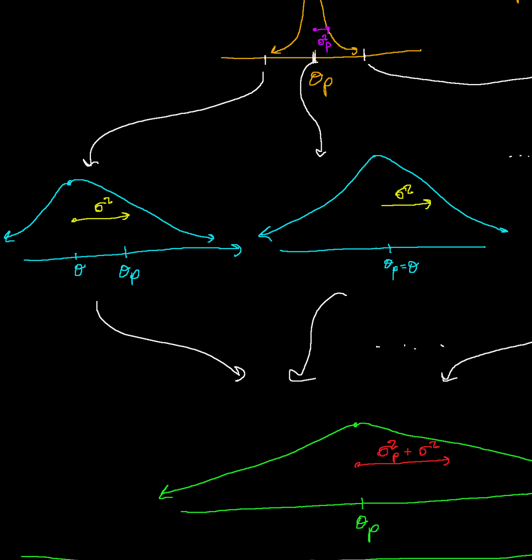
$P(\theta|X,\sigma^2) = N(\theta_p, \sigma_p^2)$



An overdispersed normal (with known variance is also normal.

Now consider the iid normal model with θ known, sigsq unknown i.e.

$Y_1 \overset{iid}{\sim} N(\theta, \sigma^2)$, θ known

$$P(X|\theta,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\theta)^2} = (2\pi)^{-n/2} (\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\overbrace{\sum(x_i-\theta)^2}^{\text{constant}}}$$

$$P(\sigma^2|X,\theta) \propto P(X|\theta,\sigma^2)\,P(\sigma^2|\theta) \propto P(X|\theta,\sigma^2) \propto (\sigma^2)^{-n/2} e^{-\frac{\sum(x_i-\theta)^2/2}{\sigma^2}}$$
$$= (\sigma^2)^{-n/2} e^{-\frac{n\hat{\sigma}^2_{MLE}/2}{\sigma^2}}$$

Consider the Laplace prior of indifference, a distribution on sigsq which has support (0, ∞). This prior would be $P(\sigma^2|\theta) \propto 1$

Let's take a break and find the MLE for sigsq.

$$\ell(\sigma^2; X,\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum(x_i-\theta)^2$$

$$\ell'(\sigma^2; X,\theta) = -\frac{n}{\cancel{2}\cancel{\sigma^2}} + \frac{\sum(x_i-\theta)^2}{\cancel{2}(\sigma^2)^2} \overset{set}{=} 0 \Rightarrow \frac{\sum(x_i-\theta)^2}{\sigma^2} = n \Rightarrow \hat{\sigma}^2_{MLE} = \frac{\sum(x_i-\theta)^2}{n}$$

average sqd deviation

Let's explore the kernel of the posterior using probability theory.

$K(y) = y^{-a} e^{-\frac{b}{y}}$ where $y \in (0,\infty)$

Let's try to find the actual density by finding the norm. constant c:

$$\frac{1}{c} = \int_0^\infty K(y)\,dy = \int_0^\infty y^{-a} e^{-\frac{b}{y}}\,dy = \int_\infty^0 z^{a} e^{-bz}(-z^{-2})\,dz = \int_0^\infty z^{(a-1)-1} e^{-bz}\,dz$$

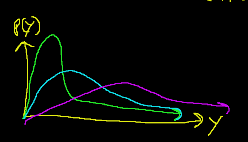
let $z = \frac{1}{y} \Rightarrow y = \frac{1}{z} \Rightarrow \frac{dy}{dz} = -z^{-2} \Rightarrow dy = -z^{-2}dz$, $y=0 \Rightarrow z=\infty$, $y=\infty \Rightarrow z=0$

Traditionally, $\alpha = a-1 \Rightarrow a = \alpha+1$, $\beta = b$

$$\overset{u\text{-subst}}{=} \frac{\Gamma(a-1)}{b^{a-1}} \Rightarrow P(y) = \frac{b^{a-1}}{\Gamma(a-1)} y^{-a} e^{-\frac{b}{y}} \Rightarrow P(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-\alpha-1} e^{-\frac{\beta}{y}}$$

This is called the "inverse gamma" distribution $= InvGamma(\alpha,\beta)$, $\alpha,\beta > 0$

Note: $W \sim Gamma(\alpha,\beta) \iff \frac{1}{W} \sim InvGamma(\alpha,\beta)$



$Y \sim InvGamma(\alpha,\beta)$

$E[Y] = \frac{\beta}{\alpha-1}$ for $\alpha > 1$

$Med[Y] = qinvgamma(0.5, \alpha, \beta)$

$Mode[Y] = \frac{\beta}{\alpha+1}$ for all $\alpha,\beta > 0$.

Back to the regularly scheduled program...

$$P(\sigma^2|X,\theta) \propto (\sigma^2)^{-n/2} e^{-\frac{n\hat{\sigma}^2_{MLE}/2}{\sigma^2}} = (\sigma^2)^{-\frac{n-2}{2}-1} e^{-\frac{n\hat{\sigma}^2_{MLE}/2}{\sigma^2}} \propto InvGamma\left(\frac{n-2}{2}, \frac{n\hat{\sigma}^2_{MLE}}{2}\right)$$

$$-\frac{n}{2} = -\frac{n}{2}+1-1 = -\left(\frac{n}{2}-1\right)-1 = -\frac{n-2}{2}-1$$

That's the posterior under Laplace's prior. Let's get the conjugate model now:

$$P(\sigma^2|X,\theta) \propto P(X|\theta,\sigma^2)\,P(\sigma^2|\theta) \propto (\sigma^2)^{-\frac{n}{2}} e^{-\frac{n\hat{\sigma}^2_{MLE}/2}{\sigma^2}} P(\sigma^2|\theta)$$

What form should the prior be so that its kernel has the same form as the posterior's kernel? It's an inverse gamma.

Let $P(\sigma^2|\theta) = InvGamma(\alpha,\beta)$

$$= (\sigma^2)^{-n/2} e^{-\frac{n\hat{\sigma}^2_{MLE}/2}{\sigma^2}} \frac{\beta^\alpha}{\Gamma(\alpha)} (\sigma^2)^{-\alpha-1} e^{-\frac{\beta}{\sigma^2}} \propto (\sigma^2)^{-\left(\frac{n}{2}+\alpha\right)-1} e^{-\frac{\frac{n\hat{\sigma}^2_{MLE}}{2}+\beta}{\sigma^2}}$$

$$\propto InvGamma\left(\frac{n}{2}+\alpha, \frac{n\hat{\sigma}^2_{MLE}}{2}+\beta\right)$$

Traditionally, we use a different parameterization of the prior:

let $\alpha = \frac{n_0}{2}$, $\beta = \frac{n_0\sigma_0^2}{2} \Rightarrow P(\sigma^2|\theta) = InvGamma\left(\frac{n_0}{2}, \frac{n_0\sigma_0^2}{2}\right)$

$$P(\sigma^2|X,\theta) = InvGamma\left(\frac{n+n_0}{2}, \frac{n\hat{\sigma}^2_{MLE}+n_0\sigma_0^2}{2}\right)$$

Bayesian point estimates for sigsq:

$$\hat{\theta}_{MMSE} = E[\sigma^2|X,\theta] = \frac{\frac{n\hat{\sigma}^2_{MLE}+n_0\sigma_0^2}{2}}{\frac{n+n_0}{2}-1} = \frac{n\hat{\sigma}^2_{MLE}+n_0\sigma_0^2}{n+n_0-2} \quad \text{if } n+n_0 > 2$$

$\alpha > 1 \Rightarrow \frac{n+n_0}{2} > 1$

$$\hat{\theta}_{MMAE} = Med[\sigma^2|X,\theta] = qinvgamma\left(0.5, \frac{n+n_0}{2}, \frac{n\hat{\sigma}^2_{MLE}+n_0\sigma_0^2}{2}\right)$$

$$\hat{\theta}_{MAP} = Mode[\sigma^2|X,\theta] = \ldots = \frac{n\hat{\sigma}^2_{MLE}+n_0\sigma_0^2}{n+n_0+2}$$

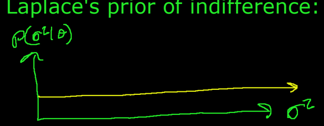
Credible Regions? Same thing... just use appropriate qinvgamma.
Hypothesis Tests? Same thing... just use appropriate pinvgamma.

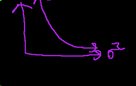
Pseudoobservation interpretation. n_0 = # of pseudo observations.
Imagine $y_1, y_2, \ldots, y_{n_0}$ sigsq_0 is guess of value of sigsq

$$\frac{n\hat{\sigma}^2_{MLE}+n_0\sigma_0^2}{2} = \frac{\sum_{i=1}^{n}(x_i-\theta)^2 + \sum_{i=1}^{n_0}(y_i-\theta)^2}{2} \Rightarrow \sigma_0^2 = \frac{1}{n_0}\sum(y_i-\theta)^2$$

Haldane's prior of absolute ignorance: n_0 = 0 => $P(\sigma^2|\theta) = InvGamma(0,0) = \frac{0^0}{\Gamma(0)}(\sigma^2)^{-0-1}e^{-\frac{0}{\sigma^2}} = \frac{1}{\sigma^2}$
sigsq_0 can be anything so by convention we say 0.

Laplace's prior of indifference: $P(\sigma^2|\theta) \propto 1$





$$\Rightarrow P(\sigma^2|X,\theta) = InvGamma\left(\frac{n}{2}, \frac{n\hat{\sigma}^2_{MLE}}{2}\right)$$

Is this a smart idea? This means that sigsq in [0,1] has the same weight as sigsq in [1000000000, 1000000001]. This is not a smart idea and no one really uses this prior to my knowledge.

What does this Laplace prior correspond to? Recall it results in a posterior of:

$$P(\sigma^2|X,\theta) = InvGamma\left(\frac{n-2}{2}, \frac{n\hat{\sigma}^2_{MLE}}{2}\right) \Rightarrow n_0 = -2,\ \sigma_0^2 = 0$$

$$\Leftarrow P(\sigma^2|\theta) = InvGamma\left(\frac{-2}{2}, \frac{0}{2}\right) = InvGamma(-1, 0)$$