# 4: Asymptotic and connections to non-Bayesian approaches

09/18/19

# Asymptotic Normality: cases of unmet assumptions

We go through some common examples where one of the above assumptions is not met. In these cases, using asymptotics is not allowed.

# Asymptotic Normality: cases of unmet assumptions

A *model* is **underidentified** given data $y$ if the likelihood, $p(y \mid \theta)$, is equal for a range of values $\theta$.

A *model* is **weakly identified** given data $y$ if the likelihood, $p(y \mid \theta)$, is close to being equal for a range of values $\theta$.

These can be problematic because $\hat{\theta}$ will not have any specific number/vector $\theta$ to which it can converge. These are violations of assumption (3).

# Asymptotic Normality: cases of unmet assumptions

$$\left[\begin{array}{c} u \\ v \end{array}\right] \Bigg| \rho \sim \text{Normal}\left(\left[\begin{array}{c} 0 \\ 0 \end{array}\right], \left[\begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array}\right]\right)$$

If $v$ is latent/hidden, then we work with the marginal likelihood $p(u \mid \rho)$:

$$u \mid \rho \sim \text{Normal}\,(0, 1)$$

Notice that this is free of $\rho$!

$$p(\rho \mid u) \propto p(u \mid \rho) p(\rho) \propto p(\rho)$$

Here we say the *parameter* is **nonidentified**.

# Asymptotic Normality: cases of unmet assumptions

Sometimes it is harder to spot nonidentifiable parameters. It may be the case that $p(y \mid \theta)$ yields the same function in $y$ for two different values of $\theta$. If this is true, then for any particular data set $y$, $p(y \mid \theta)$ will be equal for these two values of $\theta$.

Example $y \mid \theta \sim \text{Normal}(0, \theta^2)$. Then $p(y \mid \theta) = p(y \mid -\theta)$!

We can fix this easily by restricting the parameter space. The model is no longer underidentified if $\theta \in \mathbb{R}^+$. When this happens, we call this problem **aliasing**.

# Asymptotic Normality: cases of unmet assumptions

Another example of **aliasing**. If you look at a histogram of $y$ and it's bimodal, then a possibly suitable model is the **normal mixture model**:

$$
\begin{aligned}
&p(y_i \mid \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \lambda) \\
&= \lambda \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left[-\frac{1}{2\sigma_1^2}(y_i - \mu_1)^2\right] \\
&+(1-\lambda)\frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left[-\frac{1}{2\sigma_2^2}(y_i - \mu_2)^2\right]
\end{aligned}
$$

# Asymptotic Normality: cases of unmet assumptions

**Unbounded likelihoods** might also be a problem. Assume

$$p(y \mid \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{y^2}{2\sigma^2}\right].$$

If $y = 0$, then this simplifies to

$$p(y \mid \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}$$

which goes to $\infty$ as $\sigma^2 \to 0$. The theoretical probablity of you getting $y = 0$ is obviously 0, but it is possible to get 0s computationally if you have an **underflow** problem. Double precision floating point numbers give you about 15-17 digits of precision.

# Asymptotic Normality: cases of unmet assumptions

- Improper posterior distributions

$$y \sim \text{Binomial}(n, p), p \sim \text{Beta}(0, 0)$$

$$p \mid y \propto p^{y-1}(1-p)^{n-y-1}$$

  $Pr(p \mid 0)$ or $Pr(p \mid n)$ is improper!
- Prior distributions that exclude $\theta_0$
- Convergence to the edge of parameter space, $\theta_0 \in \partial\Theta$
- Tails of the distribution is less accurate

# Asymptotic Normality: difficulty in high dimension

The number of parameters increases with the sample size, the standard asymptotics won't apply.

- Consistency is hard to obtain (not impossible though): prior distribution plays a much bigger role; the sample size is not big in each dimension of $\theta$
- Normal approximation is of high dimension
- Asymptotic normality is not efficient even if we can

It makes more sense to consider other asymptotic properties (consistency, convergence rates) instead by putting restriction on $\theta_0$ and the prior as well.
When the number of parameter is large, hierarchical prior is preferred since then their common distribution can be estimated from data.

# Frequency evaluations of Bayesian inferences

- Large sample correspondence, $\hat{\theta}$, MLE

$$(nJ(\hat{\theta}))^{-1}(\theta - \hat{\theta})|y \sim N(0, I)$$

$$(nJ(\theta_0))^{-1}(\hat{\theta} - \theta_0)|\theta_0 \sim N(0, I)$$

  $\Rightarrow$ Bayesian posterior credible interval is asymptotically the same as confidence interval in repeated sampling

- Consistency/ Asymptotic unbiasedness/ Asymptotic efficiency of point estimate