}$$ (same as before)

Is there shrinkage?

$$\frac{1}{\hat{\theta}_{mmse}} = \frac{n+\alpha+\sum x_i+\beta}{n+\alpha} = \frac{\alpha+\beta}{n+\alpha}\cdot\frac{\alpha}{\alpha} + \frac{n+\sum x_i}{n+\alpha}\cdot\frac{n}{n} = \frac{\alpha+\beta}{\alpha}\cdot\underbrace{\frac{\alpha}{n+\alpha}}_{\rho} + \overset{1+\bar{x}}{\frac{n+\sum x_i}{n}}\cdot\underbrace{\frac{n}{n+\alpha}}_{1-\rho}$$

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

this is shrinkage but a bit different from before...

Jeffreys prior?

$$p(\theta) \propto \sqrt{J(\theta)} = \sqrt{\frac{1}{n(1-\theta)\theta^2}} \propto \theta^{-1}(1-\theta)^{-\frac{1}{2}} \propto Beta\left(0, \frac{1}{2}\right)$$ Improper!

Post. Pred. Distr.

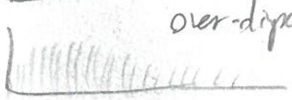
$$p(x^*|x) = \int_\Theta p(x^*|\theta)\,p(\theta|x)\,d\theta = \int_0^1 (1-\theta)^{x^*}\theta\,\frac{1}{B(n+\alpha, \sum x_i+\beta)}\,\theta^{n+\alpha-1}(1-\theta)^{\sum x_i+\beta-1}\,d\theta$$

for $n^*=1$

$$= \frac{1}{B(n+\alpha, \sum x_i+\beta)}\int_0^1 \theta^{n+\alpha+1-1}(1-\theta)^{x^*+\sum x_i+\beta-1}\,d\theta = \frac{B(n+\alpha+1,\ x^*+\sum x_i+\beta)}{B(n+\alpha, \sum x_i+\beta)}$$

geometric

$$E(x^*|x) = \frac{\sum x_i+\beta}{n+\alpha-1}$$



over-dispersed geometric

BetaGeom Distr

$BetaGeom(n+\alpha, \sum x_i+\beta) :=$

Not covered in class

$X_1, \ldots, X_n | \theta \overset{iid}{\sim} \text{NegBin}(r, \theta)$ fixed (r)

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

$\theta | X_1, \ldots, X_n \sim \text{Beta}(rn + \alpha, \sum x_i + \beta)$

$n^* = 1$, $X^* | X \sim \text{BetaNegBin}$

$X \sim \text{Bin}(n, \theta) := \binom{n}{x} \theta^x (1-\theta)^{n-x}$ $\quad \text{Supp}(X) = \{0, \ldots, n\}$, $\theta \in \Theta = (0, 1)$

if $n \to \infty$, $\theta \to 0$ s.t. $\lambda = n\theta$

$$\lim_{n \to \infty} \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} = \frac{\lambda^x}{x!} \lim \cancel{\frac{n \cdot n \cdots}{(n-x)(n-x+1) \cdots 1 \cdot n \cdot n \cdots}} \to 1$$

$$\cdot \lim \left(1 - \frac{\lambda}{n}\right)^n \cdot \lim \cancel{\left(1 - \frac{\lambda}{n}\right)^x} \to 0$$

$$= \frac{\lambda^x e^{-\lambda}}{x!}$$

$\Downarrow$

$$X \sim \text{Poisson}(\theta) := \frac{e^{-\theta} \theta^x}{x!}$$

$\text{Supp}(X) = \{0, \ldots\} = \mathbb{N}_0$, $\theta \in \Theta \in (0, \infty)$

differ θ!!!

$E[X] = \theta$

$Var(X) = \theta$

$$P(\theta | X) \propto P(X|\theta) P(\theta) = \frac{e^{-\theta} \theta^x}{x!} P(\theta) \propto e^{-\theta} \theta^x P(\theta)$$

$P(\theta)$ should have the form...

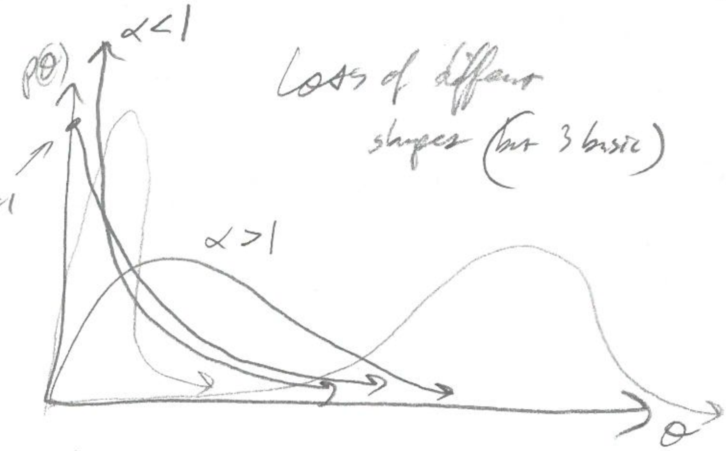
$$P(\theta) \propto e^{-b\theta} \theta^a$$

Integrate and see $\quad P(\theta) = \frac{b^{a+1}}{\Gamma(a+1)} e^{-b\theta} \theta^a \quad$ this is called the gamma distr.

Usually its parameterized via $\quad \theta \sim \text{Gamma}(\alpha, \beta) := \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta\theta} \theta^{\alpha - 1}$

$Supp(\theta) = (0, \infty)$ exactly what we want

param space $\alpha, \beta > 0$ just as before



Lots of different shapes (but 3 basic)

$$\left.\begin{aligned} E(\theta) &= \frac{\alpha}{\beta} \\ Var(\theta) &= \frac{\alpha}{\beta^2} \\ Mode(\theta) &= \frac{\alpha-1}{\beta} \text{ if } \alpha \geq 1 \end{aligned}\right\} \text{calculus exercises}$$

$Med(\theta) = qgamma(0.5, \alpha, \beta)$ (i.e., no closed form)

$$p(\theta|x) \propto p(x|\theta)\,p(\theta) = \frac{e^{-\theta}\theta^x}{x!}\frac{\beta^\alpha}{\Gamma(\alpha)}e^{-\beta\theta}\theta^{\alpha-1} \propto e^{-\theta}\theta^x e^{-\beta\theta}\theta^{\alpha-1} = e^{-(\beta+1)\theta}\theta^{x+\alpha-1}$$

$X|\theta \sim Poisson(\theta)$

$\theta \sim Gamma(\alpha, \beta)$

$\propto Gamma(x+\alpha, \beta+1)$

$\theta|X \sim Gamma(x+\alpha, \beta+1)$

"gamma is conjugate prior for Poisson likelihood"

Now lets say you have n iid Poissons

$X_1, \ldots, X_n | \theta \overset{iid}{\sim} Poisson(\theta)$

$\theta \sim Gamma(\alpha, \beta)$

$\theta | X_1, \ldots, X_n \sim ?$

$$p(\theta|x) \propto p(x|\theta)\,p(\theta) = \left(\prod_{i=1}^n \frac{e^{-\theta}\theta^{x_i}}{x_i!}\right)\left(\frac{\beta^\alpha}{\Gamma(\alpha)}e^{-\beta\theta}\theta^{\alpha-1}\right) = \frac{e^{-\sum_{i=1}^n \theta_i}\,\theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}\frac{\beta^\alpha}{\Gamma(\alpha)}e^{-\beta\theta}\theta^{\alpha-1}$$

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

Interpretations

α = sum total of successes seen previously

β = # of pseudo trials performed

$E(\theta) = \frac{\alpha}{\beta}$ = avg # of successes / trial

$\hat{\theta}_{MMSE} = \frac{\sum x_i + \alpha}{n+\beta}$, $\hat{\theta}_{MAE} = \text{qgamma}(0.5, \sum x_i + \alpha, n+\beta)$, $\hat{\theta}_{MAP} = \frac{\sum x_i + \alpha - 1}{n+\beta} \approx \hat{\theta}_{MMSE}$

Can we say $\theta \sim U$? No..

But... what if $P(\theta) \propto 1$ Improper since $\int_0^\infty c \, d\theta = \infty$

What happens?

$P(\theta|x) \propto P(x|\theta) P(\theta) \propto e^{-n\theta} \theta^{\sum x_i +1-1} \propto \text{Gamma}(\sum x_i +1, n)$ (always proper)

$\theta \sim \text{Gamma}(1, 0) \propto 1$ Laplace prior (improper)

What is MLE? Back to math stats...

The Haldane prior would be Gamma(0,0) which would yield a posterior of Gamma(sum x_i, n) which would be improper for all x_i = 0.

$$\mathcal{L}(\theta; x) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!} = \frac{e^{-n\theta}\theta^{\sum x_i}}{\prod x_i!}$$

$$\ell(\theta; x) = -n\theta + \sum x_i \ln(\theta) - \ln\left(\prod x_i!\right)$$

$$\ell'(\theta; x) = -n + \frac{\sum x_i}{\theta} \overset{set}{=} 0 \Rightarrow \frac{\sum x_i}{\theta} = n \Rightarrow \boxed{\hat{\theta}_{ML} = \bar{x}}$$

Jeffrey's Prior?

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

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

$$P(\theta) \propto \sqrt{I(\theta)} = \sqrt{\frac{n}{\theta}} \propto \sqrt{\frac{1}{\theta}} = \theta^{-\frac{1}{2}} \propto \text{Gamma}\left(\frac{1}{2}, 0\right)$$

Same idea as before.. see 0 trials ... but know there's 0.5 successes somewhere!

So two choices for uninformative prior: Haldane, Jeffreys. Jeffreys leads to proper post. always.

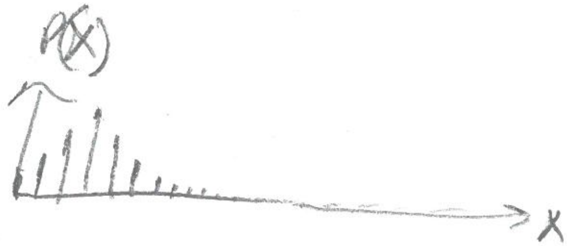
shrinkage? Yes, sensible weighted arithmetic avg shrinkage

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

$$\lim_{n\to\infty} \rho = 0 \checkmark$$

Poisson pred. distr.

$X|\theta \sim \text{Poisson}(\theta)$



Now θ is unknown ... so this will be "dispersed"

the beta-binomial is the "dispersed binomial"

the beta geometric is the "dispersed geometric"

what is a "dispersed Poisson"?

$$P(X^*|X) = \int_{\Theta} P(X^*|\theta)\, P(\theta|X)\, d\theta \qquad \text{for } n^* = 1 \ldots$$

$$= \int_0^\infty \left(\frac{e^{-\theta}\theta^{x^*}}{x^*!}\right)\left(\frac{\beta'^{\alpha'}}{\Gamma(\alpha')} e^{-\beta'\theta}\theta^{\alpha'-1}\right) d\theta$$

$$= \frac{\beta'^{\alpha'}}{\Gamma(\alpha')\, x^*!} \int_0^\infty \underbrace{e^{-(\beta'+1)\theta}\,\theta^{x^*+\alpha'-1}}_{\substack{\text{kernel}\\ \text{of}\\ \text{Gamma}(x^*+\alpha',\ \beta'+1)}} d\theta$$