Biometrical Journal **xx** (2015) xx, zzz–zzz / DOI: 10.1002/bimj.200100000

On the design of closed recapture experiments

Danilo Alunni Fegatelli[*,1] and **Alessio Farcomeni** [**,1]

[1] Department of Public Health and Infectious Diseases *(Sapienza - University of Rome)*

Received zzz, revised zzz, accepted zzz

We propose a method, based on the expected length of the profile likelihood confidence interval, to plan the number of occasions of recapture experiments for population size estimation. The general method is detailed in closed form for models assuming homogeneous, time-varying, subject-specific capture probabilities. Additionally, we outline models with behavioural response to capture and combining behavioural response with subject-specific effects. The principle we propose can be applied numerically to plan any other model specification. We formally show the validity of the approach by proving distributional convergence. We illustrate with simulations and examples. We report that in many cases adding as few as two sampling occasions may substantially reduce the length of confidence intervals around population size estimates.

Key words: Profile confidence interval; planning of experiments; population size estimation.

1 Distributional Convergence

Let $l(N;T)$, as in the main paper, denote the log profile likelihood for any model, as a function of N and a possibly multidimensional sufficient statistic T. In the following we treat N as a continuous parameter. We formally show in this supporting information the statements at the end of Section 2 of the manuscript.

Theorem 1.1 *Assuming the log-profile likelihood is continuous and differentiable at* $E[T]$*,* $\Pr(\sum_j X_{ij} > 0) > 0 \; \forall \; i$*,* $N > 0$ *and* p *is in the intern of the parameter space,*

$$\sqrt{N}(l(N;T)/\sigma_l - \mu_l) \xrightarrow{d} N(0,1),$$

where $\xrightarrow{d}$ *denotes convergence in distribution,*

$$\mu_l = l(N;E[T]) \tag{1}$$

and

$$\sigma_l = \sqrt{\nabla l(N;E[T])' \Sigma_T \nabla l(N;E[T])}, \tag{2}$$

where $\nabla l(N;\cdot)$ *denotes the gradient of the profile likelihood and* Σ_T *the asymptotic covariance matrix of the sufficient statistics* T*.*

P r o o f. The profile log-likelihood can in general be expressed as

$$l(N;T) \propto l(N;\mathbf{X}) = \sup_p \sum_{i=1}^{N} \log(\Pr(X_i|p)). \tag{3}$$

*e-mail: danilo.alunnifegatelli@uniroma1.it

**Corresponding author: e-mail: alessio.farcomeni@uniroma1.it

Given that subjects are independent and identically distributed, the central limit theorem implies that for any p

$$\frac{\sum_{i=1}^{N} \log(\Pr(X_i|p))}{\sqrt{N}}$$

converges in distribution to a normal random variable. It is straightforward to check that the $\sup$ operator in (3) leads to a continuous mapping, therefore we can use Slutsky's theorem to conclude that $l(N;T)$ converges in distribution to a normal random variable. It is now a matter of computing the asymptotic mean and variance. A direct application of the Delta method yields (1) and (2). □

Corollary 1.2 *Under the same hypotheses as Theorem 1.1 we have that*

$$\sqrt{N}(L_{CI}(N;T)/\sigma_L - \mu_L) \xrightarrow{d} N(0,1),$$

where $\mu_L = L_{CI}(N; E[T])$ *and* $\sigma_L = \sqrt{\nabla L_{CI}(N; E[T])' \Sigma_T \nabla L_{CI}(N; E[T])}$.

Proof. Distributional convergence and expression of μ_L are a direct consequence of Theorem 1.1 and Slutzky theorems. In order to obtain the expression of σ_L we use the Delta method. □