Pois lik. [Lecture 11 3/16/16 Math 390.03-02]

$X|\theta \sim \text{Poisson}(\theta) := \frac{e^{-\theta}\theta^x}{x!}$, $\theta \in (0, \infty)$ param space

cont. gen of bin.

$X|\theta \sim \text{Exp}(\theta) := \theta e^{-\theta x}$ cont. gen. of geom

Poisson is limit of $\text{Bin}(n,p)$ s.t. $\theta = np$ $n \approx \infty, p \approx 0$

$\text{Supp}[X] = \{0, 1, \dots\} = \mathbb{N}_0$

$P(X|\theta) \propto \frac{\theta^x}{x!}$ why?

Given θ. Image $\theta = 3.12$

$P(X|\theta = 3.12) = \frac{e^{-3.12}\, 3.12^x}{x!} \propto \frac{3.12^x}{x!}$

a const

why not a const? Changes with x... it is a $f(x)$

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

Now given X

Image $X = 17$

$P(\theta|X=17) = \frac{e^{-\theta}\,\theta^{17}}{17!} \propto e^{-\theta}\theta^{17}$

not a const it is a $f(\theta)$

a const

different kernels

$\text{Supp}(\theta) = (0, \infty)$

$\int_0^\infty \theta^a e^{-b\theta}\, d\theta$

let $t = b\theta$

$\frac{dt}{d\theta} = b$

$\theta = \frac{t}{b}$

$\int_0^\infty \left(\frac{t}{b}\right)^{(a-1)-1} e^{-t} \frac{dt}{b}$

$\| $

$\frac{\Gamma(a+1)}{b^{a+1}}$

kernel

Bayesian setup. We want to infer θ. We have a prior idea on θ

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

seems to be $\theta^a e^{-b\theta}$

~~$\theta \sim P(\theta)$? Needs to have $supp(\theta) = ParamSpace(X) = (0, \infty)$~~

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

Shape hyperparam — Rate hyperparam

It's integral $\int_0^\infty f(\theta)\,d\theta = 1$ due to the gamma function!!

Param space
$\alpha, \beta \in (0, \infty)$
just like beta

$Med(\theta)$ no closed form

$Mode(\theta) = \frac{\alpha-1}{\beta}$ $\alpha \geq 1$

Kind of like the beta except for supp $(0, \infty)$

$E(\theta) = \frac{\alpha}{\beta}$
$Var(\theta) = \frac{\alpha}{\beta^2}$ } Calculus exercises

$P(\theta; \alpha, \beta) \propto \underbrace{\theta^{\alpha-1} e^{-\beta\theta}}_{\text{Kernel of the gamma PDF}}$ the $\frac{\beta^\alpha}{\Gamma(\alpha)}$ is a const. of integration

Why can't I use $k(X|\theta)$ since θ is non ignorant!

$$P(\theta|x) \propto \underset{\substack{\uparrow \\ \text{no } \frac{1}{P(x)}}}{P(X|\theta)P(\theta)} \propto P(X|\theta)\,k(\theta) = \left(\frac{e^{-\theta}\theta^x}{x!}\right)\left(\theta^{\alpha-1}e^{-\beta\theta}\right)$$

$$= \left(\frac{1}{x!}\right) \theta^{x+\alpha-1} e^{-\beta\theta-\theta}$$

const!!

$$\propto \theta^{x+\alpha-1} e^{-(\beta+1)\theta} \Big\}$$ Kernel of gamma

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

Gamma is <u>conjugate</u> prior for Poisson likelihood

But this is only for one data pt.

$X_1, \dots, X_n \overset{each}{\sim} \text{Poisson}(\theta)$, $\theta \sim \text{Gamma}(\alpha, \beta)$

$$P(\theta|x) \propto P(x|\theta)P(\theta) \propto P(x|\theta)\,k(\theta)$$

$$= \left(\prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!}\right)\left(\theta^{\alpha-1}e^{-\beta\theta}\right)$$

$$= \frac{e^{-n\theta}\,\theta^{\Sigma x_i}}{\prod_{i=1}^{n} x_i!}\,\theta^{\alpha-1}e^{-\beta\theta}$$

$$\propto e^{-n\theta}\theta^{\Sigma x_i}\theta^{\alpha-1}e^{-\beta\theta}$$

$$= \theta^{\Sigma x_i + \alpha - 1}e^{-(\beta+n)\theta}$$

$$\propto \underbrace{\text{Gamma}\left(\Sigma x_i + \alpha,\ n+\beta\right)}_{\text{Posterior for iid Poisson } \theta} \rightarrow \text{Always proper ??}$$

$$\hat{\theta}_{MMSE} := E[\theta|x] = \frac{\alpha'}{\beta'} = \frac{\Sigma x_i + \alpha}{n+\beta}$$

$\hat{\theta}_{MAE} := \text{Med}[\theta|x] = \text{qgamma}(0.5, \Sigma x_i + \alpha, n+\beta)$ need computer

$\hat{\theta}_{MAP} := \dfrac{\Sigma x_i + \alpha - 1}{n+\beta}$ as long as $\Sigma x_i + \alpha - 1 \geq 1$ (Usually not a prob.)

Principle of Indifference / Obj. / Ref / Uninformative prior

$P(\theta) = 1 \, \mathbb{1}_{\theta \geq 0}$

Indifferent to where θ is on the positive reals.

What's the problem? Improper $\int_0^\infty 1 \, d\theta = \infty$ not finite!

not a distr!

But ...

$$P(\theta|X) \propto P(X|\theta) \underbrace{P(\theta)}_{=1}$$

$$= P(X|\theta)$$

$$= \frac{e^{-n\theta} \, \theta^{\Sigma x_i}}{\prod_{i=1}^{n} x_i!}$$

$$\propto e^{-n\theta} \, \theta^{\Sigma x_i (+1 - 1)}$$

$$\propto \text{Gamma}\left(\Sigma x_i + 1, \; n\right) = \text{Gamma}(\Sigma x_i + \underset{=1}{\alpha}, \; n + \underset{=0}{\beta})$$

Always Proper posterior!!

Uninformative prior

is Gamma$(1, 0)$ not legal since α, β need > 0

$$:= \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha - 1} e^{-\beta\theta} = \frac{0^1}{\Gamma(1)} \theta^{1-1} e^{-(0)\theta} = \frac{0}{1}(1)(1) = 0 \text{ illegal!}$$

What is MLE?

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

$$\ell(\theta; x) = -n\theta + \Sigma x_i \ln\theta - \sum_{i=1}^{n} \ln(x_i!)$$

$$\ell'(\theta; x) = \frac{d}{d\theta}\left[\ell(\theta; x)\right] = -n + \frac{\Sigma x_i}{\theta} \overset{\text{set}}{=} 0 \Rightarrow \frac{\Sigma x_i}{\theta} = n \Rightarrow \boxed{\hat\theta_{MLE} = \frac{\Sigma x_i}{n} = \bar{x}}$$

Under uniform prior...

$$\hat\theta_{MMSE} = \frac{\Sigma x_i + \alpha}{n+\beta} = \frac{\Sigma x_i + 1}{n} = \hat\theta_{MLE} + \frac{1}{n}$$

$$\hat\theta_{MAP} = \frac{\Sigma x_i + \alpha - 1}{n+\beta} = \frac{\Sigma x_i}{n} = \hat\theta_{MLE}$$

$\hat\theta_{MAE}$ need computer...

$\theta \sim \text{Gamma}(0,0)$ totally illegal but... $\theta | x \sim \text{Gamma}(\Sigma x_i, n)$ sometimes improper!! since Σx_i can $= 0$!

$$\hat\theta_{MMSE} = \hat\theta_{MLE}$$

$$\hat\theta_{MAP} = \frac{\Sigma x_i - 1}{n} = \hat\theta_{MLE} - \frac{1}{n}$$

examine...

$$\hat\theta_{MMSE} = \frac{\Sigma x_i + \alpha}{n+\beta} = \frac{\Sigma x_i}{n+\beta} + \frac{\alpha}{n+\beta} = \frac{\Sigma x_i}{n+\beta}\left(\frac{n}{n}\right) + \frac{\alpha}{n+\beta}\left(\frac{\beta}{\beta}\right)$$

Is it a shrinkage estimator?

$$= \frac{n}{n+\beta}\hat\theta_{MLE} + \frac{\beta}{n+\beta} E(\theta)$$

$$= \rho E(\theta) + (1-\rho)\hat\theta_{MLE} \text{ s.t. } \rho = \frac{\beta}{n+\beta}$$

YES...

Jeffreys Prior...

$$l'(\theta; x) = -n + \frac{\sum x_i}{\theta}$$

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

$$I(\theta) := E\left[-l''(\theta; x)\right] = E\left(-\frac{n\bar{x}}{\theta^2}\right) = -\frac{n}{\theta^2}E[\bar{x}] = -\frac{n}{\theta^2}\theta = -\frac{n}{\theta} \propto \frac{1}{\theta} = \theta^{-1}$$

What is this? $\quad j(\theta) \propto \sqrt{I(\theta)} = \theta^{-\frac{1}{2}}$

$$\text{Gamma}(\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}$$

$$\propto \theta^{\alpha-1} e^{-\beta\theta} \quad \text{if } \beta = 0, \ \alpha = 0.5$$

$$\propto \theta^{-1/2}$$

Still improper!

$$\hat{\theta}_{MMSE} = \frac{\sum x_i + \frac{1}{2}}{n + \beta}$$

$$\hat{\theta}_{MAP} = \frac{\sum x_i - \frac{1}{2}}{n + \beta}$$

Gamma(α, β) → general

Gamma$(1, 0)$ → uniform

Gamma$(0, 0)$ → total ignorance

Gamma$(\frac{1}{2}, 0)$ → Jeffreys

<u>Analogue in Beta-Bin model</u>

Beta(α, β)

Beta$(1,1) = U(0,1)$ Laplace

Beta$(0,0)$ Haldane

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

Recall in Beta-Bin model

$$p(x) = \int_{\Theta} p(x|\theta)\, p(\theta)\, d\theta$$ prior predictive distr

if $\theta \sim Beta(\alpha,\beta) \Rightarrow p(x) = BetaBin(n,\alpha,\beta)$

$$p(x^*|x) = \int_{\Theta} p(x|\theta)\, p(\theta|x)\, d\theta$$ posterior predictive distr.

if $\theta \sim Beta(\alpha,\beta) \Rightarrow p(x^*|x) = BetaBin(m,\alpha+x,\beta+n-x)$

why are these two distr's the same?? Conjugacy...

What is the prior & post. pred distr's in the Poisson model? Assume $n=1$

$$p(x) = \int_{\Theta} p(x|\theta)\, p(\theta)\, d\theta$$

$$= \int_0^{\infty} \frac{e^{-\cancel{n}\theta}\, \theta^{\cancel{\Sigma} x}}{x!} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}\, d\theta$$

$$= \frac{\beta^{\alpha}}{\Gamma(\alpha)} \frac{1}{x!} \int_0^{\infty} \theta^{\cancel{\Sigma x_i}\, x+\alpha-1} e^{-(\beta+\cancel{n}\,1)\theta}\, d\theta$$

Recall $\Gamma(t) = \int_0^{\infty} y^{t-1} e^{-y}\, dy$

let $y = (\beta+n)\theta \Rightarrow \theta = \frac{y}{\beta+n}$, $d\theta = \frac{1}{\beta+n}\, dy$

$$\int_0^{\infty} \left(\frac{y}{\beta+n}\right)^{\Sigma x_i+\alpha-1} e^{-y} \frac{1}{\beta+n}\, dy$$

$t = \Sigma x_i + \alpha$

$$\int_0^{\infty} \frac{y^{\Sigma x_i+\alpha-1}\, e^{-y}}{(\beta+n)^{\Sigma x_i+\alpha-1+1}}\, dy = \frac{1}{(\beta+n)^{\Sigma x_i+\alpha}} \int_0^{\infty} y^{t-1} e^{-y}\, dy$$

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

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

$$\int_{Supp(\theta)} f(\theta)\, d\theta = 1$$

$$= \frac{\beta^\alpha}{(\beta+1)^\alpha} \frac{\Gamma(x+\alpha)}{\Gamma(\alpha)} \frac{1}{x!} \frac{1}{(\beta+1)^x}$$

$$= \left(\frac{\beta}{\beta+1}\right)^\alpha \left(\frac{1}{\beta+1}\right)^x \frac{(x+\alpha-1)!}{(\alpha-1)!\, x!}$$ if $\alpha \in \mathbb{N}$ only!

let $p = \frac{\beta}{\beta+1} \in (0,1)$, $r = \alpha$

$$= \binom{x+r-1}{x} p^r (1-p)^x := \text{NegBin}(r,p)$$

If $\alpha \notin \mathbb{N}$

$$= \frac{\Gamma(x+r)}{\Gamma(\alpha)} p^r (1-p)^x$$ "extended negative binomial"

Posterior pred distr (for one data pt)

$$p(x^*|x) = \int_\Theta p(x^*|\theta)\, p(\theta|x)\, d\theta$$

this has to be the same as the prior pred. distr with param changed since

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

$\theta|X \sim Gamma(\Sigma x_i + \alpha, n+\beta)$

$\Rightarrow X^*|X \sim NegBin(r, p)$ where $r = \Sigma x_i + \alpha$

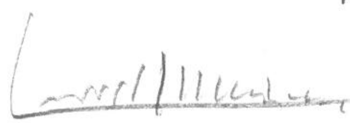
$p = \frac{n+\beta}{n+\beta+1}$

(HW exercise)

Math 241: NegBin: waiting time to achieve r successes from a Geom(p) process $= \sum_r Geom(p)$

But this's not what it is here!!

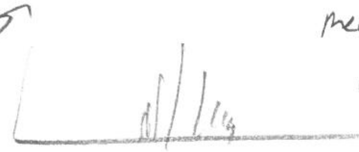
Recall: $X^*|X \sim BetaBin(n, \alpha, \beta)$



mean, expectation can be indep. with 2 params

but

$X^*|\theta \sim Bin(n, \theta)$

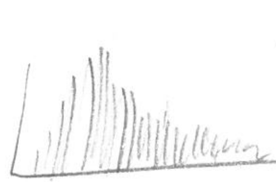


mean, expectation one param

Can be more dispersed due to uncertainty

NOW:

$X^*|X \sim NegBin\left(\Sigma x_i + \alpha, \frac{n+\beta}{n+\beta+1}\right)$



mean, expect can be independ 2 params

but

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

mean, expectation one param

Can be more dispersed as well

$X_1, \ldots, X_n \overset{\text{iid}}{\sim} Exp(\theta) := \theta e^{-\theta x}$

$$p(\theta|x) \propto p(X|\theta)\, p(\theta)$$

$$= \prod_{i=1}^{n} \theta e^{-\theta x_i} \; p(\theta)$$

$$= \underbrace{\theta^n e^{-(\sum x_i)\theta}}_{\text{Kernel for Gamma}} \; p(\theta)$$

Kernel for Gamma

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

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

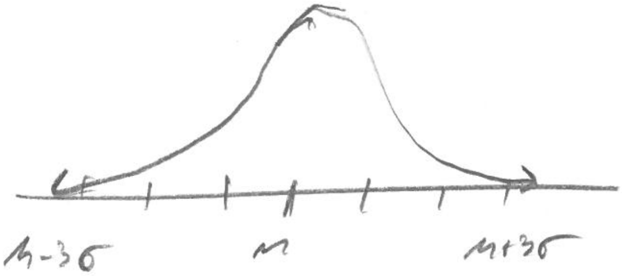
$$= \theta^{n+\alpha-1} e^{-(\sum x_i+\beta)\theta}$$

$$\propto Gamma(n+\alpha, \sum x_i + \beta)$$

everything else for HW....

Normal Model

$$X \sim N(\theta,\sigma^2) := \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2\sigma^2}(x-\theta)^2}$$



Features
symmetric
unimodal
"thin tails" e.g. $P(X > \theta+3\sigma) \approx 0.15\%$

$E(X)=\theta$, $Var(X)=\sigma^2$, $\Leftarrow$ Convenient and intuitive parameterization by default

Param space $\left.\begin{array}{l}\theta \in \mathbb{R}\\ \sigma^2 \in (0,\infty)\end{array}\right\} \vec{\theta}$ dim 2!! $Supp(X) = \mathbb{R}$

$$P(X|\theta,\sigma^2) \propto e^{-\frac{1}{2\sigma^2}(x-\theta)^2}$$ Random?

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

$$= e^{-\frac{1}{2\sigma^2}(x^2-2\theta x)}\, e^{-\frac{1}{2\sigma^2}\theta^2}$$

$$\propto \underbrace{e^{-\frac{1}{2\sigma^2}(x^2-2\theta x)}}_{\text{Kernel}}$$ Fun?

$$P(\theta|X,\sigma^2) \propto e^{-\frac{1}{2\sigma^2}(x-\theta)^2}$$

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

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

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

$$= \underbrace{e^{\left(\frac{x}{\sigma^2}\right)\theta}\, e^{\left(-\frac{1}{2\sigma^2}\right)\theta^2}}_{\text{Kernel}}$$