

$$E[\theta] = \frac{a+\alpha}{2(\alpha+\beta)} = \frac{\frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}$$

$\mathcal{F}: X_1, \ldots, X_n \mid \theta, \sigma^2 \overset{iid}{\sim} N(\theta, \sigma^2)$

$\Rightarrow P(\theta) = N(\mu_0, \tau^2)$ is the conjugate prior

$$\Rightarrow P(\theta \mid X, \sigma^2) = N\left(\frac{\frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}},\ \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\right)$$

this was proven just with 10th grade algebra and kernels!

probably one of the most famous formulas in Bayesian statistics!

Pt. Estimates

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

normal distr. is symmetric & unimodal! mean, median, mode are all the same!!

Credible regions..

$$CR_{\theta, 1-\alpha_0} := \left[qnorm\left(\frac{\alpha_0}{2}, \frac{\frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\right),\ qnorm\left(1 - \frac{\alpha_0}{2}, \frac{\frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\right) \right]$$

Hypothesis tests at level α_0, σ^2 known

$H_0: \theta \geq \theta_0$

$H_1: \theta < \theta_0$

$$pval := P(H_0 \mid X, \sigma^2) = P(\theta \geq \theta_0 \mid X, \sigma^2) = 1 - pnorm(\theta_0, \ , \ )$$

Shrinkage? Let's compute $\hat{\theta}_{MLE}$ first. You did this on the HW..

$$\mathcal{L}(\theta; X, \sigma^2) = (2\pi\sigma^2)^{-n/2} e^{-\frac{\sum x_i^2}{2\sigma^2} + \frac{n\bar{x}\theta}{\sigma^2} - \frac{n\theta^2}{2\sigma^2}}, \quad \ell'(\theta; X, \sigma^2) = \frac{n\bar{x}}{\sigma^2} - \frac{n\theta}{\sigma^2} \overset{set}{=} 0$$

$$\ell(\theta; X, \sigma^2) = \ln(\ ) - \frac{\sum x_i^2}{2\sigma^2} + \frac{n\bar{x}\theta}{\sigma^2} - \frac{n\theta^2}{2\sigma^2}$$

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

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

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

$$= \left(\frac{\tau^2\sigma^2}{\tau^2\sigma^2}\right)\underbrace{\frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}}_{1-\rho}\hat{\theta}_{MLE} + \left(\frac{\tau^2\sigma^2}{\tau^2\sigma^2}\right)\underbrace{\frac{\frac{1}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}}_{\rho} E[\theta|\sigma^2]$$

$$= \underbrace{\frac{n\tau^2}{n\tau^2+\sigma^2}}_{1-\rho}\hat{\theta}_{MLE} + \underbrace{\frac{\sigma^2}{n\tau^2+\sigma^2}}_{\rho} E[\theta|\sigma^2]$$

Yes... it is a shrinkage estimator

If $n \to \infty \Rightarrow \rho \to 0$

Laplace Prior $P(\theta|\sigma^2) \propto 1$

we did this before!!

$P(\theta|x,\sigma^2) \propto P(x|\theta,\sigma^2)(1) \propto N\left(\bar{x}, \frac{\sigma^2}{n}\right)$ what is μ_0, τ^2?

$\bar{x} = \frac{\frac{n\bar{x}}{\sigma^2}+\frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}$, $\frac{\sigma^2}{n} = \frac{1}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}$ $\tau^2 = \infty$, $\mu_0 =$ anything! $= 0$

Laplace prior: $N(\theta|\sigma^2) = N(0,\infty)$ def. improper!!!

Jeffreys Prior: $P_J(\theta) \propto \sqrt{I(\theta)} = \sqrt{\frac{n}{\sigma^2}} \propto 1 \Rightarrow P_J(\theta)$ is same as Laplace Prior is $N(0,\infty)$

$I(\theta) :=$

$-\ell''(\theta;x) = \frac{n}{\sigma^2}$, $\underset{x}{E}[-\ell''(\theta;x)] = \underset{x}{E}\left[\frac{n}{\sigma^2}\right] = \frac{n}{\sigma^2}$

Pseudocount Interpretation of μ_0, τ^2.

Imagine previous data $Y_1, \ldots, Y_{n_0}$

let $\mu_0 = \bar{y}$

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

if $\frac{1}{\tau^2} = \frac{n_0}{\sigma^2}$ then $\hat{\theta}_{MMSE} = \dfrac{\frac{n\bar{x}}{\sigma^2} + \frac{n_0\bar{Y}}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{n_0}{\sigma^2}} = \dfrac{n\bar{x} + n_0\bar{Y}}{n+n_0} = \frac{n}{n+n_0}\bar{x} + \frac{n_0}{n+n_0}\bar{Y}$

Shrinkage expression simple!

Convex comb. of data avg. and pseudodata avg.

$= \dfrac{\Sigma x_i + \Sigma y_i}{n+n_0}$ avg of all data and pseudodata

$\Rightarrow \tau^2 = \dfrac{\sigma^2}{n_0}$

$\Rightarrow P(\theta|\sigma^2) = N\left(\bar{y}, \dfrac{\sigma^2}{n_0}\right)$ need σ^2 known for this to work!!!

now just specify $\bar{Y}$ and strength n_0

$\Rightarrow P(\theta|x,\sigma^2) = N\left(\frac{n}{n+n_0}\bar{x} + \frac{n_0}{n+n_0}\bar{Y}, \dfrac{\sigma^2}{n+n_0}\right)$ much nicer!

Haldane prior: complete ignorance let $n_0 = 0$

$\Rightarrow P(\theta|\sigma^2) = N(\bar{Y}, \infty) = N(0, \infty)$ $\bar{Y}$ doesn't matter so $= 0$

Laplace = Jeffreys = Haldane = $N(0,\infty)$! So this is the only principled uninformative prior to use!

σ^2 unknown

$P(\theta) = N(\mu_0, \tau^2) \Rightarrow$ still okay

$= N\left(\bar{Y}, \frac{\tau_0^2}{n_0}\right)$ 3 hyperparameters can be shortened... Haldane same...

Posterior Predictive Distr. for one X^* and fixed σ^2.

let $\theta_p := \dfrac{\frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}$, $\sigma_p^2 = \dfrac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}$

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

$$\propto \int_{\mathbb{R}} e^{-\frac{1}{2\sigma^2}(x^*-\theta)^2} e^{-\frac{1}{2\sigma_p^2}(\theta-\theta_p^2)} d\theta = \int_{\mathbb{R}} \overset{\text{constant wrt } \theta}{e^{-\frac{x^{*2}}{2\sigma^2}}} e^{\frac{x^*\theta}{\sigma^2}} e^{-\frac{\theta^2}{2\sigma^2}} e^{-\frac{\theta^2}{2\sigma_p^2}} e^{\frac{\theta\theta_p}{\sigma_p^2}} \overset{\text{constant wrt } x^*, \theta}{e^{-\frac{\theta_p^2}{2\sigma_p^2}}} d\theta$$

$$\propto e^{-\frac{x^{*2}}{2\sigma^2}} \int_{\mathbb{R}} \underbrace{e^{\left(\frac{x^*}{\sigma^2} + \frac{\theta_p}{\sigma_p^2}\right)\theta - \left(\frac{1}{2\sigma^2} + \frac{1}{2\sigma_p^2}\right)\theta^2}}_{e^{a\theta - b\theta^2} \;\propto\; N\left(\frac{a}{2b}, \frac{1}{2b}\right)} d\theta$$

To avoid doing this integral... ~~lets~~ do some kernel stuff again

$$N\left(\tfrac{a}{2b}, \tfrac{1}{2b}\right) = \frac{1}{\sqrt{2\pi\frac{1}{2b}}} e^{-\frac{1}{2\frac{1}{2b}}\left(\theta - \frac{a}{2b}\right)^2} = \frac{1}{\sqrt{\frac{\pi}{b}}} e^{-b\left(\theta^2 - 2\theta\frac{a}{2b} + \frac{a^2}{4b^2}\right)}$$

$$= \frac{1}{\sqrt{\frac{\pi}{b}}} e^{-b\theta^2 + a\theta - \frac{a^2}{4b}} = \underbrace{\left(\frac{1}{\sqrt{\frac{\pi}{b}}} e^{-\frac{a^2}{4b}}\right)}_{\text{normalization constant}} e^{a\theta - b\theta^2}$$

$$\Rightarrow \int_{\mathbb{R}} e^{a\theta - b\theta^2} d\theta = \left(\frac{1}{\sqrt{\frac{\pi}{b}}} e^{-\frac{a^2}{4b}}\right)^{-1} \underbrace{\int_{\mathbb{R}} \left(\frac{1}{\sqrt{\frac{\pi}{b}}} e^{-\frac{a^2}{4b}}\right) e^{a\theta - b\theta^2} d\theta}_{1} = \sqrt{\frac{\pi}{b}}\, e^{\frac{a^2}{4b}}$$

$$\to = e^{-\frac{x^*}{2\sigma^2}} \sqrt{\frac{\pi}{\frac{1}{2\sigma^2} + \frac{1}{2\sigma_p^2}}}\; e^{\left(\frac{x^*}{\sigma^2} + \frac{\theta_p}{\sigma_p^2}\right)^2 \Big/ \left(\frac{1}{\sigma^2} + \frac{1}{\sigma_p^2}\right)}$$

$$= e^{-\frac{x^*}{2\sigma^2}} e^{\frac{\frac{(x^*)^2}{(\sigma^2)^2}}{A}} e^{\frac{2\frac{x^*}{\sigma^2}\frac{\theta_p}{\sigma_p^2}}{A}} e^{\frac{\frac{\theta_p^2}{(\sigma_p^2)^2}}{A}} \propto e^{-\frac{x^*}{2\sigma^2}} e^{\frac{\left(\frac{x^*}{\sigma^2} + \frac{\theta_p}{\sigma_p^2}\right)^2}{\frac{1}{\sigma^2} + \frac{1}{\sigma_p^2}}\Big/ A} \propto e^{\left(\frac{\frac{2}{\sigma^2}\frac{\theta_p}{\sigma_p^2}}{A} - \frac{1}{2\sigma^2}\right)x^*}$$