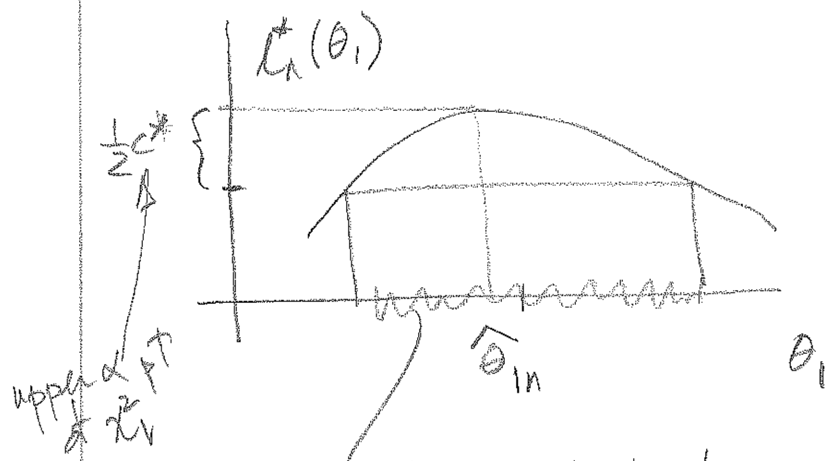
$l = 1$ picture



the approximate confidence set for θ_1

Insert $l=2$ picture

Summary of the big picture of large sample facts about 3 basic testing methods (LRT, Wald, and Rao/Score tests) in k-dimensional parameter problems

$$\underset{k\times 1}{\theta} = \begin{pmatrix} \underset{l\times 1}{\theta_1} \\ \underset{(k-l)\times 1}{\theta_2} \end{pmatrix} \qquad \text{Testing} \quad H_0 : \theta_1 = \theta_{10}$$

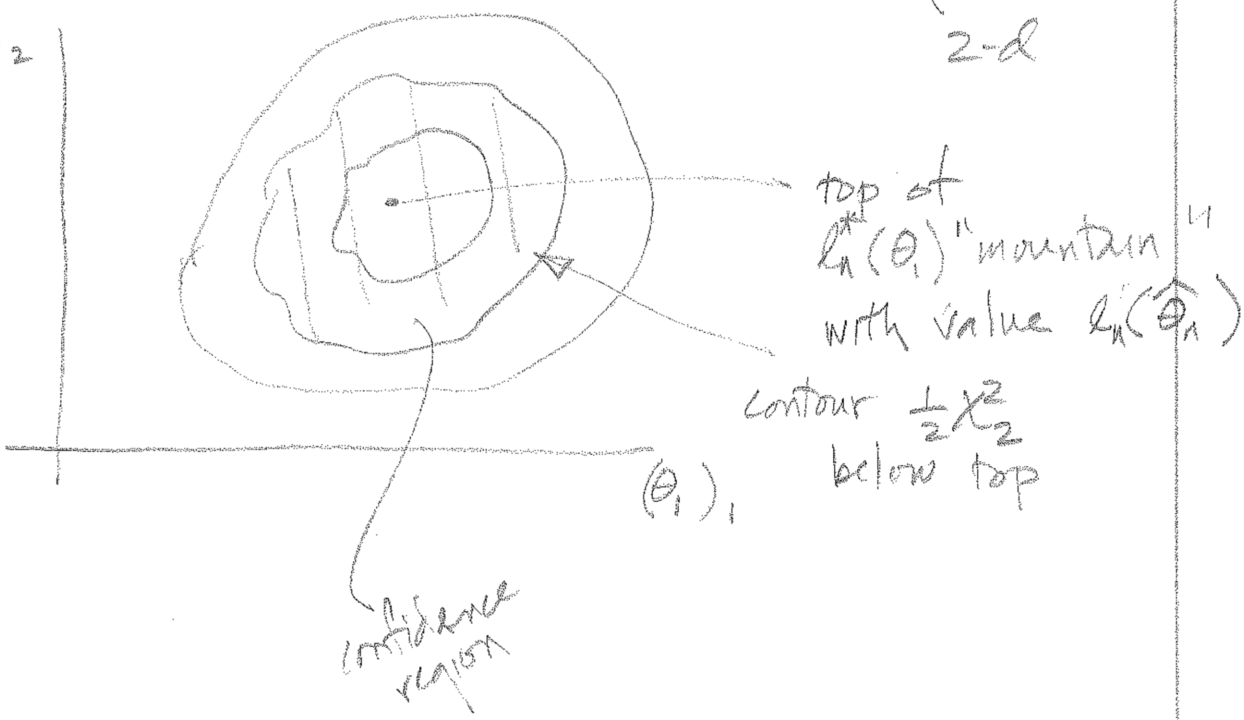
<u>LRT</u>

$$2\left(\underset{\substack{\uparrow \\ \text{maximum} \\ \text{log likelihood}}}{l_n(\hat{\theta}_n)} - \underset{\substack{\uparrow \\ \text{profile} \\ \text{log-likelihood}}}{l_n^*(\theta_{10})} \right) \xrightarrow{\mathcal{L}_{H_0}} \chi^2_l$$

$$\overbrace{2\left(\underset{\substack{\uparrow \\ \text{requires computing} \\ \text{of MLE}}}{\sup_\theta l_n(\theta_n)} - \underset{\substack{\uparrow \\ \text{requires computing} \\ \text{"restricted MLE"} \\ \text{for } \theta_2}}{\sup_{\theta_2} (\theta_{10}, \theta_2)} \right)}$$

$\lambda = 2$ picture contour plot of $\ell_n^*(\theta_1)$

2-d



Wald Test

$$(\hat{\theta}_{1n} - \theta_{10})' \begin{pmatrix} \text{estimated} \\ \text{covariance} \\ \text{matrix for} \\ \hat{\theta}_{1n} \end{pmatrix}^{-1} (\hat{\theta}_{1n} - \theta_{10}) \xrightarrow{\mathcal{L}_{H_0}} \chi^2_\ell$$

This requires computation of the (full) MLE

Rao (Score) Test

Motivation: If H_0 is true and $\theta^*_{2n}(\theta_{10})$ is a maximizer of $l_n(\theta_{10}, \cdot)$ then the score function at $(\theta_{10}, \theta^*_{2n}(\theta_{10}))$ should be nearly 0 (it's 0 at the MLE) — a plausible test statistic

$$R_n = \left(\frac{\partial}{\partial \theta_i} l_n(\theta) \Big|_{\theta = \begin{pmatrix} \theta_{10} \\ \theta^*_{2n}(\theta_{10}) \end{pmatrix}} \right)' \begin{pmatrix} \text{estimated} \\ \text{covariance} \\ \text{for } \hat{\theta}_n \text{ evaluated} \\ \text{at } \begin{pmatrix} \theta_{10} \\ \theta^*_{2n}(\theta_{10}) \end{pmatrix} \end{pmatrix} \Big(\quad \Big)$$

also $\xrightarrow{\mathcal{L}_{H_0}} \chi^2_\ell$

day 40

estimate of $(nI_1(\theta))^{-1}$ which is the inverse of $nI_1(\theta)$, which is the covariance matrix of the score function at "the truth"

Say a bit about what happens to Bayes posterior dens for large n —

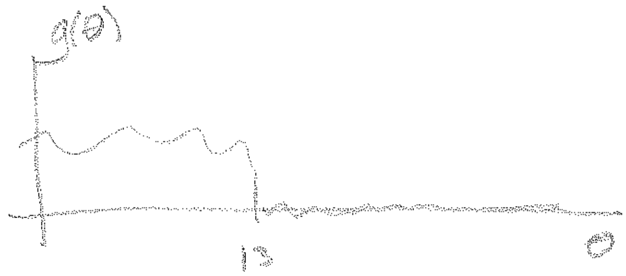
$L(\theta)$ likelihood

$g(\theta)$ prior

posterior $g(\theta|\text{data}) \propto L(\theta)g(\theta)$

What if the amount of data is large (say, e.g., L_n for an iid model and $n \to \infty$)? There are 2 points

1) The posterior cannot put any probability where the prior fails to do so



the prior probability that $\theta > 13 \Rightarrow$ the posterior probability that $\theta > 13$ is 0 — $\frac{1}{}$ const

$$g(\theta) = 0 \implies L(\theta)g(\theta) = 0$$

2) At least for large samples provided $g(\theta)$ spreads its probability, $g(\theta|x)$ will

posterior consist ⟹ a) tend to pile up near θ_0 under the θ_0 model

b) tend to look approximately normal in cases where Θ is cont's and this is possible

asymptotic normality of posterior

Lemma In an iid model if the θ and θ' dsns for X_1 are different

$$\frac{L_n(\theta')}{L_n(\theta)} \xrightarrow{P_\theta} 0$$

Pf: As in the proof of Thm 1 on ML handout (root of the likelihood equation near θ_0 w.p. $\to 1$ under θ_0)

$$\frac{L_n(\theta')}{L_n(\theta)} = \exp\left(-\log\frac{L_n(\theta)}{L_n(\theta')}\right)$$

$$= \exp\left(-\left(\ell_n(\theta) - \ell_n(\theta')\right)\right)$$

and for $\gamma > 0$ I can choose M large enough so that for $n > M$

$$P_\theta\left[\ell_n(\theta) - \ell_n(\theta') \geq \frac{n}{2}K(\theta,\theta')\right] \geq 1-\gamma$$

i.e. so that

$$P_\theta\left[\frac{L_n(\theta')}{L_n(\theta)} \leq \exp\left(-\frac{n}{2}K(\theta,\theta')\right)\right] \geq 1-\gamma$$

and since $\exp\left(-\frac{n}{2}K(\theta,\theta')\right)$ is eventually $\leq$ any positive number, we're done ∎

Corollary If Θ is finite, in an iid model if $g(\theta) > 0$ for each $\theta \in \Theta$ and no two dsns for X_1 (specified by $f(x|\theta)$) are the same, the posterior dsn $g_n(\cdot|X)$ is consistent in the sense that

$$g_n(\theta|X) \xrightarrow{P_\theta} 1$$

while $g_n(\theta'|X) \xrightarrow{P_\theta} 0$ for any $\theta' \neq \theta$

Pf:
$$g_n(\theta|X) = \frac{g(\theta) L_n(\theta)}{g(\theta) L_n(\theta) + \sum_{\theta' \neq \theta} g(\theta') L_n(\theta')}$$

$$= \frac{1}{1 + \sum_{\theta' \neq \theta} \frac{g(\theta')}{g(\theta)} \frac{L_n(\theta')}{L_n(\theta)}}$$

There are a finite # of these converging to 0 in θ probability

$$\xrightarrow{P_\theta} 1$$

$g_n(\theta|X)$ specifies a dsn on Θ. So $\theta' \neq \theta$ must have $g_n(\theta'|X) \xrightarrow{P_\theta} 0$. ∎

There are (more complicated) non-finite Θ versions of this kind of result that say that under appropriate conditions in an iid model with diffuse G, for $A \subset \Theta$ with $\theta \in A$

(G) posterior probability assigned to A $\xrightarrow{P_\theta} 1$

So RV's

(G) posterior probability assigned to A^c $\xrightarrow{P_\theta} 0$

That is, there are Thms about "consistency of posteriors" that really come from the fact that for large n, $L_n(\theta)$ tends to be increasingly peaked near the value θ under which one is computing -

There are also theorems that allow one to make large sample approximations to posterior dsns - e.g. there are theorems like

"Result" In an iid model where a prior dsn for $\theta \in \mathbb{R}^1$ has a density that is cont$\underline{s}$ and positive at θ_0 and regularity conditions hold, if $\delta_n(X)$ is the MLE of θ, under the θ_0 dsn for X the posterior density ← a random function of θ

of

$$\sqrt{-l_n''(\delta_n(X))}\,\left(\theta-\delta_n(X)\right)$$

converges to the std normal density $\left(\frac{1}{\sqrt{2\pi}}\exp-\frac{\theta^2}{2}\right)$

day 41

i.e. for large n the posterior density of θ tends to look like a normal density with

mean $\delta_n(X)$ and variance $\dfrac{1}{-l_n''(\delta_n(X))}$

the observed FI in X about θ

(This really comes from the fact that $l_n(\theta)$ tends to look quadratic with maximum near θ_0 - that makes the log posterior density (which is roughly proportional to it) also look quadratic with a maximum near θ_0

Example (of what this means) Suppose

$$\underbrace{X_1, X_2, \cdots, X_n \text{ iid Bernoulli}(p) \text{ and } p \sim \text{Beta}(\alpha, \beta)}_{\text{a Bayes model}}$$

Then the posterior dsn for p is $\text{Beta}(\alpha + \Sigma X, \beta + n - \Sigma X_i)$ - This is a random dsns with a random density proportional to

$$p^{\alpha + \Sigma X_i - 1} \quad (1-p)^{\beta + n - \Sigma X_i - 1}$$

Note that $\hat{\theta}_n(X) = \overline{X}_n$ and $-l_n''(\overline{X}_n) = \dfrac{n}{\overline{X}_n(1-\overline{X}_n)}$

What happens to the $\text{Beta}(\underbrace{\alpha + \Sigma X_i - 1}_{\tilde{\alpha}_n}, \underbrace{\beta + n - \Sigma X_i - 1}_{\tilde{\beta}_n})$ density as $n \to \infty$?

That is, what can be said about

$$g_{\tilde{\alpha}_n, \tilde{\beta}_n}(p) \ ?$$

Well, density for $\sqrt{\dfrac{n}{\overline{X}_n(1-\overline{X}_n)}}\,(p - \overline{X}_n) = u$

namely

$$\sqrt{\frac{\overline{X}_n(1-\overline{X}_n)}{n}}\; g_{\tilde{\alpha}_n, \tilde{\beta}_n}\left(\frac{u}{\sqrt{\frac{n}{\overline{X}_n(1-\overline{X}_n)}}} + \overline{X} \right) \xrightarrow{P_{p_0}} \phi(u)$$

i.e. The function on the left converges at every u to $\phi(u)$ and this happens $\forall\ p_0 \in (0,1)$ at every u (no matter what be (α,β)) —
the (α,β) washes out! Any Beta prior will have a posterior density that is approximately normal in this sense — of course how big n has to be for this approximation to be any good with large p_0 probability depends upon all of p_0, u, α, and β —

Anyway convergence of posterior density (pointwise in u) is pretty easy to prove (and there are multiparameter versions of it) —

Some insight into where this comes from is this (this provides no tight argument, just some rough motivation)

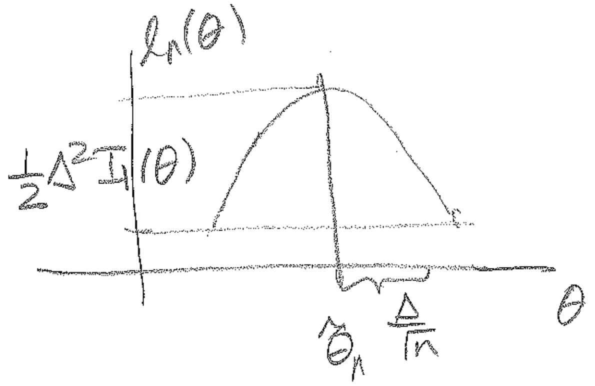
Lemma (not really enough to show posterior is normal, for that we'd need some uniformity in Δ)
Under regularity conditions in an iid model, where $\Theta \in \mathbb{R}^1$ and $\hat{\theta}_n$ is a consistent MLE of θ, for a fixed $\Delta \in \mathbb{R}^1$,

$$Q_n(\Delta) = l_n(\hat{\theta}_n) - l_n\left(\hat{\theta}_n + \frac{\Delta}{\sqrt{n}}\right) \xrightarrow{P_\theta} \frac{1}{2}\Delta^2 I_1(\theta)$$

Pf Expand $l_n(\cdot)$ around $\hat{\theta}_n$ in a Taylor series ...

∎

The point here is that if you stayed a fixed distance from $\hat{\theta}_n$, with increasing n, you expect the log-likelihood to drop further and further below the maximum log-likelihood — this Q_n thing captures the local shape of the loglikelihood — as you come into $\hat{\theta}_n$ at rate $\frac{1}{\sqrt{n}}$ it looks quadratic — of course it does! locally near its maximum it is flat, so it must then be quadratic (1st Taylor term is 0, so 2nd is what matters)



Corollary Supposing that a prior G has a pdf $g(\cdot)$ on $\Theta \subset \mathbb{R}$, $g(\theta_0) > 0$ and $g(\cdot)$ is cont$^{\underline{s}}$ at θ_0, if $\hat{\theta}_n$ is an MLE of θ consistent at θ_0, the posterior density $g_n(\theta \mid X)$ has the property that

$$R_n(\Delta) = \log \frac{g_n(\hat{\theta}_n | X)}{g_n(\hat{\theta}_n + \frac{\Delta}{\sqrt{n}} | X)} \xrightarrow{P_{\theta_0}} \frac{1}{2}\Delta^2 I_1(\theta_0)$$

<u>Pf</u>:

$$R_n(\Delta) = \log \frac{g(\hat{\theta}_n)\, L_n(\hat{\theta}_n)}{g(\hat{\theta}_n + \frac{\Delta}{\sqrt{n}})\, L_n(\hat{\theta}_n + \frac{\Delta}{\sqrt{n}})}$$

$$= \log g(\hat{\theta}_n) - \log g(\hat{\theta}_n + \tfrac{\Delta}{\sqrt{n}})) \xrightarrow{P_{\theta_0}} 0$$

$$+ \underbrace{\left(\ell_n(\hat{\theta}_n) - \ell_n(\hat{\theta}_n + \tfrac{\Delta}{\sqrt{n}})\right)}_{Q_n(\Delta)}$$

What does this say? It says that near the MLE the shape of the log posterior density is the same as the shape of the log-likelihood, namely quadratic

Think what this would mean if the approximations were exact (and uniform, etc.)

Q: If Y is cont$\underline{s}$ with pdf f and

$$\ln \frac{f(0)}{f(y)} = \frac{c}{2} y^2$$ what is the dsn of Y?

A: $$\ln\left(\frac{f(y)}{f(0)}\right) = -\frac{c}{2}y^2$$

$$\frac{f(y)}{f(0)} = \exp - \frac{c}{2} y^2$$

$$f(y) = f(0) \exp - \frac{c}{2} y^2$$

and clearly Y is normal with mean 0 and variance $\frac{1}{c}$

This at least strongly suggests that under P_{θ_0} I can expect posterior densities to start to look normal with mean $\hat{\theta}_n$ and variance $\frac{1}{nI(\theta_0)}$