INFINITE DIAMETER CONFIDENCE SETS IN A MODEL FOR PUBLICATION BIAS

A PREPRINT

Jonas Moss
Department of Mathematics
University of Oslo
PB 1053, Blindern, NO-0316, Oslo, Norway
jonasmgs@hotmail.no

December 20, 2019

ABSTRACT

There is no confidence set of guaranteed finite diameter for the mean and heterogeneity parameters in the selection function publication bias meta-analysis model where the selection function is a step function. The proof is based on a minor generalization of the Gleser-Hwang theorem.

Keywords publication bias · meta-analysis

1 Introduction

Meta-analysis is the quantitative combination of results from different studies [Lipsey and Wilson, 2001]. The purpose of a meta-analysis is to pool estimates across studies in order to reduce sampling error. Publication bias occurs when the published research literature is not representative of all the research that has actually been done [Rothstein et al., 2006]. An important case of publication bias is when publication decisions are made based on p-values [Sterling, 1959]. Such p-value based publication bias (and its sister phenomenon, p-hacking [Simmons et al., 2011]) can cause seriously biased conclusions when they are not accounted for [Simmons et al., 2011, Moss and De Bort, 2019].

One of the most important models for meta-analysis is the normal random effects model with normal likelihoods where the standard deviations are assumed to be known [Hedges and Vevea, 1998]. The likelihood for an observation in this model is

\[ \phi \left(x_i \mid \theta_0, \left(\sigma_i^2 + \tau^2\right)^{1/2}\right) \]

where \( \phi (x \mid \mu, \sigma) \) is the normal density with mean \( \mu \) and standard deviation \( \sigma \). The parameter \( \tau \) is the standard deviation of the random effects distribution, and is known as the heterogeneity parameter, while \( \sigma_i \) are the study specific standard errors and \( \theta_0 \) is the mean of the effect size distribution. A modification of the random effects meta-analysis model essentially due to [Hedges, 1984] allows us to account for p-value based publication bias.

Let \( u = \Phi \left(-x_i / \sigma_i\right) \) be the standard one-sided p-value for the null hypothesis \( \theta_i = 0 \), and \( w(u) \) a probability for each \( u \). The selection function meta-analysis model [Hedges, 1984, 1992] based on the p-value \( u \) is

\[ f \left(x_i \mid \theta_0, \left(\sigma_i^2 + \tau^2\right)^{1/2}\right) \propto \phi \left(x_i \mid \theta_0, \left(\sigma_i^2 + \tau^2\right)^{1/2}\right) w(u) \] (1.1)

The model is also known as the weighting function publication bias model. Here is an interpretation of it:

Alice is an editor who receives a study with p-value \( u \). Her publication decision is a random function of this p-value. That is, she will publish the study with some probability \( w(u) \). Every study you will ever read in Alice’s journal has survived this selection mechanism, the rest are lost forever.

This is essentially describes a rejection sampling [von Neumann, 1951, Flury, 1990] procedure. The publication probability function is likely to be well approximated by a step function, with cutoffs such as 0.01, 0.025 and 0.05
being especially important. To model this, let \( \alpha \) be a vector with \( 0 = \alpha_0 < \alpha_1 < \cdots < \alpha_J = 1 \) and \( \rho \) be a non-negative vector in \([0, 1]^J\) with \( \rho_1 = 1 \). Now define \( w(u \mid \rho, \alpha) = \sum_{j=1}^{J} \rho_j 1(\alpha_{j-1}, \alpha_j) \) \( u \). This is a step function where the value of \( w \) on each step \( [\alpha_{j-1}, \alpha_j] \) is the probability of acceptance for a study with a \( p \)-value \( u \) falling inside the interval \( (\alpha_{j-1}, \alpha_j] \). I will call the density proportional to \( \phi \left( x_{i} \mid \theta_{0}, (\sigma_{i}^{2} + \tau^{2})^{1/2} \right) w(u \mid \rho, \alpha) \) the step function publication bias model and denote it \( f \left( x_{i} \mid \theta_{0}, (\sigma_{i}^{2} + \tau^{2})^{1/2} \right) \).

The step function publication bias model was first used by Hedges [1984], who used an \( F \) distribution instead of a the normal distribution and a single step in \( w(u \mid \rho, \alpha) \). Iyengar and Greenhouse [1988] studied other choices of \( w \) while Ciklewicz and Vevera [2017] used a beta density publication probability function instead of a step function. Hedges [1992] is an accessible paper about the model.

Frequentist estimation of the step function publication bias model is problematic, as noted by for instance McShane et al. [2016, appendix, 1]. The purpose of this note is to formalize and prove exactly how frequentist estimation of this model can be. I do this by proving there are no confidence sets of guaranteed finite diameter for the mean parameter \( \theta \) for any coverage \( \tau^2 \) like \( [0.5, \infty) \), as they have no practical value. (2) It shows that the automatic confidence sets procedures that are guaranteed to yield finite confidence sets of some positive nominal coverage, such as bootstrapped confidence sets, likelihood-ratio based confidence sets, and subsampling confidence sets never have true coverage greater than 0 [see Gleser, 1996].

The main result of the paper is

**Theorem 1.** Let \( h(x) = \prod_{i=1}^{n} f(x_{i} \mid \theta_{0}, \tau, \sigma) \) be the density of \( n \) independent observations from a step function publication bias model. Then \( h \) has no almost surely finite diameter confidence set for \( \theta \) or \( \sigma \) of non-zero coverage.

### 2 Definitions and Proofs

Let \( P_{\theta}, \theta \in \Theta \) be a family of dominated probability measures with densities \( p_{\theta} \). Let \( g : \Theta \rightarrow E \) for some set \( E \) equipped with a positive function \( \| \cdot \| : E \rightarrow [0, \infty) \). Here \( g \) maps \( \theta \) to the parameter we wish to form a confidence set for. The function \( \| \cdot \| \) is a distance measure and could be a norm if \( E \) is a vector space. Below it is the absolute value \( |\cdot| \). A random set \( C \) is a confidence set for \( g(\theta) \) with coverage probability \( 1 - \alpha \) if \( P_{\theta} \left( g(\theta) \in C \right) \geq 1 - \alpha \) for all \( \theta \). An acceptance set with level \( \alpha \) has the property that \( P_{\theta} \left( A_{g(\theta)} \right) \geq 1 - \alpha \). There is a well known duality between confidence sets and acceptance sets: For each confidence set \( C \) there is a collection of acceptance sets \( A_{g(\theta)} \) satisfying \( \omega \in A_{g(\theta)} \iff g(\theta) \in C(\omega) \). The classical reference for these concepts is Lehmann and Romano [2006, chapter 3.5].

We will need terminology for the diameter of the confidence set according to the function \( \| \cdot \| \). The diameter is a random variable, \( D = \text{sup}_{\omega \in C} \| \eta \| \). A confidence set has infinite diameter at \( \theta \) with positive probability if \( P_{\theta} (D = \infty) > 0 \). If \( P_{\theta} (D = \infty) > 0 \) for all \( \theta \), we will say that \( D \) has infinite diameter with positive probability.

The main ingredient in the proof of theorem [1] is theorem [3], an extension and reformulation of the Gleser-Hwang theorem [Gleser and Hwang, 1987], a result they used to prove the non-existence of confidence sets with guaranteed finite diameter for models such as the errors-in-variables regression model and Fieller’s ratio of means problem. Berger et al. [1999, theorem 1] is a similar extension of the Gleser-Hwang theorem.

The problem with non-existence of confidence sets with guaranteed finite diameter for finite-dimensional parameters has not received much attention in the statistical literature, as lamented by Gleser [1996]. Some references in this area are Bahadur and Savage [1956], who studied infinitely large confidence sets in non-parametric estimation of the mean, Romano’s [2004] extension of their results, Donoho [1988] and Pfanzagl [1998].

**Lemma 2.** Let \( C \) be a confidence set and \( A_{g(\theta)} \) its associated acceptance sets. If there exists a sequence \( \{\theta_{n}\}_{n=1}^{\infty} \) with \( \| g(\theta_{n}) \| \rightarrow \infty \) such that \( \lim_{n \rightarrow \infty} P_{\theta} \left( A_{g(\theta_{n})} \right) > 0 \), then the diameter of \( C \) at \( \theta \) is infinite with positive probability.

**Proof.** Assume without loss of generality that \( \| g(\theta_{n}) \| \geq n \) for each \( n \); such a sequence can always be found by appropriately filtering the original \( \{\theta_{n}\}_{n=1}^{\infty} \). Since \( \{D \geq n\} \) is a decreasing sequence of sets, \( P_{\theta} (D = \infty) = \lim_{n \rightarrow \infty} P_{\theta} (D \geq n) \) for each \( \theta \). From \( A_{g(\theta_{n})} \subseteq \{D \geq n\} \) it follows that \( P_{\theta} \left( A_{g(\theta_{n})} \right) \leq P_{\theta} (D \geq n) \), and thus \( P_{\theta} (D = \infty) = \lim_{n \rightarrow \infty} P_{\theta} (D \geq n) \geq \lim_{n \rightarrow \infty} P_{\theta} (A_{g(\theta_{n})}) > 0 \).}

This is the extension of the Gleser-Hwang theorem.
Theorem 3. If there exists a sequence \( \{ \theta_i \}_{i \in \mathbb{N}} \) with \( \| g(\theta_i) \| \to \infty \) such that \( p_{\theta_i} \to f^* \) pointwise and \( \text{supp} p_{\theta_i} = \text{supp} f^* \) for each \( \theta \), then every confidence set with coverage \( 1 - \alpha > 0 \) has infinite diameter with positive probability.

Proof. Assume without loss of generality that \( \| g(\theta_n) \| \geq n \). By a variant of the dominated convergence theorem [Billingsley, 1995, exercise 16.4a, I], \( f_{\theta_n} \to f^* \) pointwise and \( \text{supp} f_{\theta_n} = \text{supp} f^* \) for each \( \theta \). Since \( A_{g(\theta_n)} \subseteq \{ D \leq n \}, 1 - \alpha \leq f_{A_{g(\theta_n)}} p_{\theta_n} d\mu \leq \int_{D \geq n} p_{\theta_n} d\mu \) for all \( n \), thus \( \int_{D = \infty} f^* d\mu > 1 - \alpha > 0 \). The result follows from \( P_{\theta}(D \geq n) = \int_{D \geq n} (p_{\theta}/p_{\theta_n}) p_{\theta_n} d\mu > \int_{D = \infty} (p_{\theta}/f^*) f^* d\mu > 0 \).

This an extension of theorem 5 to mixture distributions.

Proposition 4. Let \( f_{\theta_1}, f_{\theta_2}, \ldots \) be a finite or infinite sequence of densities, \( \pi(\theta) \) a countable probability vector and \( \sum_{n=1}^{\infty} \pi_n(\theta) f_{\theta_n} \) a mixture distribution. Assume there is a sequence \( \{ \theta_i \}_{i \in \mathbb{N}} \) with \( \| g(\theta_i) \| \to \infty \) such that \( f_{\theta_i} \to f^* \) pointwise and \( \text{supp} f_{\theta_i} = \text{supp} f^* \) for each \( \theta \). If there is a parameter \( \theta^* \in \Theta \) satisfying \( \pi_1(\theta^*) > \alpha \), the mixture \( \sum_{n=1}^{\infty} \pi_n(\theta) f_{\theta_n} \) admits no almost surely finite diameter confidence set of coverage \( 1 - \alpha \) for \( g(\theta) \).

Proof. Let \( C \) be confidence set for \( g(\theta) \) with coverage \( 1 - \alpha \). Since \( \pi_1(\theta^*) > \alpha \) by assumption, \( C \) must include a confidence \( C_1 \) for \( f_{\theta_1} \) of some positive coverage. But by theorem 5, \( C_1 \) has infinite diameter with positive probability for all \( \theta \). Since \( C \supseteq C_1 \), \( C \) has infinite diameter with positive probability too.

The following lemma is a proof of proposition 1 for the special case when no non-significant studies are published. Let \( \phi_{a,b}(x | \theta, \sigma) \) denote the density of a normal variable with mean \( \theta \) and standard deviation \( \sigma \) truncated to \( [a, b] \).

Lemma 5. Let \( f \) be a normal density truncated to \([c, \infty)\) with underlying mean \( \theta_n = -n - c \) and standard deviation \( \sigma_n = \sqrt{n} \). Then \( \phi_{c,\infty}(x | \theta_n, \sigma_n) \) converges pointwise to \( \exp(-x - c) \), the density of a shifted exponential.

Proof. The formula for \( \phi_{c,\infty}(x | \theta_n, \sigma_n) \) is \( \Phi \left( -\frac{n + \phi^{1/2}(1 - \alpha_j) \phi^{-1}(1 - \alpha_j - 1)}{\sqrt{n}} \right)^{-1} \phi(x; -n - c, \sqrt{n}) \). The normal density part equals
\[
(2\pi)^{-1/2} n^{-\frac{1}{2}} \exp \left( -\frac{n^2 - 2n (c + x) + [2cx - c^2 - x^2]}{2n} \right)
\]
When \( n \) is large compared to \( x \) and \( c \), \( [2cx - c^2 - x^2] / 2n \) is negligible, hence
\[
\phi(x; \theta_n, \sigma_n) \approx (2\pi)^{-1/2} n^{-\frac{1}{2}} \exp \left( -\frac{n}{2} \right) \exp(-x - c)
\]
which equals \( \exp(-x - c) \phi(n^{1/2}) / n^{1/2} \). Since \( \Phi(-n / \sqrt{n}) = \Phi(\sqrt{n}) = \sqrt{n} / \Phi(\sqrt{n}) \) as \( n \) grows (see e.g. Borjesson and Sundberg [1979, equation 5]), we end up with \( \Phi(-n / \sqrt{n}) \approx \phi(x; \theta_n, \sigma_n) \to \exp(-x - c) \).

This lemma gives a mixture representation of the publication bias model. The proof is omitted.

Lemma 6. The density of an observation from the step function publication bias model with parameters \( \theta_0, \tau, \sigma \) is
\[
f(x | \theta_0, \tau, \sigma) = \sum_{j=1}^{N} \pi_j \phi_{\theta_1, (1 - \alpha_j)} \phi_{\theta_j, (1 - \alpha_j - 1)} \left( x | \theta_0, (\tau^2 + \sigma_j^2)^{1/2} \right)
\]
where
\[
\pi_j(\theta_j, \tau, \sigma) = \frac{\rho_j \Phi(c_j - 1 | \theta_0, (\tau^2 + \sigma_j^2)^{1/2}) - \Phi(c_j | \theta_0, (\tau^2 + \sigma_j^2)^{1/2})}{\sum_{j=1}^{J} \rho_j \Phi(c_j - 1 | \theta_0, (\tau^2 + \sigma_j^2)^{1/2}) - \Phi(c_j | \theta_0, (\tau^2 + \sigma_j^2)^{1/2})}
\]

Proof of theorem 7. Using the representation 2.1 of lemma 6 and 4 it is enough to show the result for \( \phi_{\theta_1, (1 - \alpha_j)} \left( x | \theta_0, (\tau^2 + \sigma_j^2)^{1/2} \right) \) where \( \alpha_j + 1 = 1 \). To this end, let \( c = \Phi(1 - \alpha_j) \). Then \( \theta_n = -n - c \) and \( \tau_n^2 = n \) does the trick by lemma 5. For the case with more than one observation, observe that
\[
\prod_{i=1}^{n} \sum_{j=1}^{N} \pi_j \phi_{\theta_1, (1 - \alpha_j)} \left( x_i | \theta_0, (\tau^2 + \sigma_j^2)^{1/2} \right)
\]
is mixture model. (Just expand it.) The component belonging to the mixture probability \( \pi_j^n \) fulfills the demands of proposition 4 for \( \alpha_j \) when \( \alpha_j + 1 = 1 \), as its density is a product of densities converging to shifted exponentials.
Acknowledgements

I am grateful to Riccardo De Bin and Emil Aas Stoltenberg for reading through the manuscript and giving helpful comments.

References

Raghu R Bahadur and Leonard J Savage. The nonexistence of certain statistical procedures in nonparametric problems. *Annals of Mathematical Statistics*, 27(4):1115–1122, 1956.

James O Berger, Brunero Liseo, Robert L Wolpert, et al. Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*, 14(1):1–28, 1999.

Patrick Billingsley. *Probability and Measure*. John Wiley & Sons, 1995.

P Borjesson and C-E Sundberg. Simple approximations of the error function q(x) for communications applications. *IEEE Transactions on Communications*, 27(3):639–643, 1979.

Martyna Citkowicz and Jack L Vevea. A parsimonious weight function for modeling publication bias. *Psychological Methods*, 22(1):28, 2017.

David L Donoho. One-Sided inference about functionals of a density. *Annals of Statistics*, 16(4):1390–1420, December 1988. ISSN 0090-5364, 2168-8966. doi: 10.1214/aos/1176351045. URL https://projecteuclid.org/euclid.aos/1176351045.

Bernard D Flury. Acceptance–rejection sampling made easy. *SIAM Review*, 32(3):474–476, 1990.

Leon Jay Gleser. Bootstrap confidence intervals: Comment. *Statistical Science*, pages 219–222, 1996.

Leon Jay Gleser and Jiunn T Hwang. The nonexistence of 100 (1-\(\alpha\))% confidence sets of finite expected diameter in errors-in-variables and related models. *Annals of Statistics*, pages 1351–1362, 1987.

Larry V Hedges. Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational Statistics*, 9(1):61–85, 1984.

Larry V Hedges. Modeling publication selection effects in meta-analysis. *Statistical Science*, pages 246–255, 1992.

Larry V Hedges and Jack L Vevea. Fixed-and random-effects models in meta-analysis. *Psychological Methods*, 3(4):486, 1998.

Satish Iyengar and Joel B Greenhouse. Selection models and the file drawer problem. *Statistical Science*, pages 109–117, 1988.

Erich L Lehmann and Joseph P Romano. *Testing statistical hypotheses*. Springer Science & Business Media, 2006.

Mark W Lipsey and David B Wilson. *Practical Meta-analysis*. Sage Publications, Inc, 2001.

Blakeley B McShane, Ulf Böckenholt, and Karsten T Hansen. Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science*, 11(5):730–749, 2016.

Jonas Moss and Riccardo De Bin. Modelling publication bias and p-hacking. *arXiv preprint arXiv:1911.12445*, 2019.

J Pfanzagl. The nonexistence of confidence sets for discontinuous functionals. *Journal of Statistical Planning and Inference*, 75(1):9–20, November 1998. ISSN 0378-3758. doi: 10.1016/S0378-3758(98)00121-9. URL http://www.sciencedirect.com/science/article/pii/S0378375898001219.

Joseph P Romano. On non-parametric testing, the uniform behaviour of the t-test, and related problems. *Scandinavian Journal of Statistics*, 31(4):567–584, 2004.

Hannah R Rothstein, Alexander J Sutton, and Michael Borenstein. *Publication bias in meta-analysis: Prevention, assessment and adjustments*. John Wiley & Sons, 2006.

Joseph P Simmons, Leif D Nelson, and Uri Simonsohn. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11):1359–1366, 2011.

Theodore D Sterling. Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American statistical association*, 54(285):30–34, 1959.

John von Neumann. Various techniques used in connection with random digits. *Applied Math Series*, 12(36-38):1, 1951.