S4 Appendix

A Identifiability

The probability of observing an infection with configuration $i$ is

$$Q_i = \frac{1}{e^\lambda - 1} \prod_{j=1}^n (e^{\lambda p_j} - 1)^{i_j}.$$  \hspace{1cm} (S1)

**Proof of Remark 1.** Let $Q_i^{(1)}$ and $Q_i^{(2)}$ denote the distributions (S1) for two parameter vectors $\theta_1, \theta_2 \in \Theta$. Suppose $Q_i^{(1)} = Q_i^{(2)}$ for all $i$.

First, assume $\theta_1 = (\lambda, p_1)$ and $\theta_2 = (\lambda, p_2)$ with $p_1 \neq p_2$. Without loss of generality let $p_1^{(1)} \neq p_2^{(2)}$, i.e., the first components of $p_1$ and $p_2$ differ. Then clearly $Q_{e_1}^{(1)} \neq Q_{e_1}^{(2)}$, where $e_1 = (1, 0, \ldots, 0)$, a contradiction.

Now, suppose $\theta_1 = (\lambda_1, p_1)$ and $\theta_2 = (\lambda_2, p_2)$ with $\lambda_1 \neq \lambda_2$ and possibly $p_1 \neq p_2$. Without loss of generality let $\lambda_1 < \lambda_2$, $p_1^{(1)} \geq p_2^{(1)}$ and $p_2^{(2)} \leq p_2^{(1)}$. This implies $e^{\lambda_1 p_2^{(2)}} - 1 < e^{\lambda_2 p_2^{(2)}} - 1$. Because $Q_{e_1}^{(1)} = Q_{e_1}^{(2)}$ it follows that $Q_{e_1 + e_2}^{(1)} < Q_{e_1 + e_2}^{(2)}$ ($e_2 = (0, 1, 0, \ldots, 0)$), a contradiction. Hence, $Q_i^{(1)} = Q_i^{(2)}$ for all $i$ only if $\theta_1 = \theta_2$. \hfill $\Box$

B Constraining the data and the parameter space

Here, we derive the lineages’ prevalences and the probability of observing irregular data, i.e., $N_k = N$ for at least one $k$ or $\sum_{k=1}^n N_k = N$. Remember that a blood sample is represented by a 0-1 vector $i = (i_1, \ldots, i_n)$. The probability of observing $i$ is denoted by $Q_i$ and given by (S1).

First, we derive the prevalence of lineage $j$, i.e., the probability to observe lineage $j$ in a blood sample. Note that by using (S1)

$$1 = \sum_{i \in \{0,1\}^n \setminus \{0\}} Q_i = \sum_{i \in \{0,1\}^n \setminus \{0\} : i_n = 1} Q_i + \sum_{i \in \{0,1\}^n \setminus \{0\} : i_n = 0} Q_i$$

$$= Q_{e_n} + (e^{\lambda p_n} - 1) \sum_{i \in \{0,1\}^n \setminus \{0\} : i_n = 0} Q_i$$

$$= e^{\lambda p_n} - 1 \frac{e^{\lambda p_n} - 1}{e^\lambda - 1} + e^{\lambda p_n} \sum_{i \in \{0,1\}^n \setminus \{0\} : i_n = 0} Q_i$$

$$= e^{\lambda p_n} - 1 \frac{e^{\lambda p_n} - 1}{e^\lambda - 1} + e^{\lambda p_n} \left(1 - \sum_{i_n = 1} Q_i\right).$$
Hence,

\[ Q_i = \frac{e^\lambda (e^{\lambda p_i} - 1)}{e^{\lambda p_i} (e^\lambda - 1)}, \]

and by replacing \( n \) with \( j \) we obtain the probability that lineage \( j \) is observed in a blood sample as

\[ q(j) := q_j = \sum_{i \in \{0,1\}^n \setminus \{0\}: i_n = 1} Q_i = \frac{e^\lambda (e^{\lambda p_j} - 1)}{e^{\lambda p_j} (e^\lambda - 1)}. \]  

(S2)

Note, that \( p_j \) is the frequency of lineage \( j \) in the population of infective agents and differs from the prevalence \( q_j \). However, in the limit \( \lambda \to 0 \), these probabilities coincide as is easily seen by applying de l’Hospitals rule. This is not surprising, because in this limit every host is infected by exactly one lineage.

Next, we show by induction that the probability of observing lineages \( j_1, \ldots, j_k \) together in a blood sample is given by

\[ q\{j_1, \ldots, j_k\} := \sum_{i \in \{0,1\}^n \setminus \{0\}: \sum_{i_j = 1}^k i_j = 1} Q_i = \frac{e^\lambda}{e^\lambda - 1} \prod_{s=1}^k (1 - e^{-\lambda p_{j_s}}). \]  

(S3)

This probability corresponds to the prevalence of the combination of lineages \( j_1, \ldots, j_k \).

If \( k = 1 \) the formulas (S3) and (S2) obviously coincide. By relabelling it suffices to show that (S3) holds for lineages \( n-k, \ldots, n \). Assume \( q\{n-k+1, \ldots, n\} \) is given by (S3).

Hence,

\[ q\{n-k+1, \ldots, n\} = \sum_{i \in \{0,1\}^n \setminus \{0\}: i_{n-k+1} = \ldots = i_n = 1} Q_i = \sum_{i \in \{0,1\}^{n-k-1}: i_{n-k} = 1} Q\{i,1,\ldots,1\} + \sum_{i \in \{0,1\}^{n-k}: i_{n-k} = 0} Q\{i,1,\ldots,1\} \]

\[ = (e^{\lambda p_{n-k}} - 1) \sum_{i \in \{0,1\}^{n-k-1}: i_{n-k} = 1} Q\{i,1,\ldots,1\} + \sum_{i \in \{0,1\}^{n-k}: i_{n-k} = 0} Q\{i,1,\ldots,1\} \]

\[ = e^{\lambda p_{n-k}} \sum_{i \in \{0,1\}^{n-k}: i_{n-k} = 0} Q\{i,1,\ldots,1\} \]

\[ = e^{\lambda p_{n-k}} \left( q\{n-k+1, \ldots, n\} - \sum_{i \in \{0,1\}^{n-k}: i_{n-k} = 1} Q\{i,1,\ldots,1\} \right) \]

\[ = e^{\lambda p_{n-k}} \left( q\{n-k+1, \ldots, n\} - q\{n-k, \ldots, n\} \right). \]

Hence,

\[ q\{n-k, \ldots, n\} = \frac{e^\lambda}{e^\lambda - 1} \prod_{j=n-k}^n (1 - e^{-\lambda p_j}) \]

follows, which proves (S3).
Now, the probability that $N_k = N$ for at least on $k$ is found by a standard inclusion exclusion argument and is given by

$$
\sum_{A \subseteq \{1, \ldots, n\}} (-1)^{|A|+1} \left( \sum_{i \in \{0,1\}^n \setminus \{0\}: \sum_{k \in A} Q_i} \right)^N = \sum_{A \subseteq \{1, \ldots, n\}, A \neq \emptyset} (-1)^{|A|+1} q^N_A
$$

$$
= \frac{1}{(1 - e^{-\lambda})^N} \sum_{A \subseteq \{1, \ldots, n\}} (-1)^{|A|+1} \prod_{j \in A} (1 - e^{-\lambda p_j})^N
$$

$$
= \frac{1}{(1 - e^{-\lambda})^N} \left( 1 - \prod_{j=1}^n (1 - (1 - e^{-\lambda p_j})^N) \right).
$$

(S4)

Next, we will derive the probability of observing only single infections. A single infection with lineage $k$ is represented by the standard base vectors $\mathbf{e}_k$. The probability of a single infection is therefore $Q_{\mathbf{e}_k} + \ldots + Q_{\mathbf{e}_n}$. Hence, the probability of observing only single infections is given by $\left( \sum_{k=1}^n Q_{\mathbf{e}_k} \right)^N$.

Summarizing, the probability that $N_k = N$ for at least one $k$ or $\sum_{k=1}^n N_k = N$, is given by

$$
q := \frac{1}{(1 - e^{-\lambda})^N} \left( 1 - \prod_{j=1}^n (1 - (1 - e^{-\lambda p_j})^N) \right) + \left( \sum_{j=1}^n Q_{\mathbf{e}_j} \right)^N - \sum_{j=1}^n Q_{\mathbf{e}_j}^N,
$$

where the probabilities that all samples are only infected by lineage $j$, i.e., $Q_{\mathbf{e}_j}^N$, are subtracted because they are accounted for in both irregular cases, i.e., $N$ single infections with the same lineage ($N_j = N$ for exactly one $j$ implying $\sum_{k=1}^n N_k = N$).

Clearly, the probability $q$ vanishes as $N \to \infty$. However, if $N$ and $\lambda$ are small and the lineage frequencies are very skewed, it might be rather larger.

Next we present the proof of Result [1].

Proof of Result [1]. For regular data a unique MLE $(\hat{\lambda}, \hat{\mathbf{p}})$ exists and $\lambda = 0$ is neither the MLE nor the true parameter (cf. [1]).

Because $\hat{\Theta}$ is compact, $L(\lambda, \mathbf{p})$ attains a global maximum at $(\hat{\lambda}, \hat{\mathbf{p}}) \in \hat{\Theta}$. If $\hat{\lambda} \notin [\lambda_{\text{min}}, \lambda_{\text{max}}]$, then $(\hat{\lambda}, \hat{\mathbf{p}}) \notin \hat{\Theta}$. Hence, $(\hat{\lambda}, \hat{\mathbf{p}}) \neq (\hat{\lambda}, \hat{\mathbf{p}})$. Therefore, $\hat{\lambda} = \lambda_{\text{min}}$ or $\hat{\lambda} = \lambda_{\text{max}}$ must hold.

Let $c = L(\hat{\lambda}, \hat{\mathbf{p}}) - L(\hat{\lambda}, \hat{\mathbf{p}}) > 0$. The parameter vector $(\hat{\lambda}, \hat{\mathbf{p}})$ maximizes $L(\lambda, \mathbf{p})$ subject to the equality constraint $L(\hat{\lambda}, \hat{\mathbf{p}}) = L(\lambda, \mathbf{p}) - c = 0$. This corresponds exactly to finding the profile-likelihood confidence interval $(\hat{\lambda}, \hat{\mathbf{p}})$ for $\lambda$ at the $(1 - \alpha)$-level, where $\alpha$ is chosen such that $2c$ is the $(1 - \alpha)$-quantile of the $\chi^2$-distribution with 1 degree of freedom.

As shown in [1], unique confidence points $\hat{\lambda} < \hat{\lambda} < \bar{\lambda}$ exists, implying $\hat{\lambda} = \frac{\lambda}{\bar{\lambda}}$ or $\hat{\lambda} = \frac{\lambda}{\bar{\lambda}}$.

Assume $\lambda < \lambda_{\text{min}}$. Then $\hat{\lambda} = \lambda_{\text{max}}$ is impossible, because the uniqueness of $\bar{\lambda} > \hat{\lambda}$ maximizing the profile-likelihood subject to the constraint $L(\lambda, \mathbf{p}) - L(\lambda, \mathbf{p}) = c$ would imply $L(\hat{\lambda}, \hat{\mathbf{p}}) = \max_p L(\lambda_{\text{min}}, \mathbf{p}) < c$. The latter is a contradiction to $(\hat{\lambda}, \hat{\mathbf{p}})$ being the
point, where the log-likelihood function attains its global maximum on $\Theta$. Thus, $\hat{\lambda} = \lambda_{\min}$. The same reasoning yields $\hat{\lambda} = \lambda_{\max}$ if $\hat{\lambda} > \lambda_{\max}$.

In the case that only one lineage is found in the data, i.e., $\sum_{k=1}^{n} N_k = N$ and $N_j = N$ for some $j$, it is immediately clear from the log-likelihood function that any estimate of the form $(\lambda, e_j)$ is equally likely. Hence, we further assume that at least two lineages are present in the data.

If $\sum_{k=1}^{n} N_k = N$, $\hat{\lambda} = 0$ is the MLE. The argument with the profile-likelihood still applies (however only the upper confidence point $\lambda$ exists), yielding $\hat{\lambda} = \lambda_{\min}$.

Next, we consider $\lambda$ as a fixed constant and maximize the log-likelihood function (eq. 3 in the main text) over the simplex. It is more convenient to introduce a nuisance parameter $\beta$ and maximize the function

$$\Lambda(\mathbf{p}, \beta) = L(\lambda, \mathbf{p}) - \beta \left(1 - \sum_{k=1}^{n} p_k\right),$$

over $\mathbb{R}^{n+1}$ because

$$\frac{\partial \Lambda}{\partial \beta} = 0 \iff \sum_{i=1}^{n} p_i = 1.$$

The equations $\frac{\partial \Lambda}{\partial p_k} = 0$ yields $N_k \frac{\lambda e^{\lambda p_k}}{e^{\lambda p_k} - 1} = \beta$ for all $k$ (cf. eq. 817b in Appendix D below). Solving this equation with respect to $p_k$ yields

$$p_k = -\frac{1}{\lambda} \log \left(1 - \frac{\lambda N_k}{\beta}\right),$$

which is in essence eq. 6 in the main text. Substituting this into $\sum_{k=1}^{n} p_k = 1$, yields after a little rearrangement,

$$0 = \lambda + \sum_{k=1}^{n} \log \left(1 - \frac{\lambda N_k}{\beta}\right) =: f(\beta).$$

Clearly, $f(\beta)$ is defined only for $\beta > \max_k \lambda N_k =: \beta_0$, $\lim_{\beta \to \beta_0} f(\beta) = -\infty$ and $\lim_{\beta \to \infty} f(\beta) = \lambda$. Taking the derivative yields

$$f'(\beta) = \frac{1}{\beta} \sum_{k=1}^{n} \frac{\lambda N_k}{\beta - \lambda N_k} > 0.$$

Thus, $f$ is strictly monotonically increasing and consequently $f(\hat{\beta}) = 0$ for exactly one $\hat{\beta} > \beta_0$. This solution is found by applying Newton’s method, which yields exactly eq. 6b in the main text for $\lambda = \lambda_{\min}$. Choosing $\lambda = \lambda_{\max}$ yields the rest of the proof. □

C Fisher information matrix

For the derivations here it is convenient to eliminate one redundant parameter by substituting $p_n = 1 - \sum_{k=1}^{n-1} p_k$. The first order derivatives of the log-likelihood function
where $\delta$ is uniformly bounded on any compact subset $\Theta$. Therefore, by the theorem of maximum and minimum we obtain: 

$$\frac{\partial L}{\partial \lambda} = -N \frac{e^\lambda}{e^\lambda - 1} + \sum_{k=1}^{n} N_k \frac{p_k e^{\lambda p_k}}{e^{\lambda p_k} - 1}, \quad (S5a)$$

$$\frac{\partial L}{\partial p_k} = N_k \frac{\lambda e^{\lambda p_k}}{e^{\lambda p_k} - 1} - N_n e^{\lambda p_n} \frac{e^{\lambda p_n}}{e^{\lambda p_n} - 1}, \quad (S5b)$$

and the second order derivatives are

$$\frac{\partial^2 L}{\partial \lambda^2} = N \frac{e^\lambda}{(e^\lambda - 1)^2} - \sum_{k=1}^{n} N_k \frac{p_k^2 e^{\lambda p_k}}{(e^{\lambda p_k} - 1)^2}, \quad (S6a)$$

$$\frac{\partial^2 L}{\partial \lambda \partial p_k} = \frac{\partial^2 L}{\partial p_k \partial \lambda} = N_k \frac{e^{\lambda p_k}}{e^{\lambda p_k} - 1} \left(1 - \frac{\lambda p_k}{e^{\lambda p_k} - 1}\right) - N_n \frac{e^{\lambda p_n}}{e^{\lambda p_n} - 1} \left(1 - \frac{\lambda p_n}{e^{\lambda p_n} - 1}\right), \quad (S6b)$$

$$\frac{\partial^2 L}{\partial p_k^2} = -N_k \frac{\lambda^2 e^{\lambda p_k}}{(e^{\lambda p_k} - 1)^2} - N_n \frac{\lambda^2 e^{\lambda p_n}}{(e^{\lambda p_n} - 1)^2}, \quad (S6c)$$

$$\frac{\partial^2 L}{\partial p_k \partial p_j} = -N_n \frac{\lambda^2 e^{\lambda p_n}}{(e^{\lambda p_n} - 1)^2} \text{ for } k \neq j. \quad (S6d)$$

Clearly, the third-order derivatives exist for any admissible parameter value $\theta \in \tilde{\Theta}$ and could be calculated analogously. Obviously, these are linear in $N$ and the $N_k$’s. Therefore, by the theorem of maximum and minimum we obtain:

**Remark 1** The third-order derivatives of the log-likelihood function satisfy that $\frac{1}{N} \frac{\partial^3 L}{\partial \theta^3}$ is uniformly bounded on any compact subset $\tilde{\Theta} \subset \tilde{\Theta}$.

Notably, $\tilde{\Theta}$ is not compact and the third-order derivatives are indeed not bounded on $\tilde{\Theta}$.

In order to derive the information matrix and prove its properties we need the following lemma.

**Lemma 1** The following identities hold for $k, \ell = 1, \ldots, n$:

$$\sum_{i \in \{0,1\}^n \setminus \{0\}} i_k Q_i = \sum_{i \in \{0,1\}^n \setminus \{0\}} i_k^2 Q_i = \frac{1 - e^{-\lambda p_k}}{1 - e^{-\lambda}}, \quad (S7a)$$

$$\sum_{i \in \{0,1\}^n \setminus \{0\}} i_k i_\ell Q_i = \frac{(1 - e^{-\lambda p_k})(1 - e^{-\lambda p_\ell})}{1 - e^{-\lambda}} \quad \text{for } k \neq \ell, \quad (S7b)$$

$$\mathbb{E} N_k = N \frac{1 - e^{-\lambda p_k}}{1 - e^{-\lambda}}, \quad (S7c)$$

$$\mathbb{E} N_k N_\ell = \frac{(1 - e^{-\lambda p_k})(1 - e^{-\lambda p_\ell})}{1 - e^{-\lambda}} \left(1 + \frac{N^2 - N}{1 + N} + \delta_{k\ell} N \frac{e^{-\lambda p_k}(1 - e^{-\lambda p_\ell})}{1 - e^{-\lambda}}\right). \quad (S7d)$$

where $\delta_{k\ell}$ denotes the Kronecker delta.

**Proof.** Clearly, (S7a) is equivalent to (S2). The second identity is a special case of (S3).

Denoting a 0-1 vector by $i = (i_1, \ldots, i_n)$ and by $n_i$ the number of blood samples corresponding to $i$ we clearly have $N_k = \sum_{i \in \{0,1\}^n \setminus \{0\}} i_k n_i$. Moreover, the $n_i$ are
multinomially distributed with parameters $N$ and $Q_i$, where $i \in \{0, 1\}^n \setminus \{0\}$.

Therefore, $E_n = NQ_i$ and hence

$$E_N = E\left(\sum_{i \in \{0, 1\}^n \setminus \{0\}} i_k n_i\right) = \sum_{i \in \{0, 1\}^n \setminus \{0\}} i_k E_n = N \sum_{i \in \{0, 1\}^n \setminus \{0\}} i_k Q_i = N\frac{1 - e^{-\lambda p_k}}{1 - e^{-\lambda}}$$

according to the first identity.

From the multinomial distribution of the $n_i$ we obtain $E_{n_i n_j} = (N^2 - N)Q_i Q_j$ for $i \neq j$ and $E_{n_i^2} = (N^2 - N)Q_i^2 + NQ_i$. Next, we aim to derive $E_{N_k N_{\ell}}$. Hence,

$$E_{N_k N_{\ell}} = E\left(\sum_{i,j \in \{0, 1\}^n \setminus \{0\}} i_k j_{n_i n_j}\right) = \sum_{i,j \in \{0, 1\}^n \setminus \{0\}} i_k j_{n_i n_j}$$

$$= (N^2 - N) \sum_{i,j \in \{0, 1\}^n \setminus \{0\}} i_k j_{n_i Q_j} + N \sum_{i \in \{0, 1\}^n \setminus \{0\}} i_k i_{n_i Q_i}$$

$$= (N^2 - N) \left(\sum_{i \in \{0, 1\}^n \setminus \{0\}} i_k Q_i\right) \left(\sum_{j \in \{0, 1\}^n \setminus \{0\}} j_{Q_j}\right) + N \sum_{i \in \{0, 1\}^n \setminus \{0\}} i_k i_{n_i Q_i}.$$

Using (S7a)-(S7c) we obtain

$$E_{N_k N_{\ell}} = \frac{(1 - e^{-\lambda p_k})(1 - e^{-\lambda p_{\ell}})}{1 - e^{-\lambda}} \left(\frac{N^2 - N}{1 - e^{-\lambda}} + N\right)$$

for $k \neq \ell$ and

$$E_{N_k^2} = \frac{1 - e^{-\lambda p_k}}{1 - e^{-\lambda}} \left(\frac{N^2 - N}{1 - e^{-\lambda}} + N\right)$$

$$= \frac{(1 - e^{-\lambda p_k})^2}{1 - e^{-\lambda}} \left(\frac{N^2 - N}{1 - e^{-\lambda}} + N\right) + N e^{-\lambda p_k} \left(\frac{1 - e^{-\lambda p_k}}{1 - e^{-\lambda}} - 1\right)$$

for $k = \ell$, finishing the proof.

Lemma 1 enables us to derive the information matrix immediately from (S6), which is presented in Result 2 in the Main Text. Moreover, we are now able to prove an important regulatory condition.

**Theorem 1** The Fisher information matrix satisfies

$$I = -E \frac{\partial^2 L}{\partial \theta^2} = E\left(\left(\frac{\partial L}{\partial \theta}\right)^T \cdot \left(\frac{\partial L}{\partial \theta}\right)\right). \quad (S8)$$

**Proof.** From (S5) one obtains

$$\frac{\partial L}{\partial \lambda} \frac{\partial L}{\partial \lambda} = N^2 \frac{1}{(1 - e^{-\lambda})^2} - 2N \frac{1}{1 - e^{-\lambda}} \sum_{k=1}^{n} N_k \frac{p_k}{1 - e^{-\lambda p_k}}$$

$$+ \sum_{k, \ell=1}^{n} N_k N_{\ell} \frac{p_k p_{\ell}}{(1 - e^{-\lambda p_k})(1 - e^{-\lambda p_{\ell}})}.$$
Hence, Lemma 1 yields
\[
\mathbb{E} \frac{\partial L}{\partial \lambda} \frac{\partial L}{\partial \lambda} = \frac{N^2}{(1-e^{-\lambda})^2} - \frac{2N}{1-e^{-\lambda}} \sum_{k=1}^{n} N \frac{p_k}{1-e^{-\lambda}} + \sum_{k,\ell=1}^{n} p_k p_{\ell}(N^2 - N) \\
+ \sum_{k=1}^{n} \frac{p_k p_{\ell} N}{1-e^{-\lambda}} + \sum_{k=1}^{n} \frac{p_k^2 N e^{-\lambda p_k}}{(1-e^{-\lambda p_k})(1-e^{-\lambda})} \\
= \frac{N^2}{(1-e^{-\lambda})^2} - \frac{2N^2}{(1-e^{-\lambda})^2} + \frac{N^2 - N}{(1-e^{-\lambda})^2} + \frac{N}{1-e^{-\lambda}} \sum_{k=1}^{n} \frac{p_k^2 e^{-\lambda p_k}}{1-e^{-\lambda p_k}} \\
= \frac{N}{1-e^{-\lambda}} \left( \frac{1}{1-e^{-\lambda}} \right) + \frac{N e^{\lambda}}{1-e^{-\lambda}} \sum_{k=1}^{n} \frac{p_k^2}{e^{\lambda p_k} - 1} = I_{1,1}.
\]

Straightforward calculation yields,
\[
\frac{\partial L}{\partial \lambda} \frac{\partial L}{\partial p_k} = -\frac{N \lambda}{1-e^{-\lambda}} \left( \frac{N_k}{1-e^{-\lambda p_k}} - \frac{N_n}{1-e^{-\lambda p_n}} \right) \\
+ \sum_{j=1}^{n} \frac{\lambda p_j}{1-e^{-\lambda p_j}} \left( \frac{N_j N_k}{1-e^{-\lambda p_k}} - \frac{N_j N_n}{1-e^{-\lambda p_n}} \right).
\]

By using Lemma 1 we obtain
\[
\mathbb{E} \frac{\partial L}{\partial \lambda} \frac{\partial L}{\partial p_k} = \frac{-N^2 \lambda}{(1-e^{-\lambda})^2} (1-1) + \frac{\lambda}{1-e^{-\lambda}} \left( \frac{N^2 - N}{1-e^{-\lambda}} + N \right) \sum_{j=1}^{n} p_j (1-1) \\
+ \frac{N \lambda p_k e^{-\lambda p_k}}{(1-e^{-\lambda p_k})(1-e^{-\lambda})} - \frac{N \lambda p_n e^{-\lambda p_n}}{(1-e^{-\lambda p_n})(1-e^{-\lambda})} \\
= \frac{N \lambda}{1-e^{-\lambda}} \left( \frac{p_k}{e^{\lambda p_k} - 1} - \frac{p_n}{e^{\lambda p_n} - 1} \right) = I_{1,k+1} = I_{k+1,1}.
\]

Next, note that
\[
\frac{\partial L}{\partial p_k} \frac{\partial L}{\partial p_k} = \lambda^2 \left( \frac{N_k^2}{(1-e^{-\lambda p_k})^2} - \frac{2N_k N_n}{(1-e^{-\lambda p_k})(1-e^{-\lambda p_n})} + \frac{N_n^2}{(1-e^{-\lambda p_n})^2} \right)
\]
yielding
\[
\mathbb{E} \frac{\partial L}{\partial p_k} \frac{\partial L}{\partial p_k} = \frac{\lambda^2 (1-2-1)}{1-e^{-\lambda}} \left( \frac{N^2 - N}{1-e^{-\lambda}} + N \right)^2 + \frac{\lambda^2 N}{1-e^{-\lambda}} \left( \frac{e^{-\lambda p_k}}{1-e^{-\lambda p_k}} + \frac{e^{-\lambda p_n}}{1-e^{-\lambda p_n}} \right) \\
= \frac{\lambda^2 N}{1-e^{-\lambda}} \left( \frac{1}{e^{\lambda p_k} - 1} + \frac{1}{e^{\lambda p_n} - 1} \right) = I_{k+1,k+1}.
\]

Furthermore, we have
\[
\frac{\partial L}{\partial p_k} \frac{\partial L}{\partial p_j} = \frac{N_k N_j \lambda^2}{(1-e^{-\lambda p_k})(1-e^{-\lambda p_j})} \\
+ \frac{1}{1-e^{-\lambda p_n}} \lambda^2 \left( \frac{N_n^2}{1-e^{-\lambda p_n}} - \frac{N_k N_n}{1-e^{-\lambda p_k}} - \frac{N_j N_n}{1-e^{-\lambda p_j}} \right)
\]
so that

\[
\mathbb{E} \frac{\partial L}{\partial p_k} \frac{\partial L}{\partial p_j} = \left( \frac{N^2 - N}{1 - e^{-\lambda}} + N \right) \frac{\lambda^2}{1 - e^{-\lambda}} (1 + 1 - 1 - 1) + \frac{\lambda^2 N}{(e^{\lambda p_k} - 1)(1 - e^{-\lambda})}
\]

\[= I_{k+1,j+1}.\]

Finally, we derive

\[
\frac{\partial L}{\partial p_k} \frac{\partial L}{\partial p_j} = \frac{N_k N_j \lambda^2}{(1 - e^{-\lambda p_k})(1 - e^{-\lambda p_j})} + \frac{1}{1 - e^{-\lambda p_k}} \lambda^2 \left( \frac{N_k - N}{1 - e^{-\lambda p_k}} - \frac{N_j N_n}{1 - e^{-\lambda p_j}} \right)
\]

which gives

\[
\mathbb{E} \frac{\partial L}{\partial p_k} \frac{\partial L}{\partial p_j} = \left( \frac{N^2 - N}{1 - e^{-\lambda}} + N \right) \frac{\lambda^2}{1 - e^{-\lambda}} (1 + 1 - 1 - 1) + \frac{\lambda^2 N}{(e^{\lambda p_k} - 1)(1 - e^{-\lambda})}
\]

\[= I_{k+1,j+1},\]

and finishes the proof. □

To prove positive definiteness of the information matrix, we need the following lemma first.

**Lemma 2** For \(\lambda > 0\) and \(0 < p < 1\) the inequality \(\frac{\lambda p - 1}{e^{\lambda p - 1}} p < 1\) holds, for \(p = 0, 1\) equality holds.

**Proof.** The statement is clear for \(p = 0, 1\). Let \(x = e^{\lambda}\), hence \(x > 1\). Moreover, let \(f(x, p) := \frac{e^x - 1}{x - 1}\) for \(x > 1\) and \(0 \leq p \leq 1\). We have \(\lim_{x \to 0} f(x, p) = p\) by de l’Hospital’s rule, \(\lim_{x \to \infty} f(x, p) = 0\) for \(0 \leq p < 1\) and \(\lim_{x \to 1} f(x, 1) = 1\). Hence, the prove the statement it suffices to show that \(f(x, p)\) is strictly monotonically decreasing for any fixed \(p \in (0, 1)\).

We have \(\frac{\partial f}{\partial x} = \frac{1 - px^{p-1} - (1-p)x^p}{(x-1)^2}\). To show that \(\frac{\partial f}{\partial x} < 0\) it suffices to consider the numerator. Let \(g(p) = 1 - px^{p-1} - (1-p)x^p\) for any fixed \(x\). We have \(g'(p) = -(1-x)x^{p-1} - x^{p-1}(x(1-p) + p)\log x\). Clearly, \(g(0) = g(1) = 0\). Solving \(g'(p) = 0\) yields the unique solution \(\hat{p} = \frac{1}{x} + \frac{x}{1-x}\log x\). Since \(g(0) = g(1) = 0\) the mean value theorem implies \(0 < \hat{p} < 1\). Moreover, a local minimum is attained at \(\hat{p}\), because \(g'(1) = x - 1 - \log x > 0\) for \(x > 1\). Hence, \(g(p) < 0\) for \(0 < p < 1\). Therefore, \(\frac{\partial f}{\partial x} < 0\) is shown, which finishes the proof. □

Now, the next important property of the information matrix can be proved.

**Theorem 2** The Fisher information is positive definite.

**Proof.** Positive definiteness is proved by the Ruth-Hurwitz criterium. The (leading) minors of the Fisher information need to have positive determinants. For convenience we define the leading minors from right to left. The first leading minor, \(I_{n,n}\), is clearly positive. The \(k\)th leading minor \((k < n)\), by first adding a redundant row and column and then successively using that adding multiples of columns does not change the
where the determinant, becomes

\[
\det M_k = \begin{vmatrix}
I_{n-k+1,n-k+1} & \ldots & I_{n-k+1,n} \\
\vdots & \ddots & \vdots \\
I_{n,n-k+1} & \ldots & I_{n,n}
\end{vmatrix} = \begin{vmatrix}
I_{n-k+1,n-k+1} & \ldots & I_{n-k+1,n} & 0 \\
\vdots & \ddots & \vdots & \vdots \\
I_{n,n-k+1} & \ldots & I_{n,n} & 0 \\
0 & \ldots & 0 & 1
\end{vmatrix}
\]

\[
= \frac{1}{a} \begin{vmatrix}
I_{n-k+1,n-k+1} & \ldots & I_{n-k+1,n} & 0 \\
\vdots & \ddots & \vdots & \vdots \\
I_{n,n-k+1} & \ldots & I_{n,n} & 0 \\
0 & \ldots & 0 & a
\end{vmatrix} = \frac{1}{a} \begin{vmatrix}
I_{n-k+1,n-k+1} & \ldots & I_{n-k+1,n} & a \\
\vdots & \ddots & \vdots & \vdots \\
I_{n,n-k+1} & \ldots & I_{n,n} & a \\
0 & \ldots & 0 & a
\end{vmatrix}
\]

\[
= \frac{1}{a} \begin{vmatrix}
I_{n-k+1,n-k+1} - a & \ldots & I_{n-k+1,n} - a & a \\
\vdots & \ddots & \vdots & \vdots \\
I_{n,n-k+1} - a & \ldots & I_{n,n} - a & a \\
-a & \ldots & -a & a
\end{vmatrix}
\]

By choosing \( a = \frac{N\lambda^2}{1-e^{-\lambda}} \frac{1}{e^{\lambda p_n-1}} \) we obtain

\[
\det M_k = \frac{1}{a} \left| \begin{array}{cc}
D & B^T \\
-B & 1
\end{array} \right|, \tag{S9}
\]

where \( D = (e^{\lambda p_n} - 1) \text{diag}(e^{\lambda p_n}, \ldots, e^{\lambda p_n-1}), \) and \( B = (1, \ldots, 1). \) Successively adding multiples of all but the last to the last row yields an upper triangular matrix. Hence, one obtains

\[
\det M_k = a^{-k-1} \left( 1 + \sum_{j=n-k+1}^{n-1} \frac{e^{\lambda p_j} - 1}{e^{\lambda p_n} - 1} \right) \prod_{j=n-k+1}^{n} \frac{e^{\lambda p_n} - 1}{e^{\lambda p_j} - 1} > 0. \tag{S10}
\]

By similar steps as above, the determinant of the Fisher information is

\[
\det I = \frac{1}{a \lambda} \begin{vmatrix}
\lambda I_{1,1} + bp_n & I_{1,2} - b & \ldots & I_{1,n} - b & b \\
\lambda I_{2,1} + ap_n & I_{2,2} - a & \ldots & I_{2,n} - a & a \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
\lambda I_{n,1} + ap_n & I_{n,2} - a & \ldots & I_{n,n} - a & a \\
ap_n & -a & \ldots & -a & a
\end{vmatrix}. \tag{S11}
\]

Choosing \( a = \frac{N\lambda^2}{1-e^{-\lambda}} \frac{1}{e^{\lambda p_n-1}} \) and \( b = -\frac{N\lambda p_n}{1-e^{-\lambda}} \frac{1}{e^{\lambda p_n-1}} \) yields

\[
\det I = \frac{a^{n-1}}{\lambda^2} \left| \begin{array}{ccc}
f & V & -p_n \\
V^T & D & B^T \\
p_n & -B & 1
\end{array} \right|, \tag{S12}
\]

where \( D = (e^{\lambda p_n} - 1) \text{diag}(e^{\lambda p_1}, \ldots, e^{\lambda p_1-1}), \) \( B = (1, \ldots, 1), \) \( V = (e^{\lambda p_n} - 1)(p_1(e^{\lambda p_1} - 1), \ldots, p_{n-1}(e^{\lambda p_{n-1}} - 1)^{-1}), \) and

\[
f = -\frac{e^{\lambda p_n-1} - e^{-1}}{e^{\lambda p_n-1}} + \sum_{k=1}^{n-1} \frac{p_k(e^{\lambda p_k-1})}{e^{\lambda p_k-1}}. \]
column for \( k = 1, \ldots, n - 1 \) gives

\[
\frac{\lambda^2}{a^{n-1}} \det I = \begin{vmatrix}
g & V & -p_n \\
0^T & D & B^T \\
\sum_{k=1}^n p_k & -B & 1
\end{vmatrix} = \begin{vmatrix}
g & V & -p_n \\
0^T & D & B^T \\
1 & -B & 1
\end{vmatrix},
\]

(S13)

where \( g = -\frac{e^{\lambda p_n} - 1}{e^{\lambda} - 1} \). Subtracting \( \frac{e^{\lambda p_k} - 1}{e^{\lambda} - 1} \) times column \( k + 1 \) from the last column yields

\[
\frac{\lambda^2}{a^{n-1}} \det I = \begin{vmatrix}
g & V & -\sum_{k=1}^n p_k \\
0^T & D & 0^T \\
1 & -B & h
\end{vmatrix} = \begin{vmatrix}
g & 0 & -1 \\
0^T & D & 0^T \\
1 & -B & h
\end{vmatrix} = (-1)^{n-1} \begin{vmatrix}
g & 0 & -1 \\
0^T & D & 0^T \\
1 & -B & h
\end{vmatrix}
\]

(S14)

\[
= (-1)^{2n-2} \begin{vmatrix}
g & -1 & 0 \\
1 & h & -B \\
0^T & 0^T & D
\end{vmatrix} = (gh + 1) \prod_{k=1}^{n-1} \frac{e^{\lambda p_k} - 1}{e^{\lambda} - 1},
\]

(S15)

where \( h = 1 + \sum_{k=1}^{n-1} \frac{e^{\lambda p_k} - 1}{e^{\lambda} - 1} = \sum_{k=1}^n \frac{e^{\lambda p_k} - 1}{e^{\lambda} - 1} \). We have

\[ gh + 1 = 1 - \sum_{k=1}^n \frac{e^{\lambda p_k} - 1}{e^{\lambda} - 1} > 1 - \sum_{k=1}^n p_k = 0 \quad \text{according to Lemma 2} \]

Thus, \( \det I > 0 \), implying that the Fisher information is positive definite.

## D Derivation of the Cramér-Rao lower bound

To derive the Fisher information and especially its properties it was convenient to reduce the parameter space \( \Theta \) to the lower dimensional parameter space \( \tilde{\Theta} \) by eliminating a redundant parameter. However, inverting the information given by Result 2 is cumbersome. To derive the Cramér-Rao lower bound, it is more convenient to embed the parameter space \( \Theta \) into a higher dimensional space by introducing a nuisance parameter. Namely, consider the log-likelihood function

\[
\Lambda = \Lambda(p, \lambda, \beta) = \Lambda(p, \lambda, \beta|X) = L(\lambda, p|X) + \beta(1 - \sum_{k=1}^n p_k).
\]

(S16)

Note that \( \Lambda(p, \lambda, \beta) = L(\lambda, p|X) \) since the term that has been added vanishes for any admissible set of parameters. The advantage of \( \Lambda \) is that \( p \) can be regarded as an element of \( \mathbb{R}^n \) rather than of the \((n - 1)\)-dimensional simplex when calculating the derivatives. However, the parameter space itself does not change. Moreover, \( \beta \) is regarded as a scalar, although we impose that the true value is given by \( \beta = N \frac{e^{\lambda}}{e^{\lambda} - 1} \).
Straightforward calculation yields

\[
\frac{\partial \Lambda}{\partial \lambda} = -N \frac{e^\lambda}{e^\lambda - 1} + \sum_{k=1}^{n} \frac{N_k p_k e^{\lambda p_k}}{e^{\lambda p_k} - 1}, \tag{S17a}
\]

\[
\frac{\partial \Lambda}{\partial p_k} = N_k \frac{\lambda e^{\lambda p_k}}{e^{\lambda p_k} - 1} - \beta, \tag{S17b}
\]

\[
\frac{\partial \Lambda}{\partial \beta} = 1 - \sum_{i=1}^{n} p_i. \tag{S17c}
\]

The second derivatives become

\[
\frac{\partial^2 \Lambda}{\partial \lambda^2} = N \frac{e^\lambda}{(e^\lambda - 1)^2} - \sum_{k=1}^{n} \frac{p_k^2 e^{\lambda p_k}}{(e^{\lambda p_k} - 1)^2}, \tag{S18a}
\]

\[
\frac{\partial^2 \Lambda}{\partial \lambda \partial p_k} = \frac{\partial^2 \Lambda}{\partial p_k \partial \lambda} = N_k \frac{e^{\lambda p_k}}{e^{\lambda p_k} - 1} \left(1 - \frac{\lambda p_k}{e^{\lambda p_k} - 1}\right), \tag{S18b}
\]

\[
\frac{\partial^2 \Lambda}{\partial p_k^2} = -N_k \frac{\lambda^2 e^{\lambda p_k}}{e^{\lambda p_k} - 1}, \tag{S18c}
\]

\[
\frac{\partial^2 \Lambda}{\partial p_k \partial p_j} = 0 \text{ for } k \neq j, \tag{S18d}
\]

\[
\frac{\partial^2 \Lambda}{\partial \beta \partial p_k} = \frac{\partial^2 \Lambda}{\partial p_k \partial \beta} = -1 \text{ for } k \neq j, \tag{S18e}
\]

\[
\frac{\partial^2 \Lambda}{\partial \lambda \partial \beta} = 0, \tag{S18f}
\]

\[
\frac{\partial^2 \Lambda}{\partial \beta^2} = 0. \tag{S18g}
\]

Application of Lemma [4] yields the following entries of the information matrix:

\[
-\mathbb{E} \left( \frac{\partial^2 \Lambda}{\partial \lambda^2} \right) = -\frac{N e^\lambda}{e^\lambda - 1} \left( \frac{1}{e^\lambda - 1} - \sum_{k=1}^{n} \frac{p_k^2}{e^{\lambda p_k} - 1}\right), \tag{S19a}
\]

\[
-\mathbb{E} \left( \frac{\partial^2 \Lambda}{\partial \lambda \partial p_k} \right) = -\mathbb{E} \left( \frac{\partial^2 \Lambda}{\partial p_k \partial \lambda} \right) = -\frac{N e^\lambda}{e^\lambda - 1} \left(1 - \frac{\lambda p_k}{e^{\lambda p_k} - 1}\right), \tag{S19b}
\]

\[
-\mathbb{E} \left( \frac{\partial^2 \Lambda}{\partial p_k^2} \right) = \frac{N e^\lambda \lambda^2}{e^\lambda - 1} \frac{1}{e^{\lambda p_k} - 1}, \tag{S19c}
\]

\[
-\mathbb{E} \left( \frac{\partial^2 \Lambda}{\partial p_k \partial p_j} \right) = 0 \text{ for } k \neq j, \tag{S19d}
\]

\[
-\mathbb{E} \left( \frac{\partial^2 \Lambda}{\partial \beta \partial p_k} \right) = -\mathbb{E} \left( \frac{\partial^2 \Lambda}{\partial p_k \partial \beta} \right) = 1 \text{ for } k \neq j, \tag{S19e}
\]

\[
-\mathbb{E} \left( \frac{\partial^2 \Lambda}{\partial \lambda \partial \beta} \right) = -\mathbb{E} \left( \frac{\partial^2 \Lambda}{\partial \lambda \partial \beta} \right) = 0, \tag{S19f}
\]

\[
-\mathbb{E} \left( \frac{\partial^2 \Lambda}{\partial \beta^2} \right) = 0. \tag{S19g}
\]

Hence, the Fisher information has the following structure

\[
\tilde{I}_N = \begin{pmatrix} D & B^T \\ B & A \end{pmatrix} \tag{S20a}
\]
with
\[ A = \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} b_1 & \cdots & b_n \\ 1 & \cdots & 1 \end{pmatrix} \quad \text{and} \quad D = \text{diag}(d_1, \ldots, d_n), \]
(S20b)

where
\[ a = -E \left( \frac{\partial^2 \Lambda}{\partial \lambda^2} \right), \quad b_k = -E \left( \frac{\partial^2 \Lambda}{\partial \lambda \partial p_k} \right) \quad \text{and} \quad d_k = -E \left( \frac{\partial^2 \Lambda}{\partial p_k^2} \right). \]
(S20c)

The inverse Fisher information can be derived using a blockwise inversion formula, namely
\[ \tilde{I}_N^{-1} = \begin{pmatrix} D^{-1} + D^{-1} B^T (A - BD^{-1} B^T)^{-1} B D^{-1} & -D^{-1} B^T (A - BD^{-1} B^T)^{-1} D \ 
- (A - BD^{-1} B^T)^{-1} B D^{-1} & (A - BD^{-1} B^T)^{-1} \end{pmatrix}. \]

The formula applies whenever, \( d_i \neq 0 \) and the \( 2 \times 2 \) matrix \( A - BD^{-1} B^T \) is invertible. Blockwise inversion is particularly simple here because \( D \) is diagonal. This is the reward of introducing the nuisance parameter \( \beta \). Without \( \beta \), \( D \) would not be diagonal and inverting the Fisher information would be more involved. Note that
\[ \tilde{I}_N^{-1} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \]
(S21)

where \( a_{11} = a - \sum_{k=1}^{n} \frac{b_k^2}{d_k^2} \), \( a_{12} = a_{21} = - \sum_{k=1}^{n} \frac{b_k}{d_k} \) and \( a_{22} = - \sum_{k=1}^{n} \frac{1}{d_k} \). Its inverse is given by
\[ (A - BD^{-1} B^T)^{-1} = \frac{1}{a_{11} a_{22} - a_{12}^2} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{12} & a_{11} \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{12} & m_{22} \end{pmatrix}. \]

Moreover,
\[ (A - BD^{-1} B^T)^{-1} B D^{-1} = \begin{pmatrix} \frac{m_{11} b_1}{d_1} + \frac{m_{12} b_2}{d_1} & \cdots & \frac{m_{11} b_n}{d_1} + \frac{m_{12} b_{n+1}}{d_1} \\ \frac{m_{12} b_1}{d_1} + \frac{m_{22} b_2}{d_1} & \cdots & \frac{m_{12} b_n}{d_1} + \frac{m_{22} b_{n+1}}{d_1} \end{pmatrix} \]
and
\[ D^{-1} B^T (A - BD^{-1} B^T)^{-1} B D^{-1} = \begin{pmatrix} 1 \\ d_i d_j \end{pmatrix} \begin{pmatrix} m_{11} b_i b_j + m_{12} (b_i + b_j) + m_{22} \\ 1 \end{pmatrix}_{ij}. \]

Hence, the entries of the inverse Fisher information \( \tilde{I}_N^{-1} \) are
\[ \begin{align*}
(\tilde{I}_N^{-1})_{ii} &= \frac{1}{d_i} + \frac{a_{22} b_i^2 - 2 a_{12} b_i + a_{11}}{a_{11} a_{22} - a_{12}^2 d_i^2} \quad \text{for} \quad i = 1, \ldots, n, \\
(\tilde{I}_N^{-1})_{ij} &= \frac{a_{22} b_i b_j - a_{12} (b_i + b_j) + a_{11}}{(a_{11} a_{22} - a_{12}^2) d_i d_j} \quad \text{for} \quad i, j = 1, \ldots, n \quad \text{and} \quad i \neq j, \\
(\tilde{I}_N^{-1})_{n+1,j} &= \frac{a_{12} b_j - a_{22} h_j}{(a_{11} a_{22} - a_{12}^2) d_j} \quad \text{for} \quad j = 1, \ldots, n, \\
(\tilde{I}_N^{-1})_{n+1,n+1} &= \frac{1}{a_{11} - a_{12}^2 / a_{22}}.
\end{align*} \]
where we do not need to calculate its last row and column because these correspond to the nuisance parameter $\beta$. After some algebraic manipulation we obtain for $i, j = 1, \ldots, n$ and $i \neq j$:

$$
(\tilde{I}^{-1}_N)_{ii} = \frac{(e^\lambda - 1)^2}{\lambda^2 Ne^\lambda} \left( e^{\lambda p_i} - 1 + \frac{p_i^2 C}{e^\lambda - 1} - \frac{2 p_i (e^{\lambda p_i} - 1) (e^{\lambda p_i} - 1)}{e^\lambda - 1 - C} \right),
$$

(S22a)

$$
(\tilde{I}^{-1}_N)_{ij} = \frac{(e^\lambda - 1)^2}{\lambda^2 Ne^\lambda} \left( p_i p_j C - p_i (e^{\lambda p_i} - 1) + \frac{(e^{\lambda p_i} - 1) (e^{\lambda p_i} - 1)}{e^\lambda - 1 - C} \right),
$$

(S22b)

$$
(\tilde{I}^{-1}_N)_{n+1,j} = \frac{(e^\lambda - 1)^2}{\lambda N e^\lambda} \left( e^{\lambda p_j} - 1 - p_j C \right),
$$

(S22c)

$$
(\tilde{I}^{-1}_N)_{n+1,n+1} = \frac{(e^\lambda - 1)^2}{N e^\lambda} \left( e^\lambda - 1 - C \right),
$$

(S22d)

where

$$
C = \sum_{k=1}^{n} (e^{\lambda p_k} - 1).
$$

(S22e)

Since we do not need the last row and column of the inverse Fisher information, let $W$ denote the matrix obtained by deleting the last row and column of $(\tilde{I}^{-1}_N)$. (Note that this is different from first deleting the last row and column of the information matrix and then inverting it.) The Cramér-Rao bound $V$ presented in Result 3 is obtained by switching rows and columns in $V$.

Let $T(X) = (T_{p_1}(X), \ldots, T_{p_n}(X), T_\lambda(X))$ be an estimator for $(p, \lambda)$ with expectation $\mathbb{E}(T(X)) = \phi(X)$. Then, $\text{Cov}(T(X)) \geq \frac{\partial \phi}{\partial(p, \lambda)} W \left( \frac{\partial \phi}{\partial(p, \lambda)} \right)^T$. (Note, that one can still incorporate the nuisance parameter $\beta$ and use $\tilde{I}^{-1}_N$ by estimating $T_\beta(X) = 0$.)

### E Variance of average MOI

MOI is the average number of super-infections, which – assuming the conditional Poisson distribution – equals

$$
\psi = f(\lambda) = \frac{\lambda}{1 - e^{-\lambda}},
$$

(S23)

a 1-1 map between $\lambda$ and average MOI. Since $\Lambda(\lambda, \beta, p) = \Lambda(f^{-1}(\psi), \beta, p)$, the chain rule implies

$$
\frac{\partial \Lambda}{\partial \psi} = \frac{\partial \Lambda}{\partial \lambda} \frac{\partial f^{-1}}{\partial \psi},
$$

(S24)
and the derivative with respect to the other parameters do not change. Further application of the chain rule yields
\[
\frac{\partial^2 \lambda}{\partial \psi^2} = \frac{\partial^2 \lambda}{\partial \lambda^2} \left( \frac{\partial f^{-1}}{\partial \psi} \right)^2 + \frac{\partial \lambda}{\partial \lambda} \frac{\partial^2 f^{-1}}{\partial \psi^2}, \tag{S25a}
\]
\[
\frac{\partial^2 \lambda}{\partial \psi \partial \psi} = \frac{\partial^2 \lambda}{\partial \psi \partial \psi} \frac{\partial f^{-1}}{\partial \psi} + \frac{\partial \lambda}{\partial \psi} \frac{\partial^2 f^{-1}}{\partial \psi^2} = \frac{\partial^2 \lambda}{\partial \psi \partial \psi} \frac{\partial f^{-1}}{\partial \psi}, \tag{S25b}
\]
\[
\frac{\partial^2 \lambda}{\partial \psi \partial \beta} = \frac{\partial^2 \lambda}{\partial \psi \partial \beta} \frac{\partial f^{-1}}{\partial \psi} + \frac{\partial \lambda}{\partial \psi} \frac{\partial^2 f^{-1}}{\partial \psi^2} = \frac{\partial^2 \lambda}{\partial \psi \partial \beta} \frac{\partial f^{-1}}{\partial \psi}, \tag{S25c}
\]
while the remaining second derivatives remain unchanged as in (S18). Note that the score function has mean zero and particularly satisfies \( \mathbb{E} \frac{\partial \lambda}{\partial \lambda} = 0 \). (That the score function has mean zero can be seen directly from (S17) and Lemma 1. Here it is important to keep in mind that \( \mathbf{p} \) is an element of the simplex and \( \beta = N \frac{\lambda e^\lambda}{e^\lambda - 1} \) for the true parameter and the MLE.) This fact and the inverse function theorem yield
\[
- \mathbb{E} \frac{\partial^2 \lambda}{\partial \psi^2} = - \left( \frac{\partial f^{-1}}{\partial \psi} \right)^2 \mathbb{E} \frac{\partial^2 \lambda}{\partial \lambda^2} = - \left( \frac{\partial f}{\partial \lambda} \right)^{-2} \mathbb{E} \frac{\partial^2 \lambda}{\partial \lambda^2}, \tag{S26a}
\]
\[
- \mathbb{E} \frac{\partial^2 \lambda}{\partial \psi \partial \psi} = - \mathbb{E} \frac{\partial^2 \lambda}{\partial \psi \partial \psi} \frac{\partial f^{-1}}{\partial \psi} - \mathbb{E} \frac{\partial^2 \lambda}{\partial \psi \partial \psi} \frac{\partial f}{\partial \lambda} = - \left( \frac{\partial f}{\partial \lambda} \right)^{-1} \mathbb{E} \frac{\partial^2 \lambda}{\partial \psi \partial \psi}, \tag{S26b}
\]
\[
- \mathbb{E} \frac{\partial^2 \lambda}{\partial \psi \partial \beta} = - \mathbb{E} \frac{\partial^2 \lambda}{\partial \psi \partial \beta} \frac{\partial f^{-1}}{\partial \psi} - \mathbb{E} \frac{\partial^2 \lambda}{\partial \psi \partial \beta} \frac{\partial f}{\partial \lambda} = - \left( \frac{\partial f}{\partial \lambda} \right)^{-1} \mathbb{E} \frac{\partial^2 \lambda}{\partial \psi \partial \beta}. \tag{S26c}
\]

The Fisher information for the parameters \((\mathbf{p}, \psi, \beta)\) is denoted by \( \mathbf{J}_N \), and given by
\[
\mathbf{J}_N = \text{diag}(1, \ldots, 1, \left( \frac{\partial f}{\partial \lambda} \right)^{-1}, 1) \mathbf{I}_N \text{diag}(1, \ldots, 1, \left( \frac{\partial f}{\partial \lambda} \right)^{-1}, 1). \tag{S27}
\]
Hence,
\[
\mathbf{J}_N^{-1} = \text{diag}(1, \ldots, 1, \left( \frac{\partial f}{\partial \lambda} \right)^{-1}, 1) \mathbf{I}_N^{-1} \text{diag}(1, \ldots, 1, \left( \frac{\partial f}{\partial \lambda} \right)^{-1}, 1). \tag{S28}
\]
Clearly, \( f'(\lambda) = \frac{\partial f}{\partial \lambda} = e^\lambda \frac{\lambda e^\lambda - 1}{e^\lambda - 1} \), so that for \( i, j = 1, \ldots, n \) and \( i \neq j \):
\[
(\mathbf{J}_N^{-1})_{ii} = (\mathbf{I}_N^{-1})_{ii}, \quad (\mathbf{J}_N^{-1})_{ij} = (\mathbf{I}_N^{-1})_{ij}, \tag{S29a}
\]
\[
(\mathbf{J}_N^{-1})_{n+1,i} = f'(\lambda)(\mathbf{I}_N^{-1})_{n+1,i} = \frac{e^\lambda - \lambda - 1}{\lambda N} e^\lambda - 1 - C, \tag{S29b}
\]
\[
(\mathbf{J}_N^{-1})_{n+1,n+1} = f'(\lambda)^2(\mathbf{I}_N^{-1})_{n+1,n+1} = \frac{e^\lambda(\lambda^2 - 1)}{N(e^\lambda - 1)^2} - \frac{e^\lambda - 1 - C}{C}, \tag{S29c}
\]
where \( C \) is given by (S22c). The last row and column of \( \mathbf{J}_N^{-1} \) are irrelevant as they correspond to the nuisance parameter. By deleting the last row and column and rearranging the remaining rows and columns, the Cramér-Rao bound of Remark 5 is obtained.

### F Constructing data

#### F.1 Dependence on true parameters and sample size

To investigate the method’s sensitivity on the true parameters we proceeded as follows. For a fixed set of parameters \( \theta = (\lambda, \mathbf{p}) \) we simulated \( K \) data sets \( \mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(K)} \) of
sample size $N$ as described below. For each data set $X^{(k)}$, we derived the MLE $\hat{\theta}^{(k)}$.

(Only regular data sets, i.e. those satisfying $N_k \neq N$ for all $k$ and $\sum_{k=1}^{n} N_k > N$ were constructed, so the MLE was well defined.) We then attained the MLE’s performance as described below. This was repeated for several parameter choices $\theta$. Particularly, the parameters’ dimension, $n$, might also affect the quality of the MLE. Therefore, parameters of different dimension were carefully chosen.

F.2 Dependence on sample size and dimension

To investigate the method’s dependency on sample size we repeated the procedure described above for several choices of sample size $N$. This provides not only a proxy for the method’s overall performance under varying sample size, but also allows identification of the necessary sample size for different parameter regions to achieve certain performance goals.

F.3 Model fit

The performance of the methods are justified only if the true model that generates the data equals the model used to derive the likelihood function. To investigate a potential dependence on model violations, we generated the data sets $X^{(k)}$ under alternative models, and repeated the procedures described above for each of them. This gives a proxy for the method’s sensitivity to model violations. Particularly, we assumed that the number of (super-)infections follows a conditional Poisson (correct model), shifted Poisson, conditional and shifted binomial, and uniform distribution as described below.

F.4 Constructing the data

We first describe how a data set $X$ is constructed under the conditional Poisson model, which is the model underlying the MLE. We then describe adaptations to construct data under different models.

**Conditional Poisson model.** The number $Y$ of parasites infecting a host follows a conditional Poisson distribution, i.e.,

$$P(Y = m) = \frac{1}{e^\lambda - 1} \frac{\lambda^m}{m!} \quad \text{for } m \geq 1. \quad (S30)$$

MOI is the average number of infecting lineages and is given by

$$\psi = \psi_{CP} = \mathbb{E} Y = \frac{\lambda}{1 - e^{-\lambda}}. \quad (S31)$$

Moreover, conditional on being infected by $m$ parasites, the infecting lineages are drawn from a multinomial distribution with parameters $m$ and $p$, i.e.,

$$P(m = (m_1, \ldots, m_n) | Y = m) = \frac{m!}{m_1! \cdots m_n!} p_1^{m_1} \cdots p_n^{m_n}, \quad (S32)$$

where $m_j \in \{0, \ldots, m\}$ is the number of times the host is infected with lineage $j$. Clearly, $m_1 + \ldots + m_n = m$. The corresponding blood sample is hence $(\text{sign } m_1, \ldots, \text{sign } m_n)$. 


Summarizing, first a random number is generated according to (S30). Then, an integer vector \( \mathbf{m} = (m_1, \ldots, m_n) \) is generated according to (S32), which yields the first blood sample as \( \mathbf{x}_1 = (\text{sign} \, m_1, \ldots, \text{sign} \, m_n) \). By repeating this procedure \( N \) times a dataset \( \mathbf{X} \) containing \( N \) blood samples is obtained.

**Shifted Poisson model.** The difference to the conditional Poisson model is the distribution of super-infecting lineages. Namely, the number of super-infecting lineages follow a Poisson distribution that is shifted by 1, i.e.,

\[
P(Y = m) = \frac{\lambda^{m-1}}{e^{\lambda} (m-1)!} \quad \text{for } m \geq 1.
\]  

(S33)

The mean number of super-infections (MOI) is

\[
\psi = \psi_{SP} = \mathbb{E} Y = \lambda + 1.
\]  

(S34)

**Binomial model.** Here, the number of super-infecting lineages is conditionally binomially distributed with parameters \( M \) and \( p \), i.e.,

\[
P(Y = m) = \binom{M}{m} p^m (1-p)^{M-m} \frac{1 - (1-p)^M}{1 - (1-p)^{M-m}} \quad \text{for } m \in \{1, \ldots, M\}.
\]  

(S35)

MOI in this case is

\[
\psi = \psi_{CB} = \mathbb{E} Y = \frac{Mp}{1 - (1-p)^M}.
\]  

(S36)

For large \( M \), small \( p \) and \( Mp \sim \lambda \) this model approximates the conditional Poisson model.

**Shifted binomial model.** As an approximation to the shifted Poisson model, the number of super-infecting lineages follows a shifted binomial distribution, namely,

\[
P(Y = m) = \binom{M}{m-1} p^{m-1} (1-p)^{M-m+1} \quad \text{for } m = 1, \ldots, M+1,
\]  

(S37)

for which MOI is

\[
\psi = \psi_{SB} = \mathbb{E} Y = Mp + 1.
\]  

(S38)

**Uniform model.** The assumption here is that a host is infected by 1 to \( M \) lineages with the same probability, i.e.,

\[
P(Y = m) = \frac{1}{M}
\]  

(S39)

for which MOI is

\[
\psi = \psi_{UN} = \mathbb{E} Y = \frac{M+1}{2}.
\]  

(S40)

**Parameter values.** We simulated data sets for all models for a wide range of parameters. Importantly, these were chosen so that the values of the MOI parameters were matching, i.e., \( \psi_{CP} = \psi_{SP} = \psi_{CB} = \psi_{SB} \), except for the uniform model, for which \( \psi_{UN} \) can attain only a limited number of values. The parameters are summarized in Table A.
F.5 Measuring MLE’s performance

The first aim is to measure the performance of the MLE for the Poisson parameter $\lambda$, or rather for MOI, $\psi = \frac{\lambda}{1 - e^{-\lambda}}$. (This is only the correct parameter for the conditional Poisson model, not for the alternative models used to simulate data.) We refer to a combination of parameters and a given model as a parameter set. For each parameter set $K = 10,000$ data sets $X^{(k)}$ were randomly generated as described above. For each data set the MLE $\hat{\lambda}^{(k)}$ is derived. We then derived the empirical mean ($E_K \hat{\lambda}$) and variance ($\text{Var}_K \hat{\lambda}$) of the $\hat{\lambda}^{(k)}$s, as well as the 0.5%, 2.5%, 5%, 95%, 97.5%, and 99.5% percentiles. These measures are informative about the distribution of bias and variance of the MLE.

To assess quality of the estimate for the lineage frequencies, we obtained the empirical means, variances and 0.5%, 2.5%, 5%, 95%, 97.5%, and 99.5% quantiles for each lineage frequency, i.e., for $p_1, \ldots, p_n$. Although these measures are important, usually the focus does not lie on single lineage frequencies but on the overall distribution, particularly because the parameters are correlated. Moreover, as $n$ increases it becomes increasingly tedious to look at measures for single frequencies, rather than on a single quantity. Therefore, we also calculated the distance $d(p, \hat{p})$ from the true and estimated parameters. For each set of parameters we then obtained the empirical means and variances of the distance, i.e., $E_K d(p, \hat{p})$ and $\text{Var}_K d(p, \hat{p})$. Since $p$ is multidimensional, it is difficult to decide on an appropriate distance. Because we do not want to decide on a “best” distance measure, we used several distance measures (Euclidian-norm, 1-norm, supreme norm, Jensen-Shannon divergence, Hellinger and Bhattacharyy distances). Qualitatively they all yielded the same results.

Tables
The table below displays the parameter values used to construct the data sets for the simulation study. Parameters were chosen so that the same MOI parameters were obtained for the conditional Poisson, shifted Poisson, conditional binomial, and shifted binomial model. In the case of the uniform model, it is impossible to accomplish this. $K$ data sets were created for each resulting combination of values of $\psi$, $p$, $N$.

| par. | description | values |
|------|-------------|--------|
| $K$  | number of data sets per par. combination | 10,000 or 100,000 |
| $\lambda$ | par. of conditional Poisson model | $\lambda = 0.1, \ldots, 2$ in steps of 0.05 |
| $\psi_{CP}$ | MOI parameter | $\frac{\lambda}{1-e^{-\lambda}}$ with $\lambda$ as above |
| $\psi_{SP}$ | MOI parameter | $\lambda + 1$ with $\lambda$ such that $\psi_{SP} = \psi_{CP}$ |
| $\psi_{CB}$ | MOI parameter | $\frac{M \lambda}{1-(1-p)^{M \lambda}}$ with $M = 8$, $p$ such that $\psi_{CB} = \psi_{CP}$ |
| $\psi_{UN}$ | MOI parameter | $\frac{M+1}{2}$ with $M = 2, 3, 4, 5, 7, 8$ |
| $N$ | sample size | 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 |
| $n$ | number of lineages | 2, 3, 4, 5, 10 |
| $p$ | lineage-frequencies for $n = 2$ | (0.5, 0.5), (0.6, 0.4), (0.7, 0.3), (0.8, 0.2), (0.9, 0.1) |
| | | (1/3, 1/3, 1/3), (0.45, 0.45, 0.1), (0.6, 0.3, 0.1), |
| | | (0.7, 0.2, 0.1), (0.8, 0.15, 0.05), (0.9, 0.05, 0.05) |
| | $n = 3$ | (0.25, 0.25, 0.25, 0.25) | (0.4, 0.2, 0.2, 0.2), |
| | | (0.4, 0.4, 0.1, 0.1) (0.5, 0.25, 0.2, 0.05), |
| | | (0.6, 0.3, 0.05, 0.05), (0.7, 0.15, 0.1, 0.05) |
| | $n = 4$ | (0.2, 0.2, 0.2, 0.2, 0.2), (0.25, 0.2, 0.2, 0.2, 0.15), |
| | | (0.3, 0.2, 0.2, 0.15, 0.15), (0.4, 0.2, 0.2, 0.1, 0.1), |
| | | (0.5, 0.2, 0.2, 0.05, 0.05), (0.6, 0.2, 0.1, 0.05, 0.05), |
| | | (0.7, 0.1, 0.1, 0.05, 0.05), (0.8, 0.05, 0.05, 0.05) |
| | $n = 5$ | (0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1), |
| | | (0.4, 0.4, 0.09, 0.05, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01) |
| | $n = 10$ | (0.7, 0.13, 0.1, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01), |
| | | (0.91, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01, 0.01) |

**Table A. Summary of parameters.** Displayed are the parameter values used to construct the data sets for the simulation study. Parameters were chosen so that the same MOI parameters were obtained for the conditional Poisson, shifted Poisson, conditional binomial, and shifted binomial model. In the case of the uniform model it is impossible to accomplish this. $K$ data sets were created for each resulting combination of values of $\psi$, $p$, $N$. 
References

1. Schneider KA, Escalante AA. A Likelihood Approach to Estimate the Number of Co-Infections. PLoS ONE. 2014;9(7):e97899. doi:10.1371/journal.pone.0097899.