# How to make a confidence/credible interval for nearly any quantity!

Matthew S. Shotwell, Ph.D.

Department of Biostatistics
Vanderbilt University School of Medicine

August 4, 2017

# Big ideas



- Quantifying statistical uncertainty is important.
- Sometimes interested in unusual quantities:
  - error variance
  - random effects variances
  - "nuisance" parameters
  - functions of parameters

# What about bootstrap?

DEPARTMENT OF BIOSTATISTICS
VANDERBILT UNIVERSITY

The bootstrap works but has drawbacks:

- computationally intense
- estimability issues (e.g., sparse categories)
- Monte carlo error (i.e., not deterministic)
- no Bayesian version
- not good for very small samples; $\binom{2n-1}{n-1}$ possible bootstraps
- complicated for correlated data

# Another solution.



This talk will focus on approximations:

- computationally easy (no bootstrap or MCMC)
- no estimability issues
- deterministic
- applies in likelihood and Bayesian context
- works for very small samples (but may not be very good)
- not complicated for correlated data

# Likelihood vs posterior



Quantifying uncertainty is based on:

- likelihood function: $L(\theta|D)$
- posterior density: $P(\theta|D) \propto L(\theta|D)P(\theta)$
- generic: $P(\theta)$

# Taylor approximation



The log of $P(\theta)$ can be approximated using a second-order Taylor approximation about $\theta'$ as follows:

$$\log P(\theta) \approx \log P(\theta') + G(\theta')(\theta - \theta') + \frac{1}{2}(\theta - \theta')^T H(\theta')(\theta - \theta')$$

where the **gradient** of $\log P(\theta)$ at $\theta'$ is

$$G(\theta') = \left. \frac{\partial \log P(\theta)}{\partial \theta^T} \right|_{\theta = \theta'}$$

and the **Hessian** of $\log P(\theta)$ at $\theta'$ is

$$H(\theta') = \left. \frac{\partial^2 \log P(\theta)}{\partial \theta^T \partial \theta} \right|_{\theta = \theta'}$$

# Normal approximation



If we let $\theta' = \hat{\theta}$ be the value that maximizes $\log P(\theta)$ (i.e., a maximum likelihood estimate [MLE] or maximum *a posteriori* [MAP] estimate) and exponentiate both sides, we have the following:

$$P(\theta) \approx K \exp\left[-\frac{1}{2}(\theta - \hat{\theta})^T \hat{\Sigma}^{-1} (\theta - \hat{\theta})\right]$$

where $K$ is a constant with respect to $\theta$ and $\hat{\Sigma}^{-1} = -H(\theta')$. From this, we draw the following conclusions/connections:

- The posterior density can be approximated by a multivariate normal density with mean $\hat{\theta}$ and variance-covariance $\hat{\Sigma}$ (this is identical to a Laplace approximation).
- It's no coincidence that the approximate sampling distribution of the MLE is the same multivariate normal distribution.
- Easy confidence/credible intervals: $\hat{\theta}_j \pm 1.96\sqrt{\hat{\Sigma}_{jj}}$

# Uncertainty captured



In both the likelihood (MLE) and Bayesian context, our uncertainty about $\theta$ is captured (approximately) by the multivariate normal density with mean $\hat{\theta}$ and variance-covariance $\hat{\Sigma}$.

# What about functions of parameters?



Say we want to quantify uncertainty about a function $g(\theta)$. Approximate (essentially the delta method)!

$$g(\theta) \approx g(\theta') + G(\theta')(\theta - \theta')$$

Since uncertainty about $\theta$ (Bayesian) or $\theta' = \hat{\theta}$ (likelihood) is captured by the normal distribution, then the same is approximately true for $g(\theta)$. In both the likelihood and Bayesian case, uncertainty about $g(\theta)$ is quantified by the following:

$$g(\hat{\theta}) \sim N(\hat{\theta}, G(\hat{\theta})\hat{\Sigma}G(\hat{\theta})^T)$$

So, a 95% CI is given as follows

$$g(\hat{\theta}) \pm 1.96\sqrt{G(\hat{\theta})\hat{\Sigma}G(\hat{\theta})^T}$$

# Examples in R!



Examples in R!

- computing likelihood can be hard (e.g., mixed effects models)
- some model fitting routines give us parts for free (e.g., `vcov`)
- usually necessary to code likelihood for Bayesian models