Shrinkage — a similar shrink one

$\hat{\theta}_S = \rho\theta_0 + (1-\rho)\hat{\theta}_{MLE}$

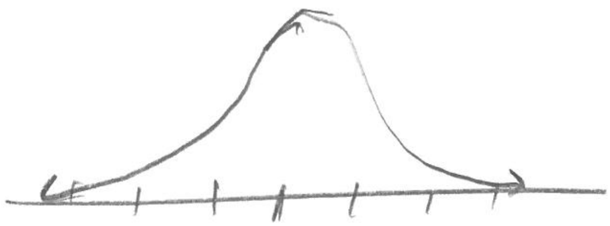
$= \rho\theta_0 + \hat{\theta}_{MLE} - \rho\hat{\theta}_{MLE} = \hat{\theta}_{MLE} + \rho(\theta_0 - \hat{\theta}_{MLE})$

$= \hat{\theta}_{MLE} - \rho(\hat{\theta}_{MLE} - \theta_0)$ If $\theta \Rightarrow$ $+ \downarrow$ $- \uparrow$

L1

Lecture 12 3/21/16

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



$\mu - 3\sigma$ μ $\mu + 3\sigma$

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

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

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?

Gives you back a normal!

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

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

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

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

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

Targets

$\theta \mid X, \sigma^2$ one

$\sigma^2 \mid X, \theta$ one

$\theta, \sigma^2 \mid X$ both

$\theta \mid X$ one but diff?!

~~$\sigma^2 \mid X$ won't do~~

Goal: pretend we know σ^2, infer μ. $X_1, \dots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$

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

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

$$\ell(\theta; x) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum \underbrace{(x_i-\theta)^2}_{f(g(\theta))} \qquad \frac{d}{d\theta} f(g(\theta)) = f'(g(\theta))\, g'(\theta)$$

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

$$= \frac{1}{\sigma^2}\left(\sum x_i - n\theta\right) \overset{\text{set}}{=} 0$$

$= \frac{n}{\sigma^2}(\bar{x} - \theta) = 0 \quad \Rightarrow \quad \bar{x} = \hat{\theta}_{MLE}$

Bayesian: now..

$$\underbrace{P(\theta \mid X, \sigma^2)}_{} = \frac{P(X \mid \theta, \sigma^2)\, P(\theta \mid \sigma^2)}{P(X \mid \sigma^2)} \propto P(X \mid \theta, \sigma^2)\, P(\theta \mid \sigma^2)$$

$X = x_1, \dots, x_n$

Looks a bit different due to this conditioning but it's the same thing...

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

$$\propto e^{-\frac{1}{2\sigma^2}\sum -2\theta x_i + \theta^2} \quad P(\theta \mid \sigma^2)$$

X, σ^2 given

$$= e^{-\frac{1}{2\sigma^2}(-2\theta \bar{x} n + n\theta^2)} \quad P(\theta \mid \sigma^2)$$

$$= e^{\frac{\bar{x}n}{\sigma^2}\theta}\, e^{-\frac{n}{2\sigma^2}\theta^2}\, P(\theta \mid \sigma^2)$$

What is conjugate prior??

Recall Poisson mode

$$p(\theta|x) \propto \theta^{\Sigma x_i} e^{-n\theta} \, p(\theta)$$

It seems if $P(\theta)$ is of the form $\theta^{a} e^{b\theta}$

$$\Rightarrow \theta^{\Sigma x_i} e^{-n\theta} \left(\theta^{a} e^{b\theta}\right)$$

$$= \theta^{\Sigma x_i + a} e^{(b-n)\theta}$$

is the same form

... then hope it's a real distr.

Here... if $P(\theta|\sigma^2) = e^{a\theta} e^{b\theta^2} = P(\theta)$ ... then... ← not a function of σ^2

$$P(\theta|x,\sigma^2) \propto e^{\left(\frac{\bar{x} n}{\sigma^2} + a\right)\theta} e^{\left(b - \frac{n}{2\sigma^2}\right)\theta^2}$$

Kernel of Normal — Kernel of normal as well

but we don't know which normal yet!

$\Rightarrow$ the normal is the conjugate prior for the normal likelihood

"Self-conjugate"

Okay let $\theta \sim N(\mu_0, \tau^2)$

No dependence on σ^2 as an assumption... otherwise MUCH harder...

$$p(\theta|x,\sigma^2) \propto e^{\frac{\bar{x}n}{\sigma^2}\theta} e^{-\frac{n}{2\sigma^2}\theta^2} \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{1}{2\tau^2}(\theta-\mu_0)^2}$$

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

$$\propto e^{\frac{\bar{x}n}{\sigma^2}\theta} e^{-\frac{n}{2\sigma^2}\theta^2} e^{-\frac{1}{2\tau^2}\theta^2} e^{\frac{\mu_0}{\tau^2}\theta}$$

$$= e^{\underbrace{\left(\frac{\bar{x}n}{\sigma^2}+\frac{\mu_0}{\tau^2}\right)}_{a}\theta} e^{\underbrace{\left(-\frac{n}{2\sigma^2}-\frac{1}{2\tau^2}\right)}_{b}\theta^2}$$

We know this is a normal but which one??

Ness

$$= e^{b\theta^2 + a\theta}$$

$$= e^{b\left(\theta^2 + \frac{a}{b}\theta\right)} \quad \leftarrow \text{form}$$

Remember $\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\theta)^2}$ we need this to look like $e^{-\frac{1}{2v}(\theta-d)^2} \propto e^{-\frac{1}{2v}(\theta^2 - 2d\theta)}$

$$(\theta-d)^2 = \theta^2 - 2d\theta + d^2$$

this process is called "completing the square" and is sometimes taught in precalculus classes

$$\Rightarrow -2d = \frac{a}{b} \Rightarrow d = -\frac{a}{2b} = -\frac{\frac{\bar{x}n}{\sigma^2}+\frac{\mu_0}{\tau^2}}{2\left(-\frac{n}{2\sigma^2}-\frac{1}{2\tau^2}\right)}$$

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

and $-\frac{1}{2v} = b = -\frac{n}{2\sigma^2} - \frac{1}{2\tau^2}$

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

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

okay, let's put all this together...

Same because its $= -2b$

$$p(\theta|x,\sigma^2) = N\left(\frac{\frac{\bar{x}n}{\sigma^2}+\frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}, \frac{1}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}\right)$$

$$\underbrace{\hat{\theta}_{MMSE}}_{\text{posterior expectation}} = \underbrace{\hat{\theta}_{MAE}}_{\text{posterior median}} = \underbrace{\hat{\theta}_{MAP}}_{\text{posterior mode}} = \frac{\frac{\bar{X}n}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}$$

Why? Symmetric & unimodal $\Rightarrow$ mean = median = mode

the beta with $\alpha = \beta$ is just like this too... & $\alpha, \beta > 1$

Rewrite:

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

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

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

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

$$= \rho E(\theta) + (1-\rho)\hat{\theta}_{MLE}$$ a.k.a. shrinkage estimator

$$\lim_{n\to\infty} \rho = \lim_{n\to\infty} \frac{\sigma^2}{\tau^2 n+\sigma^2} = 0 \text{ (as usual)}$$

and as expected

Interpretation of hyperparameters

let $\tau^2 = \frac{\sigma^2}{m}$ ← known, ← subjectively assumed

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

let $\mu_0 = \bar{Y}$ avg of m measurements

$$\Rightarrow \hat{\theta}_{MMSE} = \frac{\sum x_i + \sum y_i}{n + m}$$

equivalent to seeing m measurements avg of m w/ var σ^2/m

m small $\Rightarrow$ τ^2 big, uninformative

Uniformative prior

$p(\theta) \propto 1$ improper? of course...

$$p(\theta | X, \sigma^2) \propto p(X | \theta, \sigma^2) \overbrace{p(\theta | \sigma^2)}^{1}$$

$$\propto e^{\underbrace{\frac{\bar{X} n}{\sigma^2}}_{a} \theta} e^{\underbrace{-\frac{n}{2\sigma^2}}_{b} \theta^2}$$

$$d = -\frac{a}{2b} = \frac{-\frac{\bar{X} n}{\sigma^2}}{2\left(-\frac{n}{2\sigma^2}\right)} = \bar{X}$$

$$-\frac{1}{2v} = b = -\frac{n}{2\sigma^2} \Rightarrow v = \frac{\sigma^2}{n}$$

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

$\hat{\theta}_{MMSE} = \hat{\theta}_{MAE} = \hat{\theta}_{MAP} = \hat{\theta}_{MLE}$!

and Never improper posterior! WIN!!

Jeffreys Prior...

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

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

$$I(\theta) = E[-l''(\theta; x)] = \frac{n}{\sigma^2}$$

$j(\theta) \propto \sqrt{I(\theta)} \propto 1 \Rightarrow$ the Jeffreys prior is the uniformative prior in the normal case w/ σ^2 known

What are improper priors?

$X \sim Bin(n, \theta)$

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

$\Rightarrow \theta | x \sim Beta(\alpha + x, \beta + n - x)$

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

$\Rightarrow \theta | x \sim Beta(x, n - x)$

Which can be thought of as...

$$\lim_{\alpha \to 0} \lim_{\beta \to 0} p(\theta | x) = Beta(x, n - x)$$

They are the limits of proper priors...

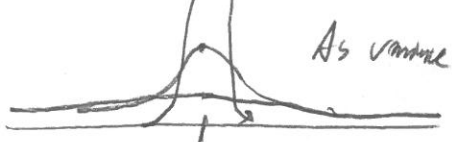
What is $p(\theta) \propto 1$ in the normal case?

(Convenient notation)

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

$$\Rightarrow \theta | x, \sigma^2 \sim N\left(\underbrace{\frac{\frac{\bar{x} n}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}}_{\theta_p}, \underbrace{\frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}}_{\sigma_p^2}\right) \longrightarrow N\left(\bar{x}, \frac{\sigma^2}{n}\right)$$

If $\tau^2 \to \infty$ (or $n \to \infty$)



μ_0

As variance goes up and up... more and more flat...

Post. pred. distr

$$N(\theta_p, \sigma_p^2)$$

$$p(x^*|x,\sigma^2) = \int_{\Theta} p(x^*|\theta,\sigma^2)\, p(\theta|x,\sigma^2)\, d\theta$$

$$= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x^*-\theta)^2} \frac{1}{\sqrt{2\pi\sigma_p^2}} e^{-\frac{1}{2\sigma_p^2}(\theta-\theta_p)^2}\, d\theta$$

Math 241+

· let $X_1 \sim U(1,2,3,4,5,6)$, $X_2 \sim U(1,2,3,4,5,6)$, IND. $\Rightarrow p(x) = \frac{1}{6}$

$S = X_1 + X_2 \sim ?$

$P(S=1) = 0$

$P(S=2) = P(X_1=1)P(X_2=1) = \frac{1}{6}\cdot\frac{1}{6}$

$P(S=3) = P(X_1=1)P(X_2=2) + P(X_1=2)P(X_1=1)$

$$\sum_{x\in Supp[X_1]} P(X_1=x)\,P(X_2=3-x)$$

$\rightarrow$ if $x=3$ is zero ... so you must be careful of the bounds

only $x=1, x=2$ are valid for both X_1, X_2

$$= \sum_{x\in\{1,2\}} P(X_1=x)\,P(X_2=3-x)$$

$$P(S=s) = \sum_{x \to \text{careful}} P(X_1=x)\,P(X_2=s-x)$$

If $X_1 \sim f_X(x)$, $X_2 \sim g_X(x)$, ind, where $Supp[X_1] = Supp[X_2] = \mathbb{R}$

$T = X_1 + X_2 \sim f * g := \int_{\mathbb{R}} f_X(x)\, f_Y(t-x)\, dx = f_T(t)$ "Convolution operator"

Let $t = X^*$, $x = \theta$

$$P(X^*|X,\sigma^2) = \int_{\mathbb{R}} \underbrace{\frac{1}{\sqrt{2\pi\sigma_p^2}} e^{-\frac{1}{2\sigma_p^2}(\theta-\theta_p)^2}}_{f_{X_1}(\theta)} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}((X^*-\theta)-0)^2}}_{f_{X_2}(X^*-\theta)} d\theta$$

$f_{X_1}(\theta) \Downarrow X_1 \sim N(\theta_p, \sigma_p^2)$

$f_{X_2}(X^*-\theta) \Downarrow X_2 \sim N(0,\sigma^2)$

$\Rightarrow X_1 + X_2 \sim N(\theta_p, \sigma_p^2 + \sigma^2)$

Assume the Thm: sum of normals is normal (Math 241)

$$\Rightarrow X^*|X,\sigma^2 \sim N(\theta_p, \sigma_p^2 + \sigma^2) = N\left(\frac{\frac{\bar{X}n}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}, \frac{1}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}} + \sigma^2\right)$$

Or... do the integral yourself (not so bad)

Your predictions look a lot like $\hat{\theta}_{MMSE}(=\dots)$ with more variance... makes sense? Inflated/over-dispersed normal

just like binomial $\uparrow$ or betaNB

$\sigma^2|X,\theta$?

Same lik.

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

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(X_i-\theta)^2} P(\sigma^2|\theta)$$

$$= \sqrt{\frac{1}{(2\pi)^n}} \sqrt{\frac{1}{(\sigma^2)^n}} e^{-\frac{1}{2\sigma^2}\sum(X_i-\theta)^2} P(\sigma^2|\theta)$$

$$\propto (\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum(X_i-\theta)^2} P(\sigma^2|\theta)$$

let $\underbrace{\hat{\sigma}^2 := \frac{1}{n}\sum(X_i-\theta)^2}_{\text{convenient notation}}$

$$= (\sigma^2)^{-\frac{n}{2}} e^{-\frac{n\hat{\sigma}^2}{2\sigma^2}} \left((\sigma^2)^{a} e^{-\frac{b}{\sigma^2}}\right) ??$$