Supporting Information for 'Estimation of Incubation Period and Generation Time Based on Observed Length-biased Epidemic Cohort with Censoring for COVID-19 Outbreak in China' by Deng et al.

Web Appendix A.

Here three parametric forms are used to fit the distribution of incubation period (inter-arrival time) I, they are Gamma distribution, Weibull distribution and Log-normal distribution. Let f_I and h be the pdf of I and V respectively, and let F_I and H be the cdf of I and V respectively.

Example 1. Suppose that $I \sim \Gamma(\alpha, \beta)$ with density

$$f_I(t;\boldsymbol{\theta}) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} \quad (t > 0),$$

where $\boldsymbol{\theta} = (\alpha, \beta)^\top$ is unknown ($\alpha > 0$, $\beta > 0$), then

$$h(t;\boldsymbol{\theta}) = \frac{\beta}{\alpha}[1 - \Gamma(t, \alpha, \beta)] \quad (t > 0),$$

$$H(t;\boldsymbol{\theta}) = \Gamma(t, \alpha + 1, \beta) + \frac{\beta t}{\alpha}[1 - \Gamma(t, \alpha, \beta)] \quad (t > 0),$$

where $\Gamma(u, \alpha, \beta)$ is the cdf of $\Gamma(\alpha, \beta)$ at u.

Example 2. Suppose that $I \sim W(k, \lambda)$ with density

$$f_I(t;\boldsymbol{\theta}) = \frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1} \exp\left\{-\left(\frac{t}{\lambda}\right)^k\right\} \quad (t > 0),$$

where $\boldsymbol{\theta} = (k, \lambda)^\top$ is unknown ($k > 0$, $\lambda > 0$), then

$$h(t;\boldsymbol{\theta}) = \frac{k}{\lambda}\Gamma\left(\frac{1}{k}\right)^{-1} \exp\left\{-\left(\frac{t}{\lambda}\right)^k\right\} \quad (t > 0),$$

$$H(t;\boldsymbol{\theta}) = \Gamma\left(\left(\frac{t}{\lambda}\right)^k, \frac{1}{k}, 1\right) \quad (t > 0).$$

Example 3. Suppose that $I \sim LN(\mu, \sigma^2)$ with density

$$f_I(t; \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}t} \exp\left\{-\frac{(\log t - \mu)^2}{2\sigma^2}\right\} \quad (t > 0),$$

where $\boldsymbol{\theta} = (\mu, \sigma^2)^\top$ is unknown ($\sigma > 0$), then

$$h(t; \boldsymbol{\theta}) = \exp\left\{-\mu - \frac{1}{2}\sigma^2\right\}\left[1 - \Phi(\log t, \mu, \sigma^2)\right] \quad (t > 0),$$

$$H(t; \boldsymbol{\theta}) = \Phi(\log t, \mu + \sigma^2, \sigma^2) + t\exp\left\{-\mu - \frac{1}{2}\sigma^2\right\}\left[1 - \Phi(\log t, \mu, \sigma^2)\right] \quad (t > 0),$$

where $\Phi(u, \mu, \sigma^2)$ is the cdf of normal distribution $N(\mu, \sigma^2)$ at u.

Web Appendix B. Likelihood approach for the mixture distribution

The log-likelihood of the mixture distribution can be derived by introducing a latent variable. Specifically, denote $\delta_j = 1$ if the jth individual in our cohort got infected at the departure, $\delta_j = 0$ if the individual got infected before departure, and t_j to be the observed time difference from the event of leaving Wuhan to symptoms onset for $j = 1, \ldots, m$. Note that only $\{t_1, \ldots, t_m\}$ are observed, and $\{\delta_1, \ldots, \delta_m\}$ are unobserved. We can rewrite this problem as a mixture distribution below,

$$\delta_j \sim Bin(1, \pi), \quad j = 1, \ldots, m.$$

$$t_j \mid (\delta_j = 1) \sim f_I^p(\cdot; \boldsymbol{\theta}), \quad t_j \mid (\delta_j = 0) \sim h^p(\cdot; \boldsymbol{\theta}),$$

and the conditional likelihood is

$$\begin{aligned} L(\boldsymbol{\theta}; t_1, \ldots, t_m \mid \delta_1, \ldots, \delta_m) &= \prod_{j=1}^{m} \{f_I^p(t_j; \boldsymbol{\theta})\}^{\delta_j} \{h^p(t_j; \boldsymbol{\theta})\}^{1-\delta_j} \\ &= \prod_{j=1}^{m} \{\delta_j f_I^p(t_j; \boldsymbol{\theta}) + (1 - \delta_j) h^p(t_j; \boldsymbol{\theta})\}. \end{aligned}$$

By integrating the unobservable $\{\delta_1, \ldots, \delta_m\}$ out, the likelihood is reduced to

$$L(\boldsymbol{\theta}, \pi; t_1, \ldots, t_m) = \prod_{j=1}^{m} \{\pi f_I^p(t_j; \boldsymbol{\theta}) + (1 - \pi) h^p(t_j; \boldsymbol{\theta})\}.$$

Hence, the estimates of parameters can be obtained by Newton-Raphson or EM algorithm(Dempster et al., 1977; Booth and Hobert, 1999). Note that extra cautions need to be taken when an EM algorithm is implemented due to the local maximum issue.

Web Appendix C. Proof of Theorem 1 and 2

To study the large-sample properties of the MLE and the likelihood ratio statistics $R_1(\boldsymbol{\theta}_0, \pi_0)$ and $R_2(\pi_0)$, we consider the behavior of $\ell(\boldsymbol{\theta}, \pi)$ for $(\boldsymbol{\theta}^\top, \pi)^\top = (\boldsymbol{\theta}_0^\top, \pi_0)^\top + n^{-1/2}\xi$ with $\xi = (\xi_1^\top, \xi_2)^\top = O_p(1)$.

By second-order Taylor expansion and weak law of large numbers, we have

$$\ell((\boldsymbol{\theta}_0^\top, \pi_0)^\top + m^{-1/2}\xi) = \ell(\boldsymbol{\theta}_0, \pi_0) + u_m^\top \xi - \frac{1}{2}\xi^\top U \xi + r_m \xi^\top \xi, \tag{S1}$$

where $r_m = o_p(1)$ uniformly for all $(\boldsymbol{\theta}, \pi)$ in a neighborhood of $(\boldsymbol{\theta}_0, \pi_0)$,

$$u_{m1} = m^{-1/2} \sum_{i=1}^{m} \frac{\pi_0 \nabla_{\boldsymbol{\theta}} f_I^p(t; \boldsymbol{\theta}_0) + (1-\pi_0)\nabla_{\boldsymbol{\theta}} h^p(t; \boldsymbol{\theta}_0)}{\pi_0 f_I^p(t; \boldsymbol{\theta}_0) + (1-\pi_0) h^p(t; \boldsymbol{\theta}_0)},$$

$$u_{m2} = m^{-1/2} \sum_{i=1}^{m} \frac{f_I^p(t; \boldsymbol{\theta}_0) - h^p(t; \boldsymbol{\theta}_0)}{\pi_0 f_I^p(t; \boldsymbol{\theta}_0) + (1-\pi_0) h^p(t; \boldsymbol{\theta}_0)},$$

and $U = (U_{ij})_{1 \le i,j \le 2}$ with

$$U_{11} = E\left\{\frac{\pi_0 \nabla_{\boldsymbol{\theta}} f_I^p(t; \boldsymbol{\theta}_0) + (1-\pi_0)\nabla_{\boldsymbol{\theta}} h^p(t; \boldsymbol{\theta}_0)}{\pi_0 f_I^p(t; \boldsymbol{\theta}_0) + (1-\pi_0) h^p(t; \boldsymbol{\theta}_0)}\right\}^{\otimes 2},$$

$$U_{12} = E\left[\frac{\{\pi_0 \nabla_{\boldsymbol{\theta}} f_I^p(t; \boldsymbol{\theta}_0) + (1-\pi_0)\nabla_{\boldsymbol{\theta}} h^p(t; \boldsymbol{\theta}_0)\}\{\nabla_{\boldsymbol{\theta}} f_I^p(t; \boldsymbol{\theta}_0) - \nabla_{\boldsymbol{\theta}} h^p(t; \boldsymbol{\theta}_0)\}^\top}{\{\pi_0 f_I^p(t; \boldsymbol{\theta}_0) + (1-\pi_0) h^p(t; \boldsymbol{\theta}_0)\}^2}\right],$$

$$U_{22} = E\left\{\frac{f_I^p(t; \boldsymbol{\theta}_0) - h^p(t; \boldsymbol{\theta}_0)}{\pi_0 f_I^p(t; \boldsymbol{\theta}_0) + (1-\pi_0) h^p(t; \boldsymbol{\theta}_0)}\right\}^2.$$

Here $A^{\otimes 2} = AA^\top$ for a vector or matrix A. It is straightforward to see that $\mathrm{var}(u_m) = U$ and $u_m \xrightarrow{d} N(0, U)$.

Let $\widehat{\xi} = (\widehat{\xi}_1^\top, \widehat{\xi}_2)^\top = \sqrt{m}((\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top, \widehat{\pi} - \pi_0)^\top$. If $(\boldsymbol{\theta}_0, \pi_0)$ is an interior point of the parameter space, it follows from (S1) that the MLE $(\widehat{\boldsymbol{\theta}}^\top, \widehat{\pi})^\top = (\boldsymbol{\theta}_0^\top, \pi_0)^\top + m^{-1/2}\widehat{\xi}$ satisfies

$$\widehat{\xi} = U^{-1} u_m + o_p(1) \xrightarrow{d} N(0, U^{-1}),$$

and that

$$R(\boldsymbol{\theta}_0, \pi_0) = 2u_m^\top \widehat{\xi} - \widehat{\xi}^\top U \widehat{\xi} + o_p(1) = u_m^\top U^{-1} u_m + o_p(1) \xrightarrow{d} \chi^2_{q_{\boldsymbol{\theta}}+1},$$

where $q_{\boldsymbol{\theta}}$ is the dimension of $\boldsymbol{\theta}$. Similarly it is straightforward to see that $R_2(\pi_0) \xrightarrow{d} \chi^2_1$. This proves Theorem 1.

To proving Theorem 2, we re-express (S1) as

$$\begin{aligned}
\ell((\boldsymbol{\theta}_0^\top, \pi_0)^\top + m^{-1/2}\xi) &= \ell(\boldsymbol{\theta}_0, \pi_0) + u_{m1}^\top \xi_1 + u_{m2}^\top \xi_2 \\
&\quad - \frac{1}{2}\xi_1^\top U_{11}\xi_1 - \xi_1^\top U_{12}\xi_2 - \frac{1}{2}\xi_2^\top U_{22}\xi_2 + r_m \xi^\top \xi \\
&= \ell(\boldsymbol{\theta}_0^\top, \pi_0)^\top + \frac{1}{2} u_{m1}^\top U_{11}^{-1} u_{m1} \\
&\quad - \frac{1}{2}\{\xi_1 - U_{11}^{-1}(u_{m1} - U_{12}\xi_2)\}^\top U_{11}\{\xi_1 - U_{11}^{-1}(u_{m1} - U_{12}\xi_2)\} \\
&\quad + u_{m1}^\top U_{11}^{-1} U_{12}\xi_2 - \frac{1}{2}\xi_2^\top (U_{22} - U_{12}^\top U_{11}^{-1} U_{12})\xi_2 + r_m \xi^\top \xi. \qquad \text{(S2)}
\end{aligned}$$

Because $\boldsymbol{\theta}_0$ is an interior point, when m is large, ξ_1 is free from any constraint. For any fixed ξ_2, taking maximum in (S2) with respect to ξ_1 leads to

$$\xi_1 = U_{11}^{-1}(u_{m1} - U_{12}\xi_2) + r_{m1}, \qquad \text{(S3)}$$

where $r_{m1} = o_p(1)$ uniformly. Putting this back in (S2), we have the profile log-likelihood

$$\ell_p(\pi_0 + m^{-1/2}\xi_2) = (u_{m2} - U_{21}U_{11}^{-1}u_{m1})\xi_2 - \frac{1}{2}(U_{22} - U_{12}^\top U_{11}^{-1} U_{12})\xi_2^2 + C + r_{m2}\xi_2^2,$$

where C does not depend on ξ_2 and $r_{m2} = o_p(1)$ uniformly.

Because $\pi \le \pi_0 = 1$, ξ_2 takes only non-positive values, and $\ell_p(\pi_0 + m^{-1/2}\xi_2)$ takes its maximum at

$$\widehat{\xi}_2 = (U_{22} - U_{12}^\top U_{11}^{-1} U_{12})^{-1}(u_{m2} - U_{21}U_{11}^{-1}u_{m1})_- + o_p(1),$$

where $x_- = \min\{x, 0\}$. Putting $\widehat{\xi}_2$ back into (S3) leads to an approximate of $\widehat{\xi}_1$, i.e.,

$$\widehat{\xi}_1 = U_{11}^{-1}\left\{u_{m1} - U_{12}(U_{22} - U_{12}^\top U_{11}^{-1} U_{12})^{-1}(u_{m2} - U_{21}U_{11}^{-1}u_{m1})_-\right\} + o_p(1).$$

The fact $u_m \xrightarrow{d} N(0, U)$ implies that $u_{m1} \xrightarrow{d} N(0, U_{11})$ and

$$u_{m2} - U_{21}U_{11}^{-1}u_{m1} \xrightarrow{d} N(0, U_{22} - U_{12}^\top U_{11}^{-1} U_{12}),$$

and that they are asymptotically independent. This immediately implies result (a) of Theorem 3.

It follows from the approximates of the MLEs and (S2) that

$$\begin{aligned} R_2(\pi_0) &= \widehat{\xi}_2^\top (U_{22} - U_{12}^\top U_{11}^{-1} U_{12}) \widehat{\xi}_2 + o_p(1) \\ &= \frac{(u_{m2} - U_{21} U_{11}^{-1} u_{m1})_-^2}{U_{22} - U_{12}^\top U_{11}^{-1} U_{12}} + o_p(1) \\ &\xrightarrow{d} \frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2, \end{aligned}$$

and

$$\begin{aligned} R_1(\boldsymbol{\theta}_0, \pi_0) &= u_{m1}^\top U_{11}^{-1} u_{m1} + \frac{(u_{m2} - U_{21} U_{11}^{-1} u_{m1})_-^2}{U_{22} - U_{12}^\top U_{11}^{-1} U_{12}} + o_p(1) \\ &\xrightarrow{d} \frac{1}{2}\chi_{q_{\boldsymbol{\theta}}}^2 + \frac{1}{2}\chi_{q_{\boldsymbol{\theta}}+1}^2. \end{aligned}$$

This proves Theorem 3. It is similar to prove the circumstance if $\pi_0 = 0$. □

Web Appendix D. Conditions and properties for deconvolution

Suppose the incubation period I follows a Gamma distribution $\Gamma(\alpha, \beta)$ with pdf f_I, then the characteristic function (chf) of I satisfies

$$\phi_I(t) = \left(1 - \frac{it}{\beta}\right)^{-\alpha} \quad \text{and} \quad |\phi_I(t)|^2 = \left(\frac{\beta^2}{\beta^2 + t^2}\right)^{\alpha}.$$

According to Liu and Taylor (1989) and Devroye (1989), under the following conditions, the estimator for f_G in (13) at its interior point is consistent:
(C1). $\phi_S(t)/|\phi_I(t)|^2$ is absolutely integrable.
(C2). The second order derivative of $f_G(y)$ exists and is continuous on $[0, +\infty)$.
(C3). The chf of I $\phi_I(t) \neq 0$ for almost all t.
(C4). The kernel chf $\phi_K(t)$ vanishes at $|t| > M$ for some M.
(C5). $M_n \to \infty$ and $h_n \to 0$ as $n \to \infty$.

The aim of the kernel $K(\cdot)$ is to smooth the empirical chf of S into an integrable function. It is known that the best kernels are those whose chfs are the flattest near the origin (Davis, 1975, 1977). The bias of $\widehat{f}_G(y)$ is

$$\text{bias}\{\widehat{f}_G(y)\} = \frac{1}{2}h^2 f_G(y) \int_{-\infty}^{y/h_n} t^2 K_c(t)dt + o(h_n^2),$$

where $K_c(t) = a_0 K(t) + a_1 K'(t)$ and

$$\begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} \int_{-\infty}^{y/h_n} K(t)dt & \int_{-\infty}^{y/h_n} K'(t)dt \\ \int_{-\infty}^{y/h_n} tK(t)dt & \int_{-\infty}^{y/h_n} tK'(t)dt \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The variance of $\widehat{f}_G(y)$ is given in Karunamuni (2009).

An alternative to nonparametric density estimation is the parametric approach. To accord with the generation mechanism of serial interval, it is more reasonable to post a parametric model on generation time rather than serial interval theoretically. A potential weakness for directly modeling the serial interval is model misspecification. An additional condition should be satisfied to make $\widehat{\phi}_G$ a proper chf:

(C6). $|\widehat{\phi}_S(t)|/|\phi_I(t)|^2 \leq 1$.

This condition requires that the norm of estimated chf for S be declining fast near the origin and is too strong for modeling observed data. For example, normal S and Gamma I may not satisfy (C6).

Web Appendix E. The goodness-of-fit test for incubation period modeling

The goodness-of-fit test is performed as follows. Divide the non-negative real line into 17 parts: $[k-1.5, k-0.5)$ for $k = 1, \ldots, 16$, and $[15.5, +\infty)$. The goodness-of-fit χ^2 statistic is

$$X^2 = \sum_{k=1}^{17} \frac{(O_k - E_k)^2}{E_k},$$

where E_k and O_k are the expected and observed number of cases in the kth interval:

$$E_k = m[\widehat{\pi}F_I(k-0.5;\widehat{\boldsymbol{\theta}}) + (1-\widehat{\pi})H(k-0.5;\widehat{\boldsymbol{\theta}})] - m[\widehat{\pi}F_I(k-1.5;\widehat{\boldsymbol{\theta}}) + (1-\widehat{\pi})H(k-1.5;\widehat{\boldsymbol{\theta}})].$$

The degree of freedom of X^2 is $17 - 3 - 1 = 13$, since there are three parameters in total. The 0.95 quantile of chi-squared distribution with 13 degrees of freedom is 22.36.

References

Booth, J. G. and Hobert, J. P. (1999). Maximizing generalized linear mixed model likelihoods with an automated monte carlo em algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **61,** 265–285.

Davis, K. B. (1975). Mean square error properties of density estimates. *The Annals of Statistics* pages 1025–1030.

Davis, K. B. (1977). Mean integrated square error properties of density estimates. *The Annals of Statistics* pages 530–535.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* **39,** 1–22.

Devroye, L. (1989). Consistent deconvolution in density estimation. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique* pages 235–239.

Karunamuni, R. J. (2009). Deconvolution boundary kernel method in nonparametric density estimation. *Journal of Statistical Planning & Inference* **139,** 2269–2283.

Liu, M. C. and Taylor, R. L. (1989). A consistent nonparametric density estimator for the deconvolution problem. *Canadian Journal of Statistics* **17,** 427–438.