UNCERTAINTY ASSESSMENT

What good is an estimate without a $\pm$?

Ad-hoc approach to *confidence intervals*: establish *sampling distribution* of estimator $\hat{\theta}$ for θ, often of form $\hat{\theta} \sim N(\theta, \nu^2)$.

Report an interval having desired (95%) chance of straddling θ (probability *wrt* data given parameter).

For instance, $\hat{\theta} \pm 1.96\nu$ (usually need $\hat{\nu}$ since ν unknown).

Always easy to work out sampling distribution?

ASIDE: interpretation of sampling distribution of estimator and CI?

UA WITH ML METHODS

Fortunately, large-sample theory can automate CI formulation:

Recall $\hat{\theta}$ maximizes log-likelihood $l(\theta)$, i.e., $l'(\theta) = 0$. Then

$$\hat{\theta} \quad \pm \quad 1.96\sqrt{\frac{1}{-l''(\hat{\theta})}}$$

is a (large n) approx. 95% CI for θ

Role of $2nd$ derivative makes sense.

Matrix analogue when $dim(\theta) > 1$.

Complete paradigm: prob model $\rightarrow$ likelihood function $\rightarrow$ point and interval estimates.

BAYESIAN UA: more direct?

Have posterior distribution for θ given data. Report an interval having 0.95 probability under this distribution as a 95% *credible interval* for θ.

For instance, take the 2.5% and 97.5% percentiles of the posterior distribution.

At least in simple problems, ML and Bayes often give similar interval estimates numerically, though the interpretation is different.

Complete paradigm: prob. models for θ and $(DATA|\theta) \rightarrow$ post. dist. for $(\theta|DATA) \rightarrow$ point & interval estimates.

HYPOTHESIS TESTING: null (H_0) versus alternative (H_1).

Ad-hoc: find a 'statistic' T such that if H_0 is true, T has known distribution. Compare observed value of T to this dist.

E.g., $Y_1, \ldots, Y_n$ iid mean μ. Test $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$.
Take $T = \bar{Y}/(S/\sqrt{n})$. If H_0 true then (large-n approx.)
$T \sim N(0,1)$.

For significance level 0.05 test, reject H_0 if $|T_{obs}| > 1.96$.

Or report P-value: $Pr\{|T| > |T_{obs}| \;\; | H_0\}$.

PROS:
- well entrenched.

CONS:
- confusion: magnitude of evidence/effect,
- asymmetry of null (specific) & alternative (general),
- can't quantify evidence *for* null,
- false non-rejection rate (power) easily ignored.

HYPOTHESIS TESTING WITH ML METHODS

Automatable via large-sample theory: Wald, likelihood-ratio, and score tests.

Wald test for null that $\theta = \theta^*$: Compare $T = (\hat{\theta} - \theta^*)/SE(\hat{\theta})$ to $N(0,1)$. Equivalent to 'inverting' confidence interval.

LR test. Have log-likelihoods $l_0()$ and $l_1()$ for null and alternative. Compare $T = 2\{l_1(\hat{\theta}_1) - l_0(\hat{\theta}_0)\}$ to χ^2_d distribution, where null had d fewer 'free' parameters than alternative.

All three tests equally justified asymptotically, but literature on 'small-sample' differences (LR better than Wald in some settings).

BAYESIAN HYPOTHESIS TESTING

Extend formulation of probability describing params given data to hypotheses and params given data.

More technically, prior dist: $Pr(H, \theta_H) = Pr(H)Pr(\theta_H|H)$.

Can set $Pr(H_0)$, then compare with $Pr(H_0|DATA)$.

CONS:
- can be computationally devilish,
- answers can be sensitive to $Pr(\theta_H|H)$ specification.

PROS:
- dealing with $(HYP|DATA)$, not $(DATA|HYP)$.

THE PHARMACY SHELF

To get regulatory approval for your new drug, you need $P < 0.05$ (relative to placebo, say).

What percentage of drugs available on the pharmacy shelf are ineffective???

Answer depends on a couple of things.

Let q be the proportion of proposed drugs that are actually effective (quite small?).

Let r be the power of the clinical trial, e.g., $Pr(|T| > 1.96|\text{effective})$. (Gross oversimplification - each study designed to have specific power for specific effect size.)

Now do the math....

$$Pr\,\{\text{ineffective}\,|\quad |T| > 1.96\,\} = \ldots$$

	$r = 0.5$	$r = 0.8$	$r = 0.95$
$q = 0.5$			
$q = 0.2$			
$q = 0.1$			

Caveat emptor!

GENERAL THOUGHTS ON STAT PRINCIPLES

Many are pragmatic, will adopt whatever techniques work well regardless of underlying principles. There are lots of criteria by which to measure the *performance* of a statistical procedure regardless of paradigm (bias, mean-squared error, coverage, avg. interval length, predictive performance). [Later in 545 will discuss simulation studies, cross-validation.]

Both ML and Bayesian analysis have some "best possible" theory to suport their use.

But in complex problems ML methods demand a lot (likelihood function can be hard to compute/maximize) and Bayesian methods even more (both computation, and prior specification). Always worth it?