On Identifiability in Capture-Recapture Models: Supporting Material–Proofs

Hajo Holzmann[1], **Axel Munk**[1,*], **and Walter Zucchini**[2]

[1] Institut für Mathematische Stochastik, Georg-August-University Göttingen, Maschmühlenweg 8-10, D–37073 Göttingen, Germany

[2] Institut für Statistik und Ökonometrie, Platz der Göttinger Sieben 5, D–37073 Göttingen, Germany

[*]*email:* munk@math.uni-goettingen.de

Abstract

In this note we provide the proofs for our paper "On identifiability in capture-recapture models".

Let us briefly recall the notation introduced in Holzmann, Munk and Zucchini (2005). We consider a mixture of the binomial $B(T, p)$-distribution with mixing distribution G,

$$\pi_G(x) = \binom{T}{x} \int_0^1 p^x (1-p)^{T-x} \, dG(p), \tag{1}$$

as well as the conditional distribution

$$\pi_G^c = \big(\pi_G^c(1), \dots \pi_G^c(T)\big), \quad \pi_G^c(x) = \frac{\pi_G(x)}{1-\pi_G(0)}, \quad x = 1, \dots, T.$$

Definition 1. We shall call a family $\mathcal{G}$ of distributions on $[0, 1]$ *identifiable* if, for each $G \in \mathcal{G}$, the vector π_G^c uniquely determines G within the class $\mathcal{G}$, i.e. if for $G, H \in \mathcal{G}$,

$$\pi_G^c = \pi_H^c \Rightarrow G = H. \tag{2}$$

Lemma 1. *Let $\pi = (\pi(0), \dots, \pi(T))$ and $\rho = (\rho(0), \dots \rho(T))$ be two probability vectors on $\{0, \dots, T\}$, and let π^c and ρ^c be the conditional probability vectors on $1, \dots, T$, given that $x \geq 1$. Then*

$$\pi^c = \rho^c \Leftrightarrow \text{there is an } \; A > 0 \; : \; \pi(x) = A\rho(x), \quad x = 1, \dots, T.$$

Proof. For $x = 1, \dots, T$,

$$\frac{\pi(x)}{1-\pi(0)} = \frac{\rho(x)}{1-\rho(0)} \Leftrightarrow \frac{\pi(x)}{\rho(x)} = \frac{1-\pi(0)}{1-\rho(0)} =: A.$$

□

We have that

$$\pi_G(x) = \binom{T}{x} \sum_{k=x}^{T} c_{k,x} \, m_G(k), \quad x = 1, \dots, T, \tag{3}$$

where $m_G(k) = \int_0^1 t^k \, dG(t)$ is the kth moment of G and

$$c_{x,k} = (-1)^{k-x} \binom{T-x}{k-x}, \quad \text{if } k \geq x, \qquad c_{k,x} = 0 \quad \text{if } k < x.$$

For our problem this implies

Theorem 1. *For two distributions G, H on $(0,1]$, $\pi_G^c = \pi_H^c$ implies that there is an $A > 0$ such that*

$$m_G(x) = A \, m_H(x), \quad x = 1, \ldots, T. \tag{4}$$

Proof. In matrix form the identity (3) can be written as

$$(\pi_G(1), \ldots \pi_G(T))' = C(m_G(1), \ldots m_G(T))',$$

where $C = (c_{x,k})_{k,x=1,\ldots,T}$, and $'$ denotes a column vector. The matrix C is invertible, because it is upper triangular with nowhere vanishing diagonal. Therefore

$$\pi_G(x) = A\pi_H(x), \quad x = 1, \ldots, T \Leftrightarrow m_G(x) = Am_H(x), \quad x = 1, \ldots, T.$$

□

Therefore, if there exist no two different $G, H \in \mathcal{G}$ such that (4) holds, then $\mathcal{G}$ is identifiable. Now consider the class of finite mixing distributions with at most m support points

$$\mathcal{G}_m = \Big\{ G = \sum_{k=1}^{m} \lambda_k \delta_{p_k}, \quad \lambda_k \geq 0, \quad \sum_k \lambda_k = 1, \quad p_k \in (0,1] \Big\}.$$

Theorem 2. *For $2m \leq T$ the class $\mathcal{G}_m$ is identifiable.*

First we prove the following lemma.

Lemma 2. *If*

$$\sum_{k=1}^{T} t_k \, p_k^x (1-p_k)^{T-x} = 0, \quad x = 1, \ldots, T,$$

for some $t_k \in \mathbb{R}$ and distinct $p_k \in (0,1]$, then it follows that $t_1 = \ldots = t_T = 0$.

Proof. The polynomials $P_x(p) = p^x(1-p)^{T-x}$, $x = 1, \ldots, T$, are linearly independent because, except for the normalization, these are the Bernstein polynomials, which are known to be linearly independent, cf. Prautzsch et al. (2002). Therefore any nontrivial linear combination, which thus is a nonzero polynomial of degree at most T, has at most T roots. Since, evidently, one of these always equals 0, there are at most $T-1$ roots within the interval $(0,1]$. Hence, for different $p_1, \ldots, p_T \in (0,1]$, if

$$\sum_{x=1}^{T} s_x p_k^x (1-p_k)^{T-x} = 0, \quad k = 1, \ldots, T,$$

it follows that the coefficients $s_1 = \ldots = s_T = 0$, all vanish. Introducing matrix notation $P = (P_{k,x}) = (p_k^x(1-p_k)^{T-x})_{k,x=1,\ldots,T}$ and $s = (s_1, \ldots, s_T)'$, this is just

$$P \cdot s = 0 \Rightarrow s = 0. \tag{5}$$

Relation (5) implies that P has full rank, hence so does its transpose P', and we get

$$P' \cdot t = 0 \Rightarrow t = 0 \quad \text{for } t \in \mathbb{R}^T, \tag{6}$$

which is the claim of the lemma. $\square$

The above lemma can also be established by arguments from the theory of Čebyšev systems (cf. Karlin and Studden, 1966). In fact, since $p^{x-1}(1-p)^{T-x}$, $x = 1, \ldots, T$, is a Čebyšev system on $\mathbb{R}$, and since the function p does not vanish in $(0, 1]$, it follows that $p^x(1-p)^{T-x}$, $x = 1, \ldots, T$, is a Čebyšev system on $(0, 1]$, which is the statement of the lemma.

Proof of Theorem 2. Suppose that $G, H \in \mathcal{G}_m$ with $\pi_G^c = \pi_H^c$. From Lemma 1, there exists an $A > 0$ with

$$\sum_{k=1}^m \lambda_{k,G}\, p_{k,G}^x \,(1 - p_{k,G})^{T-x} = A \sum_{k=1}^m \lambda_{k,H}\, p_{k,H}^x \,(1 - p_{k,H})^{T-x}, \quad x = 1, \ldots, T, \tag{7}$$

where $\lambda_{k,G}, \lambda_{k,H} \geq 0$ and

$$\sum_k \lambda_{k,G} = \sum_k \lambda_{k,H} = 1. \tag{8}$$

Subtracting the r.h.s. from the l.h.s. in (7), we get relations of the form

$$\sum_{j=1}^J \lambda_j p_j^x (1 - p_j)^{T-x} = 0, \qquad x = 1, \ldots, T,$$

where $p_1, \ldots, p_J$ are the distinct points in $\{p_{1,G}, \ldots, p_{m,G}, p_{1,H}, \ldots, p_{m,H}\}$, and the coefficients λ_j are given by one out of

$$\lambda_{k,G}, \quad -A\lambda_{k',H}, \quad \lambda_{k,G} - A\lambda_{k',H},$$

depending on whether p_j is only equal to one of the $p_{k,G}$'s, only one of the $p_{k',H}$'s or equal to one from each group.
Since $J \leq 2m \leq T$, we can apply Lemma 2 to conclude that $\lambda_j = 0$, $j = 1 \ldots, J$. Now if $\lambda_j = \lambda_{k,G}$ for some k, then this $\lambda_{k,G} = 0$, and we simple drop the corresponding $p_{k,G}$, and similarly if $\lambda_j = -\lambda_{k',H}$. Therefore, all points of support of H and G (i.e. where the corresponding weight is strictly positive) coincide. After a suitable reordering we have that

$$\lambda_{k,G} - A\lambda_{k,H} = 0.$$

Summing this relation over k and using (8) gives $A = 1$ and $\lambda_{k,G} = \lambda_{k,H}$, i.e. $H = G$, as required. $\square$

References

Holzmann, H., Munk, A. and Zucchini, W. (2005). On identifiability in capture-recapture models. *Biometrics* **62**, 934–936.

Karlin, S. and Studden, W. J. (1966). *Tchebycheff systems: With Applications in Analysis and Statistics.* New York: John Wiley & Sons.

Prautzsch, H., Boehm, W. and Paluszny, M. (2002). *Bézier and B-spline techniques.* New York: Springer.