Supplement to “Augmenting the Unreturned for Field Data with Information on Returned Failures Only”

Zhi-Sheng YE and Loon-Ching TANG

Department of Industrial & Systems Engineering

National University of Singapore, Singapore, 117576

In this supplement, we provide a proof to Theorem 1 in the main article. In addition, the asymptotic convergence of the chain \( \{\theta^{(m)}\}_{m \in \mathbb{N}} \) to an AR(1) process is heuristically justified. The notations in the main article carry over to this supplement.

1 Proof of Theorem 1

An observed pair \((x, t)\) should satisfy the observation condition that \(\tau_s < x + t < \tau_e\). Let \(f_{(X,T)}(x, t), \tau_0 \leq x \leq \tau_e, t > 0\), be the joint PDF of \(X\) and \(T\). Then \(f_{(X,T)}(x, t) = f_X(x)f_T(t)\) by the independence assumption. The joint PDF of \((X_{obs}, T_{obs})\) is given by

\[
f_{(X_{obs}, T_{obs})}(x, t) = \frac{f_{(X,T)}(x, t)}{\int_{0}^{\tau_e} P(\tau_s - x < T < \tau_e - x)f_X(x)dx} I_{[\tau_s, \tau_e]}(x + t)
= \frac{f_T(t)f_X(x)}{P(\tau_s - X < T < \tau_e - X)} I_{[\tau_s, \tau_e]}(x + t).
\]

The marginal distribution of \(X_{obs}\) can be obtained by integrating \(t\) out as

\[
f_{X_{obs}}(x) = \int_{\tau_s-x}^{\tau_e-x} f_{(X_{obs}, T_{obs})}(x, t)dt = \frac{P(\tau_s - x < T < \tau_e - x)}{P(\tau_s - X < T < \tau_e - X)f_X(x)} f_X(x), \tau_0 \leq x \leq \tau_e.
\]
On the other hand, based on the fact that $J$ is a geometric random variable conditional on $X_{\text{obs}}$, we have

$$E[J + 1|X_{\text{obs}} = x] = \frac{1}{P(\tau_s - x \leq T \leq \tau_e - x)},$$

(S.2)

and

$$E\{E[J + 1|X_{\text{obs}}]\} = \int_0^{\tau_e} \int_{\tau_s - x}^{\tau_e - x} [P(\tau_s - x \leq T \leq \tau_e - x)]^{-1} f(x, t) dt dx$$

$$= \int_0^{\tau_e} \int_{\tau_s - x}^{\tau_e - x} [P(\tau_s - x \leq T \leq \tau_e - x)]^{-1} f_T(t) f_X(x) dt dx$$

$$= \frac{1}{P(\tau_s - X < T < \tau_e - X)}.$$  

(S.3)

Combining Equations (S.1), (S.2) and (S.3) yields Equation (5) in the main article.

2 Convergence of the algorithm

Let $D_{\text{aug}}^{(m)}$ be the augmented data based on $\theta^{(m)}$. The log-likelihood of the augmented data, denoted as $Q(\theta|D, D^{(m)}, \theta^{(m)})$ and abbreviated as $Q(\theta|D, D^{(m)})$ if no confusion, is given by

$$Q(\theta|D, D^{(m)}) = \sum_{i=1}^{n} \sum_{j=0}^{J_i} \ln f_T(t_{i,j}^{(m)}) = \sum_{i=1}^{n} \left[ \ln p_i + \sum_{j=1}^{J_i} \ln g_{T_{\text{aug}}|X_{\text{obs}}}(t_{i,j}^{(m)}|x_i) \right] + \sum_{i=1}^{n} \ln \frac{f_{T_{\text{obs}}|X_{\text{obs}}}(t_i)}{p_i},$$

where $p_i = F_T(\tau_e - x_i) - F_T(\tau_s - x_i)$, and $g_{T_{\text{aug}}|X_{\text{obs}}}(\cdot|\cdot)$ and $f_{T_{\text{obs}}|X_{\text{obs}}}(\cdot|\cdot)$ are given by Equations (7) and (4) in the main text, respectively. The first summation on the right-hand side is the conditional log-likelihood of the observed lifetimes, while the second term is the summation of the conditional log-likelihood of $(J_i^{(m)}, t_{i,1}^{(m)}, \ldots, t_{i,J_i}^{(m)})$ conditional on $x_i$. Therefore, $Q(\theta|D, D^{(m)})$ can be broken down as:

$$Q(\theta|D, D^{(m)}) = Q_{\text{aug}}(\theta|D, D^{(m)}) + Q_{\text{obs}}(\theta|D).$$
Let $\hat{\theta}_n$ be the argument of the maximum of $Q_{\text{obs}}(\theta|D)$, which is the MLE for the observed data $D$. We assume that $\hat{\theta}_n$ is a $\sqrt{n}$-consistent estimator of $\theta_0$. The first and second derivatives of $Q(\theta|D,D^{(m)})$ with respect to $\theta$ are

$$\frac{\partial}{\partial \theta} Q(\theta|D,D^{(m)}) = \frac{\partial}{\partial \theta} Q_{\text{aug}}(\theta|D,D^{(m)}) + \frac{\partial}{\partial \theta} Q_{\text{obs}}(\theta|D) \equiv Z_1(\theta|D,D^{(m)}) + Z_2(\theta|D),$$

and

$$\frac{\partial^2 Q}{\partial \theta^2} = \frac{\partial^2}{\partial \theta^2} Q_{\text{aug}}(\theta|D,D^{(m)}) + \frac{\partial^2}{\partial \theta^2} Q_{\text{obs}}(\theta|D) \equiv U_1(\theta|D,D^{(m)}) + U_2(\theta|D).$$

Assume that the chain $\{\theta^{(m)}\}_{m \in \mathbb{N}}$ is ergodic so that a stationary distribution exists. In order to find out the stationary distribution, we assume that a stationary point $\theta^{(m)}$ satisfies the local parameter assumption $\sqrt{n}[\theta^{(m)} - \hat{\theta}_n] = o_p(1)$. Further assume that the third derivatives of $Q$ evaluated at $\hat{\theta}_n$ are $O_p(n)$. This assumption is valid if the third derivatives are bounded by $nh(\cdot)$ in a neighbourhood of $\theta_0$, where $h(\cdot)$ is an integrable function. We can expand $\frac{\partial}{\partial \theta} Q(\cdot|D,D^{(m)})$ around $\hat{\theta}_n$ and simplify it by noting that $\frac{\partial}{\partial \theta} Q(\theta^{(m+1)}|D,D^{(m)}) = 0$ and $Z_2(\hat{\theta}_n|D) = 0$:

$$0 = n^{-1/2} Z_1(\hat{\theta}_n|D,D^{(m)}) + \left[ n^{-1} U_1(\hat{\theta}_n|D,D^{(m)}) + n^{-1} U_2(\hat{\theta}_n|D) \right] n^{1/2}[\theta^{(m+1)} - \hat{\theta}_n] + o_p(1),$$

where $n^{-1} U_2(\hat{\theta}_n|D)$ is the observed information matrix for an observation, and $n^{-1} U_2(\hat{\theta}_n|D) \rightarrow_p I(\theta_0)$ when $n \rightarrow \infty$, $I(\theta_0)$ being the Fisher information matrix for the observed data $D$. By assuming that $I(\theta_0)$ is non-singular and noting that $[n^{-1} U_1(\hat{\theta}_n|D,D^{(m)}) + n^{-1} U_2(\hat{\theta}_n|D)]^{-1} = O_p(1)$, we can rearrange the above display as

$$\sqrt{n}[\theta^{(m+1)} - \hat{\theta}_n] = - \left[ n^{-1} U_1(\hat{\theta}_n|D,D^{(m)}) + n^{-1} U_2(\hat{\theta}_n|D) \right]^{-1} n^{-1/2} Z_1(\hat{\theta}_n|D,D^{(m)}) + o_p(1).$$

(S.4)
Further expand $n^{-1/2}Z_1(\cdot|D, D^{(m)})$ around $\theta^{(m)}$:

$$n^{-1/2}Z_1(\hat{\theta}_n|D, D^{(m)}) = n^{-1/2}Z_1(\theta^{(m)}|D, D^{(m)}) - [n^{-1}U_1(\theta^{(m)}|D, D^{(m)})] \cdot \sqrt{n}[\theta^{(m)} - \hat{\theta}_n] + o_p(1).$$

(S.5)

According to the central limit theorem, $n^{-1/2}Z_1(\theta^{(m)}|D, D^{(m)})$ is asymptotically normal with mean 0 and variance $n^{-1}U_1(\theta^{(m)}|D, D^{(m)})$, where $n^{-1}U_1(\theta^{(m)}|D, D^{(m)}) \to I_{aug}(\theta^{(m)}|D)$ and $I_{aug}(\theta^{(m)}|D)$ is the Fisher information matrix for the augmented data with the true parameter being $\theta^{(m)}$. Therefore,

$$\sqrt{n}[\theta^{(m+1)} - \hat{\theta}_n] = A_m \cdot \sqrt{n}[\theta^{(m)} - \hat{\theta}_n] + \epsilon_m + o_p(1),$$

where

$$A_m = [U_1(\hat{\theta}_n|D, D^{(m)}) + U_2(\hat{\theta}_n|D)]^{-1} U_1(\theta^{(m)}|D, D^{(m)}),$$

and $\epsilon_m$ is normal with mean 0 and variance matrix

$$n [U_1(\hat{\theta}_n|D, D^{(m)}) + U_2(\hat{\theta}_n|D)]^{-1} U_1(\theta^{(m)}|D, D^{(m)}) \cdot [U_1(\hat{\theta}_n|D, D^{(m)}) + U_2(\hat{\theta}_n|D)]^{-1}.$$
stationary distribution has been found for the chain. Under mild conditions, this stationary
distribution is unique.

Another heuristic justification of the convergence is to borrow the asymptotic arguments
used in the convergence proof of the Monte Carlo EM algorithm (Chan and Ledolter 1995).
All the above asymptotic arguments can be made rigid by deriving the joint distribution of
\[ \theta^{(m+1)} - \theta^{(m)} | \theta^{(m)} = \hat{\theta}_n \] and \[ Q_{aug}(\theta^{(m)} | D, D^{(m)}) - Q_{aug}(\hat{\theta}_n | D, D^{(m)}) | D \], and then passing
to the marginal distribution of \[ \theta^{(m+1)} - \hat{\theta}_n \] by Le Cam’s third lemma (Nielsen 2000).

References
