p-Values Without Penalties With Perfect Predictions*

Carlisle Rainey†

Separation commonly occurs in political science, usually when the presence (or absence) of a binary explanatory variable perfectly predicts the presence or absence of a binary outcome (e.g., Bell and Miller 2015; Mares 2015; Vining, Wilhelm, and Collens 2015). Under separation, maximum likelihood estimation leads to infinite coefficient estimates and standards errors. In practice, though, optimization routines converge before reaching infinite estimates and return implausibly large finite estimates and standard errors.

As an example of an implausible estimate, consider the model fit by Barrilleaux and Rainey (2014). For their application, the maximum likelihood estimates produced by the default `glm()` routine in R suggest that a governor like Deval Patrick, the Democratic governor of Massachusetts, had about a one in ten *billion* chance of opposing the Medicaid expansion under the Affordable Care Act. To give some perspective, this is *less* likely than you tossing 33 consecutive heads (around 1.2 in ten billion[1], you dealing a new poker player's first two hands as five-card straight flushes (around 2.4 in ten billion[2]),

*Draft under developement. This is the version as of September 5, 2019.

†Carlisle Rainey is Associate Professor of Political Science, Florida State University, 540 Bellamy, Tallahassee, FL, 32306. (crainey@fsu.edu).

[1]The probability of tossing 33 consecutive heads equals $\left(\frac{1}{2}\right)^{33} \approx 1.16 \times 10^{-10}$. If you tossed one coin per second for 24 hours, then you could complete 2,618 33-toss trials in one day. To obtain an all-head 33-toss sequence, you would need to continue this routine daily for about 10,000 years.]

[2]There are $\binom{52}{5} = 2,598,960$ ways to draw five cards from a 52-deck. There are ten straight flushes per

an average golfer making aces on their next two attempts on a par-3 hole (around 64 in ten billion[3]). It would take about ten billion years before a similarly situated Democratic governor would oppose the ACA–a little less than the age about the universe (about 13 billion years), but more than 30,000 *times* longer than *Homo sapiens* have existed (about 315,000 years) and about two million *times* longer than taxes have existed (about 5,000 years).

As a solution, Zorn (2005; see also Heinze and Schemper 2002) points political scientists toward the penalized maximum likelihood estimator proposed by Firth (1993). Even under separation, penalized maximum likelihood ensures finite estimates in theory and usually produces reasonably-sized estimates in practice. Conceptually, penalized maximum likelihood uses Jeffreys prior (Jeffreys 1946) to shrink the maximum likelihood estimates toward zero (Firth 1993).

Rainey (2016) points out that the parameter estimates (and especially the confidence intervals) depend largely on the chosen penalty. Indeed, other priors also guarantee finite, but different, estimates. For example, Gelman et al. (2008) recommend a Cauchy prior distribution. Rainey (2016) argues that the set of "reasonable" and "implausible" estimates depends on the substantive application, so context-free defaults (like Jeffreys and Cauchy) might not produce reasonable results. Rainey (2016) concludes that "[w]hen facing separation, researchers must *carefully* choose a prior distribution to nearly rule out implausibly large effects" (p. 354).

But researchers cannot access useful prior information in all contexts, and some scholars prefer to avoid injecting prior information into the model. How can researchers proceed in these situations? Below, I show that while maximum likelihood produces

suit and four suits, so there are $10 \times 4 = 40$ total royal flush possibilities. The chance of a straight flush is then $\frac{40}{2,598,960} \approx 1.54 \times 10^{-5}$. The chance of two consecutive straight flushes is twice this chance. If two poker players sat down with a dealer that shuffled and dealt two five-card hands per minute, those players would need to play for about 8,000 years before the dealer dealt both a straight flush.]

[3]The chance that an average golfer makes an ace is about one in 12,500. The chance of two consecutive aces is twice that. If an average golfer played a typical nine-hole course every day, it would take about 430,000 years before they would ace both par-3 holes.

implausibly large estimates under separation and standard errors, standard likelihood ratio tests behave in the usual manner. As such, researchers can produce meaningful p-values with a standard, well-known tool even while eyeing the coefficient estimates with suspicion.

Statistial Theory

Wald Test

Likelihood Ratio Test

Illustrations

To illustrate the simplicity of hypothesis testing cunder separation compared to estimation, I reanalyze data from Barrilleaux and Rainey (2014) and Bell and Miller (2015) the that Rainey (2016) considers in great detail. For Barrilleaux and Rainey (2014), the likelihood ratio test provides a useful evaluation of their substantive claims. For Bell and Miller (2015) the likelihood ratio test proves less useful.

Barrilleaux and Rainey (2014)

Barrilleaux and Rainey (2014) examine U.S. state governors decisions to support or oppose the Medicaid expansion under the 2010 Affordable Care Act. But because all Democratic governors supported the expansion, separation occurs–Democratic governors perfectly predict support for Medicaid expansion.

The R code below fits both full model from Barrilleaux and Rainey (2014) (shown in their Figure 2) and the null model where partisanship has no effect on governors' decisions to support or oppose the Medicaid expansion.

Variable	Estimator	Coefficient Est.	SE Est.	Wald p-Value	LR p-Value	Percent Change
Democratic Governor	ML with Default Precision	-20.35	3,224	0.99	0.00	-1.00
	ML with Maximum Precision	-35.22	15 million	1.00	0.00	-1.00
	PML with Jeffreys penalty	-2.68	1.42	0.06		
	PML with Cauchy penalty	-3.38	1.63	0.04		
Percent Without Health Insurance	ML with Default Precision	0.92	2.23	0.68	0.68	0.00
	ML with Maximum Precision	0.92	2.23	0.68	0.68	0.00
	PML with Jeffreys penalty	0.18	1.13	0.87		
	PML with Cauchy penalty	0.60	1.08	0.58		
Percent Favorable to ACA	ML with Default Precision	0.13	1.55	0.93	0.93	0.00
	ML with Maximum Precision	0.13	1.55	0.93	0.93	0.00
	PML with Jeffreys penalty	-0.14	1.31	0.92		
	PML with Cauchy penalty	-0.21	1.04	0.84		
GOP Legislature	ML with Default Precision	2.43	1.48	0.10	0.06	-0.38
	ML with Maximum Precision	2.43	1.48	0.10	0.06	-0.38
	PML with Jeffreys penalty	1.62	1.17	0.17		
	PML with Cauchy penalty	1.70	1.06	0.11		
Fiscal Health	ML with Default Precision	-0.05	0.85	0.95	0.95	0.00
	ML with Maximum Precision	-0.05	0.85	0.95	0.95	0.00
	PML with Jeffreys penalty	-0.12	0.73	0.87		
	PML with Cauchy penalty	0.15	0.75	0.84		
Medicaid Multiplier	ML with Default Precision	-0.35	1.19	0.77	0.77	0.00
	ML with Maximum Precision	-0.35	1.19	0.77	0.77	0.00
	PML with Jeffreys penalty	-0.33	1.02	0.75		
	PML with Cauchy penalty	-0.16	0.88	0.85		
Percent Non-White	ML with Default Precision	1.43	2.62	0.58	0.58	-0.01
	ML with Maximum Precision	1.43	2.62	0.58	0.58	-0.01
	PML with Jeffreys penalty	1.56	1.21	0.20		
	PML with Cauchy penalty	0.93	1.24	0.45		
Percent Metropolitan	ML with Default Precision	-2.76	1.69	0.10	0.06	-0.45
	ML with Maximum Precision	-2.76	1.69	0.10	0.06	-0.45
	PML with Jeffreys penalty	-1.82	1.19	0.13		
	PML with Cauchy penalty	-1.46	1.04	0.16		
Constant	ML with Default Precision	11.49	1,934	1.00	0.09	-0.90
	ML with Maximum Precision	20.42	9 million	1.00	0.09	-0.91
	PML with Jeffreys penalty	1.18	1.15	0.30		
	PML with Cauchy penalty	1.47	1.19	0.22		

References

Barrilleaux, Charles, and Carlisle Rainey. 2014. "The Politics of Need: Examining Governors' Decisions to Oppose the 'Obamacare' Medicaid Expansion." *State Politics and Policy Quarterly* 14(4): 437–60.

Bell, Mark S., and Nicholas L. Miller. 2015. "Questioning the Effect of Nuclear

Weapons on Conflict." *Journal of Conflict Resolution* 59(1): 74–92.

Firth, David. 1993. "Bias Reduction of Maximum Likelihood Estimates." *Biometrika* 80(1): 27–38.

Gelman, Andrew, Aleks Jakulin, Maria Grazia Pittau, and Yu-Sung Su. 2008. "A Weakly Informative Prior Distribution for Logistic and Other Regression Models." *The Annals of Applied Statistics* 2(4): 1360–83.

Heinze, Georg, and Michael Schemper. 2002. "A Solution to the Problem of Separation in Logistic Regression." *Statistics in Medicine* 21(16): 2409–19.

Jeffreys, H. 1946. "An Invariant Form of the Prior Probability in Estimation Problems." *Proceedings of the Royal Society of London, Series A* 186(1007): 453–61.

Mares, Isabela. 2015. *From Open Secrets to Secret Voting: Democratic Electoral Reforms and Voter Autonomy*. Cambridge: Cambridge University Press.

Rainey, Carlisle. 2016. "Dealing with Separation in Logistic Regression Models." *Political Analysis* 24(3): 339–55.

Vining, Richard L., Jr., Teena Wilhelm, and Jack D. Collens. 2015. "A Market-Based Model of State Supreme Court News: Lessons from Captial Cases." *State Politics and Policy Quarterly* 15(1): 3–23.

Zorn, Christopher. 2005. "A Solution to Separation in Binary Response Models." *Political Analysis* 13(2): 157–70.