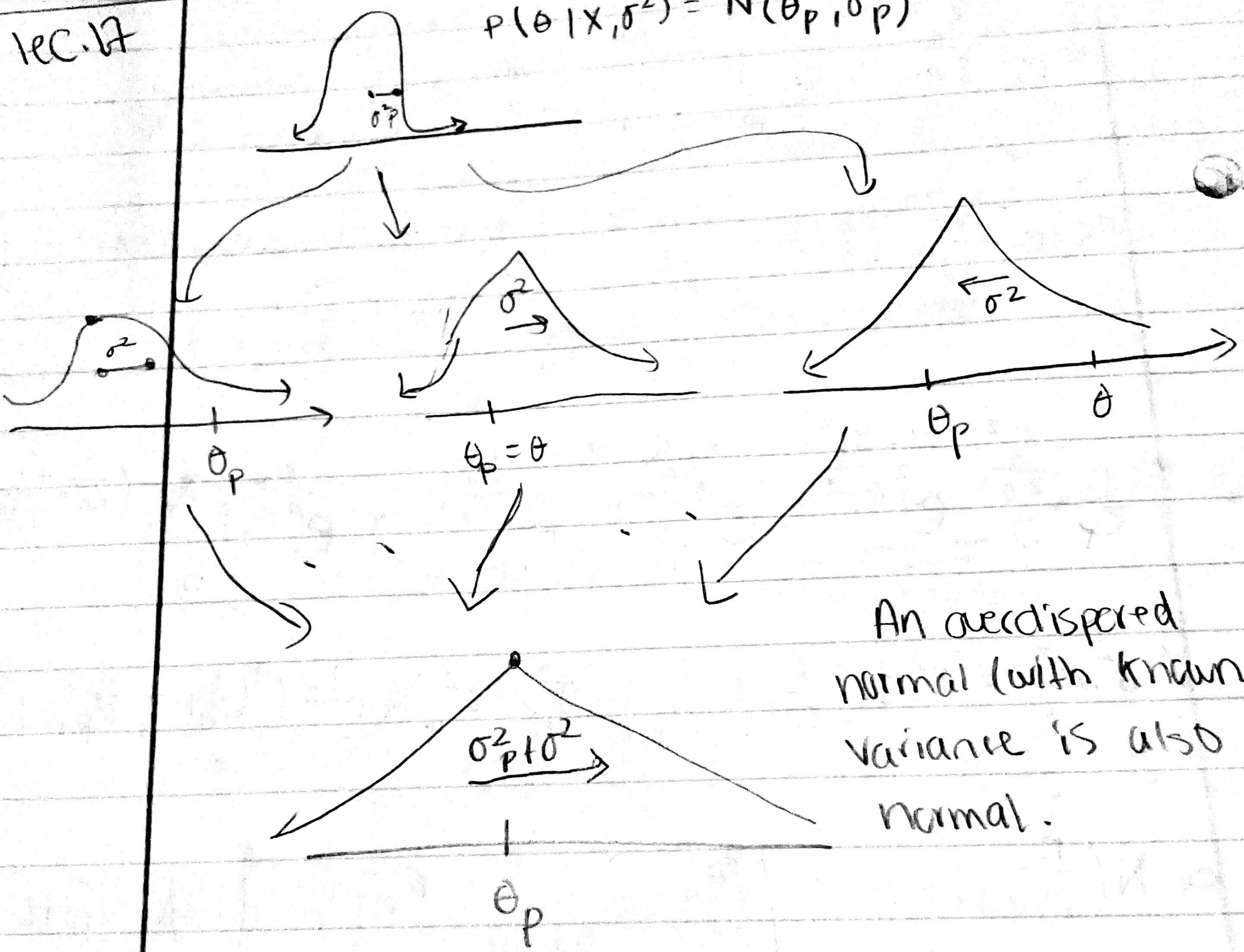
lec.17

$p(\theta \mid X, \sigma^2) = N(\theta_p, \sigma_p^2)$



An overdispersed normal (with known variance) is also normal.

Scanned with CamScanner

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

$$F: \overset{iid}{\sim} N(\theta, \sigma^2), \ \theta \text{ known}$$

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

$$= (2\pi)^{-n/2} (\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}\Sigma(x_i-\theta)^2}$$

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

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

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

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

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

MLE

$$= \frac{\Sigma(x_i-\theta)^2}{n}$$

average sqd deviation

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

$$k(y) = y^{-a} e^{-\frac{b}{y}}$$

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

$$\frac{1}{c} = \int_0^\infty k(y)dy = \int_0^\infty y^{-a} e^{-\frac{b}{y}} dy =$$

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

$y_1 = 0 \Rightarrow z = \infty$, $y = \infty = z = 0$

$$\int_\infty^0 z^a e^{-bz} (-z^{-2}) dz = \int_0^\infty$$

u-subst $\quad \frac{\Gamma(a-1)}{b^{a-1}} \Rightarrow p(y) = \frac{b^{a-1}}{\Gamma(a-1)} y^{a} e^{-\frac{b}{y}}$

traditionally $\quad \alpha = a-1 \Rightarrow a = \alpha+1$

$\beta = b$

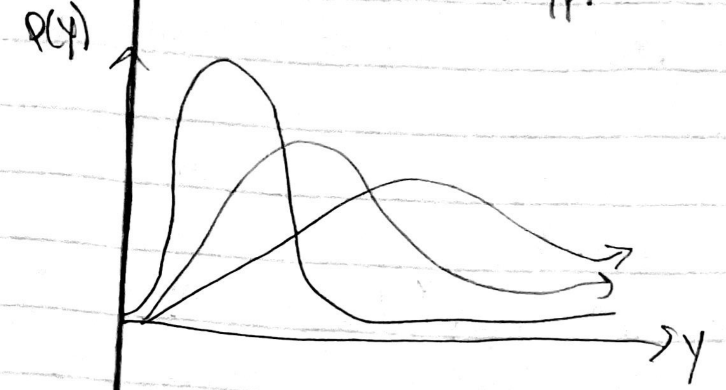
$$P(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-\alpha-1} e^{-\frac{\beta}{y}}$$

This is called the "inverse gamma" distribution

$= \text{InvGamma}(\alpha, \beta)$

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

$\alpha, \beta > 0$



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

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

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

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

Back to the regularly scheduled program...

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

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

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

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

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

what form should the prior be so that it's kernal has the same form as the posterior's kernal? It's an inverse gamma?

Let $P(\sigma^2|\theta) = \text{Inv Gamma}(\alpha, \beta)$

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

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

Traditionally, we use a different paramerization of the prior:

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

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

Bayesian point estimate for sigsq

$\alpha > 1$

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

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

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

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

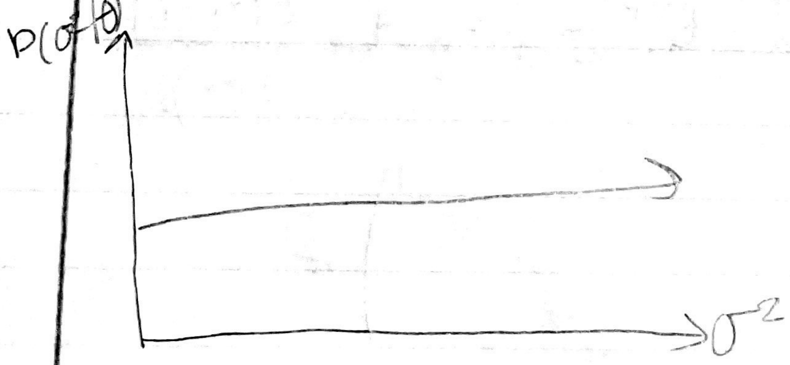
* ~~Credible Region? Same thing.. just use appropriate qinvgamma~~
Hypothesis Tests?
* Hypothesis Tests? Same thing.. just use appropriate pinvgamma

Pseudoobservation interpretation n_0 = # of pseudo observations
Imagine $y_1, y_2, \ldots, y_{n_0}$

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

Haldane's prior of absolute ignorance: $n_0 = 0 \Rightarrow P(\sigma^2|\theta) = \text{InvGamma}(0,0)$
Sigsq$_0$ can be anything so by convention we say 0.

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



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 knoweldge.

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

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

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