Lecture 13 3/23/17 Math 341

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

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

$$\theta | X_1, \ldots, X_n \sim \text{Gamma}(\underbrace{\sum x_i + \alpha}_{\alpha'}, \underbrace{n + \beta}_{\beta'})$$

Posterior Predictive Distr.

$$P(x^* | \underset{x_1, \ldots, x_n}{x}) = \int_{\Theta} P(x^* | \theta) P(\theta | x) \, d\theta$$

for $n^* = 1$

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

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

We can now put in the consts. of integration!

$$\frac{1}{c} \cdot c$$

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

$$= \left(\frac{\beta'}{(\beta'+1)}\right)^{\alpha'} \left(\frac{1}{\beta'+1}\right)^{x^*} \frac{\Gamma(x^*+\alpha')}{x^*!\, \Gamma(\alpha')}$$

Note: $\frac{\beta'}{\beta'+1} \in (0,1)$ let $p := \frac{\beta'}{\beta'+1} \Rightarrow 1-p = \frac{1}{\beta'+1}$

$$= \frac{\Gamma(x^*+\alpha')}{x^*\, \Gamma(\alpha')} (1-p)^{x^*} p^{\alpha'} = \text{ExtNegBin}(\alpha', p)$$ "fractional successes"

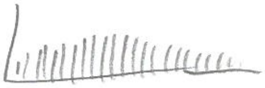
If $\alpha' \in \mathbb{N}$ note $x^* \in \mathbb{N}$ so... $\Gamma(n) = (n-1)!$

$$= \frac{(x^*+\alpha'-1)!}{x^*!\,(\alpha'-1)!} (1-p)^{x^*} p^{\alpha'} = \binom{x^*+\alpha'-1}{x^*} (1-p)^{x^*} p^{\alpha'} = \text{Negbin}(\alpha', p)$$

# of failures before α' successes where each experiment $\stackrel{iid}{\sim}$ Bern(p)

Poisson dispersal is a neg-binomial! $\cancel{E[x^*|x] = \frac{pr}{1-p} = \frac{\frac{n+\beta}{n+\beta+1}(\sum x_i+\alpha)}{\frac{1}{n+\beta+1}} = (n+\beta)}$ $\cancel{Var[x^*|x] = \frac{pr}{(1-p)^2}}$ $\cancel{\frac{\left(\frac{n+\beta}{n+\beta+1}\right)(\sum x_i+\alpha)}{\frac{1}{(n+\beta+1)^2}}}$

$$\Rightarrow x^*|x \sim \text{Negbin}\left(\alpha', \frac{\beta'}{\beta'+1}\right) = \text{NegBin}\left(\sum x_i + \alpha, \frac{n+\beta}{n+\beta+1}\right)$$



mixed of

$$X|\theta \sim \text{Gamma}(1,\theta) := \frac{\theta^1}{\underbrace{\Gamma(1)}_{1}} e^{-\theta(x)} \underbrace{\theta^{1-1}}_{1} = \theta e^{-\theta x}$$

$$= \text{Exp}(\theta)$$

Exp(θ) is a special case of the Gamma

$$p(\theta|x) \propto p(x|\theta)\,p(\theta) = \underbrace{\theta e^{-\theta x}}_{\text{Kernel of gamma}} \, \underset{\uparrow \text{ model}}{p(\theta)}$$

$p(\theta) = \text{Gamma}(\alpha,\beta)$ $\Rightarrow$ gamma is conjugate model for Exp Model

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

$$\propto \text{Gamma}(\alpha+1, \beta+1)$$

$$\Rightarrow \hat{\theta}_{MMSE} = \frac{\alpha+1}{\beta+1}$$

Jeffrey's prior? HW...

$r \in \mathbb{N}$ and <u>known</u>

$$X|\theta \sim \text{Gamma}(r,\theta) = \text{Erlang}(r,\theta) \rightarrow \text{sum of } r \text{ iid Exp}(\theta)\text{'s}$$

the cont! analogue of the NegBin(r,θ)

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

$$p(\theta|x) \propto p(x|\theta)\,p(\theta) = \left(\frac{\theta^r}{\underbrace{\Gamma(r)}_{=(r-1)!}} e^{-\theta x} x^{r-1}\right)\left(\frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta\theta}\theta^{\alpha-1}\right)$$

$$\propto \theta^r e^{-\theta x} e^{-\beta\theta}\theta^{\alpha-1}$$

$$= e^{-(x+\beta)\theta}\theta^{r+\alpha-1} \propto \text{Gamma}(r+\alpha, x+\beta)$$

$\Rightarrow$ Gamma is conjugate model for Gamma with fixed first parameter model

Normal Model

$$p(X \mid \theta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(X-\theta)^2}$$

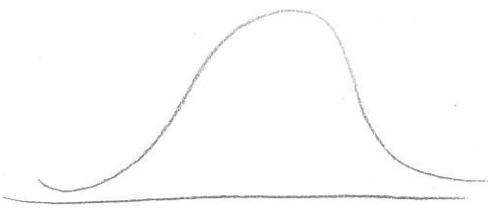
$$\propto e^{-\frac{1}{2\sigma^2}(x^2 - 2\theta x + \theta^2)}$$

$$= e^{-\frac{x^2}{2\sigma^2}} e^{\frac{\theta x}{2\sigma^2}} e^{-\frac{\theta^2}{2\sigma^2}}$$

$$\propto e^{-\frac{x^2}{2\sigma^2}} e^{\frac{\theta x}{2\sigma^2}}$$

$$= e^{-ax^2} e^{bx} \quad \text{s.t. } a \geq 0,\ b \in \mathbb{R}$$

do this later



$E(X) = \theta$

$Var(X) = \sigma^2$

$Supp(X) = \mathbb{R}$

Param space is two dimensional.

$\theta \in \mathbb{R}$

$\sigma^2 \in (0, \infty)$

given $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ and

Goal: pretend σ^2 is known, infer θ. First find MLE

$$\mathcal{L}(\theta; x, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i - \theta)^2}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum(x_i - \theta)^2}$$

$(x_i - \theta)^2 = x_i^2 - 2\theta x + \theta^2$

$\Rightarrow \sum(x_i - \theta)^2 = \sum x_i^2 - 2\theta \sum x + \sum \theta^2$

$= \sum x_i^2 - 2\theta \bar{x} n + n\theta^2$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}(\sum x_i^2 - 2\theta \bar{x} n + n\theta^2)}$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{\sum x_i^2}{2\sigma^2}} e^{\frac{\theta \bar{x} n}{\sigma^2}} e^{-\frac{n\theta^2}{2\sigma^2}}$$

$$\ell(\theta; x, \sigma^2) = n \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{\sum x_i^2}{2\sigma^2} + \frac{\theta \bar{x} n}{\sigma^2} - \frac{n\theta^2}{2\sigma^2}$$

$$\ell'(\theta; x, \sigma^2) = \frac{\bar{x} n}{\sigma^2} - \frac{\cancel{2} n\theta}{\cancel{2}\sigma^2} \overset{set}{=} 0$$

$\Rightarrow \bar{x} n - n\theta = 0$

$\Rightarrow \bar{x} - \theta = 0$

$\Rightarrow \boxed{\hat{\theta}_{MLE} = \bar{X}}$

Fun with Kernels... Just one..

$$P(X|\theta,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x_i^2}{2\sigma^2}} e^{\frac{\theta x_4}{\sigma^2}} e^{-\frac{\theta^2}{2\sigma^2}}$$

$$\propto \underbrace{e^{-\frac{x_i^2}{2\sigma^2}} e^{\frac{\theta x_4}{\sigma^2}}}_{e^{-a x_i^2} e^{bx}}$$

s.t. $a \geq 0, b \in \mathbb{R}$

$$P(X_{\text{whole}}|\theta,\sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^4 e^{-\frac{\sum x_i^2}{2\sigma^2}} e^{\frac{\theta \bar{x} 4}{\sigma^2}} e^{-\frac{4\theta^2}{2\sigma^2}}$$

$$\propto e^{-\frac{\sum x_i^2}{2\sigma^2}} e^{\frac{\theta \bar{x} 4}{\sigma^2}}$$

$$= e^{-a\sum x_i^2} e^{b\bar{x}} \quad \text{s.t. } a \geq 0, b \in \mathbb{R}$$

Now... let's figure out what the prior could be... back to Bayes Rule

$$P(\theta|X,\sigma^2) = \frac{P(X|\theta,\sigma^2)\,P(\theta|\sigma^2)}{P(X|\sigma^2)}$$

Why? σ^2 known... so it's conditioned on everywhere...

$$\propto P(X|\theta,\sigma^2)\,P(\theta|\sigma^2)$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^4 e^{-\frac{\sum x_i^2}{2\sigma^2}} e^{\frac{\theta \bar{x} 4}{\sigma^2}} e^{-\frac{4\theta^2}{2\sigma^2}} P(\theta|\sigma^2)$$

$$\propto \underbrace{e^{\frac{\theta \bar{x} 4}{\sigma^2}} e^{-\frac{4\theta^2}{2\sigma^2}}}_{e^{-a\theta^2} e^{b\theta}} P(\theta|\sigma^2)$$

s.t. $a = \frac{4}{2\sigma^2}$, $b = \frac{\bar{x} 4}{\sigma^2}$

What should $P(\theta|\sigma^2)$ be?

Let's match the kernel like we've been doing...

The kernel is a normal so let's do a normal...

$$P(\theta|\sigma^2) = N(\mu_0,\tau^2) = \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{1}{2\tau^2}(\theta-\mu_0)^2}$$

$$\propto e^{-\frac{1}{2\tau^2}(\theta^2 - 2\theta\mu_0 + \mu_0^2)}$$

$$\propto e^{-\frac{\theta^2}{2\tau^2}} e^{\frac{\theta\mu_0}{\tau^2}}$$

$$\Rightarrow p(\theta|X,\sigma^2) \propto e^{-\frac{n}{2\sigma^2}\theta^2} e^{\frac{\bar{X}n}{\sigma^2}\theta} \; e^{-\frac{1}{2\tau^2}\theta^2} e^{\frac{\mu_0}{\tau^2}\theta}$$

$$= e^{-\left(\frac{n}{2\sigma^2} + \frac{1}{2\tau^2}\right)\theta^2} e^{\left(\frac{\bar{X}n}{\sigma^2} + \frac{\mu_0}{\tau^2}\right)\theta}$$

This is normal dsn since the kernel is normal

but which one is it?

$$\propto \frac{1}{\sqrt{2\pi v^2}} e^{-\frac{1}{2v^2}(\theta - c)^2}$$

$$\propto e^{-\frac{1}{2v^2}(\theta-c)^2} = e^{-\frac{1}{2v^2}(\theta^2 - 2c\theta + c^2)} = e^{-\frac{1}{2v^2}\theta^2} e^{\frac{c}{v^2}\theta} e^{-\frac{c^2}{2v^2}}$$

$$\propto e^{-\frac{1}{2v^2}\theta^2} e^{\frac{c}{v^2}\theta}$$

solve for v^2, c

$$\Rightarrow \frac{1}{\cancel{2}v^2} = \frac{n}{\cancel{2}\sigma^2} + \frac{1}{\cancel{2}\tau^2} \Rightarrow v^2 = \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}$$

$$\Rightarrow \frac{c}{v^2} = \frac{\bar{X}n}{\sigma^2} + \frac{\mu_0}{\tau^2}$$

$$\Rightarrow c = \left(\frac{\bar{X}n}{\sigma^2} + \frac{\mu_0}{\tau^2}\right)v^2 = \frac{\frac{\bar{X}n}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}$$

$$\propto N\left(\frac{\frac{\bar{X}n}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\right)$$

Normal is conjugate prior for normal likelihood family.

$$\Rightarrow \quad X_1,\dots,X_n \mid \theta,\sigma^2 \overset{iid}{\sim} N(\theta,\sigma^2)$$

$$\theta \mid \sigma^2 \sim N(\mu_0,\tau^2)$$

$$\Rightarrow \theta \mid X_1,\dots,X_n,\sigma^2 \sim N\left(\frac{\frac{\bar{x}_n}{\sigma^2}+\frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}},\ \frac{1}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}\right)$$

$$\hat{\theta}_{MMSE} = \frac{\frac{\bar{x}_n}{\sigma^2}+\frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}$$

$\hat{\theta}_{MAE} =$ "

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

Why? Normal is symmetric & unimodal

$\Rightarrow$ mean = median = mode

Shrinkage...

$$\hat{\theta}_{MMSE} = \frac{\frac{\bar{x}_n}{\sigma^2}}{\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)}\cdot\frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n}} + \frac{\frac{\mu_0}{\tau^2}}{\left(\frac{n}{\sigma^2}+\frac{1}{\tau^2}\right)}\frac{\tau^2}{\tau^2}$$

$$= \frac{1}{1+\frac{\sigma^2}{n\tau^2}}\bar{x} + \frac{1}{\frac{n\tau^2}{\sigma^2}+1}\mu_0$$

$$= \underbrace{\frac{\sigma^2}{n\tau^2+\sigma^2}}_{\rho}\underbrace{\mu_0}_{E(\theta)} + \underbrace{\frac{n\tau^2}{n\tau^2+\sigma^2}}_{(1-\rho)}\underbrace{\bar{x}}_{\hat{\theta}_{MLE}}$$

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

weighted arithmetic avg shrinkage