Sibyl: Epidemic tracing via Belief Propagation

The Sibyl Team

I. A SEMI-CONTINUOUS INFECTION MODEL

Each pair i, j of individuals will be in mutual contact in a finite set of instants $X_{ij} \subset \mathbb{R}_\infty = \mathbb{R} \cup \{+\infty\}$. For reasons that will become clear in the following, we will always assume $\infty \in X_{ij}$. As time advances, instantaneous contagion will happen with probability λ at time $s_{ij} \in X_{ij}$ if i is infected and j is susceptible. We will assume $\lambda = \lambda_{ij,}^{s_{ij},t_i}$ to possibly depend both on the specific (absolute) contact time s_{ij}, on the direction of the contact and on the time t_i of infection of individual i. Individual i can thus become infected in one instant in the set $X_i = \cup_{j \in \partial i} X_{ij}$. We will denote by $t_i \in X_i, r_i \in \mathbb{R}$ respectively the times of infection and recovery of individual i, with $t_i = \infty$ (resp. $r_i = \infty$) if the individual did not become infected (resp. recovered) within the time-frame.

We will assume the recovery delay $r_i - t_i$ of node i to be distributed with a continuous distribution with pdf $p_{R,i}(r_i - t_i)$. We will assume a set of factorized, site-dependent observations $p_{O,i}(\mathcal{O}_i | t_i, r_i)$. Model parameters will be hidden for the moment inside functions $p_{R,i}$, $\lambda_{ij}^{s,t}$, and $p_{O,i}$ and we will include them only later to avoid cluttering the notation.

A. Example: the SIR model

The standard SIR model can be obtained with $p_{R,i}(r_i - t_i) = \mu e^{-\mu(r_i - t_i)}$, $\lambda_{ij}^{s,t} \equiv \lambda$. In this setup, the model is memory-less (Markov) on the state of infection variables $x_i^t \in \{S, I, R\}$.

II. DISTRIBUTION OF TRANSMISSION, INFECTION AND RECOVERY TIMES

In the following we will always assume that $t_i \in X_i$ and $s_{ij}, s_{ji} \in X_{ij}$. Given the times of infection and recovery t_i and r_i, the transmission delay s_{ij} has "truncated" generalized geometric distribution

$$S_{ij}\left(s_{ij}|t_i,r_i\right) = \mathbb{I}\left[t_i < s_{ij} < r_i\right] \lambda_{ij}^{s_{ij},t_i} \prod_{t_i<s<s_{ij}} \left(1-\lambda_{ij}^{s,t_i}\right) + \mathbb{I}\left[s_{ij}=\infty\right] \prod_{s\geq r_i} \left(1-\lambda_{ij}^{s,t_i}\right)$$

because i will be infectious in the open time interval (t_i, r_i) and will possibly (i.e. if j is susceptible at the time) transmit the disease at some time s_{ij} in that interval if he did not transmit it before. If time r_i arrives and he did not transmit the disease, the individual will recover and never transmit the disease through that link; then transmission time on that link will be nominally $s_{ij} = \infty$. Special attention must be taken for auto-infections (otherwise, no infection can enter into the closed system). By adding a contact with an additional extra always-infected virtual neighbor i^{t*} to node i at time instant $t \in X_i$ with probability $p_t\left(s_{i^{t*}i} = t\right) = \gamma_i^t$ (typically small), $p_t\left(s_{i^{t*}i} = \infty\right) = 1-\gamma_i^t$, i will spontaneously self-infect at time t with probability γ_i^t (if it is susceptible at that time[1]). Let us define $A_i\left(\boldsymbol{s}_{i^*}\right) = A_i\left(\{s_{i^{t*}i}\}_{t\in X_i}\right) = \prod_{t\in X_i\setminus\{\infty\}} p_t\left(s_{i^{t*}i}\right)$. For convenience, let us define $\partial^* i$ the enlarged neighborhood of i including all extra nodes $\left\{i^{t*}\right\}_{t\in X_i}$. Given $\{s_{ki}\}$ for $k \in \partial^* i$, the infection time t_i satisfies, in a deterministic way:

$$t_i = \min_{k\in\partial^* i} s_{ki} \tag{1}$$

We can now write the joint pdf of discrete variables $\boldsymbol{t}, \boldsymbol{s}$ and continuous variables $\boldsymbol{r}$ as

$$p\left(\boldsymbol{t},\boldsymbol{r},\boldsymbol{s}\right) \propto \prod_i \delta\big(t_i, \min_{k\in\partial^* i} s_{ki}\big) A_i\left(\boldsymbol{s}_{i^*}\right) R_i\left(r_i - t_i\right) \prod_{(ij)} S_{ij}\left(s_{ij}|t_i,r_i\right) \tag{2}$$

so then as $p\left(\boldsymbol{t},\boldsymbol{r},\boldsymbol{s}|\mathcal{O}\right) \propto p\left(\mathcal{O}|\boldsymbol{t},\boldsymbol{r}\right) p\left(\boldsymbol{t},\boldsymbol{r},\boldsymbol{s}\right)$, we get

$$p\left(\boldsymbol{t},\boldsymbol{r},\boldsymbol{s}|\mathcal{O}\right) \propto \prod_i \delta\big(t_i, \min_{k\in\partial^* i} s_{ki}\big) A_i\left(\boldsymbol{s}_{i^*}\right) R_i\left(r_i - t_i\right) p_{O,i}\left(\mathcal{O}_i|t_i,r_i\right) \prod_{(ij)} S_{ij}\left(s_{ij}|t_i,r_i\right) \tag{3}$$

A note about r_is: all terms in (3) except R_i are constant as functions of r_i in any interval $(\hat{r}_i, \hat{r}_i')$ of consecutive times in X_i. We will exploit this fact and write $r_i = \hat{r}_i + u_i$,

where $\hat{r}_i = \max\{r \in X_i : r < r_i\}$. Integrating away the u_i variables in (3), we obtain an all-discrete model for variables $\boldsymbol{t}, \boldsymbol{s}, \hat{\boldsymbol{r}}$, with an identical expression to (3) but in which r_i has been replaced by $\hat{r}_i$ and $R_i(r_i - t_i)$ by $\hat{R}_i(\hat{r}_i - t_i) = \int_{\hat{r}_i - t_i}^{\hat{r}'_i - t_i} p_{R,i}(u)\, du$. For simplicity of notation, we will drop the ˆ symbols in the following.

III. BELIEF PROPAGATION EQUATIONS

A naive interpretation of (3) as a graphical model would introduces many unneeded short cycles that were not present in the original contact network. For example, pairs $(t_i, s_{ji}), (t_i, s_{ij}), (t_j, s_{ij}), (t_j, s_{ji})$ share respectively factors with indices $i, (ij), j, (ji)$, effectively forming a small cycle. A simple solution consists in regrouping factors as in (4) and considering (s_{ij}, s_{ji}) as a single variable:

$$\Psi_i\left(t_i, r_i, \{s_{ki}, s_{ik}\}_{k\in\partial i}\right) \;=\; \delta\big(t_i, \min_{k\in\partial^* i} s_{ki}\big) A_i(\boldsymbol{s}_{i^*})\, R_i(r_i - t_i) \prod_{j\in\partial i} S_{ij}(s_{ij}|t_i, r_i) \tag{4}$$

which results in a factor graph for (3) in which variables (s_{ij}, s_{ji}) have degree two and live in the middle of the original edges, and vars t_i, r_i have degree 1, i.e. a topology that closely follows the one of the original contact network:

$$p(\boldsymbol{t}, \boldsymbol{r}, \boldsymbol{s}|\mathcal{O}) \;=\; \frac{1}{Z}\prod_i \Psi_i \tag{5}$$

The corresponding BP equations for Ψ_i are

$$\begin{aligned} m_{ij}(s_{ij}, s_{ji}) \;\propto\; & \sum_{t_i}\sum_{r_i} p_{O,i}(\mathcal{O}_i|t_i, r_i)\, R_i(r_i - t_i)\, S_{ij}(s_{ij}|t_i, r_i) \times \qquad (6)\\ & \times \sum_{\{s_{ki}, s_{ik}\}} A_i(\boldsymbol{s}_{i^*})\, \delta\big(t_i, \min_{k\in\partial^* i} s_{ki}\big) \prod_{k\in\partial^* i\setminus j} S_{ik}(s_{ik}|t_i, r_i)\, m_{ki}(s_{ki}, s_{ik}) \end{aligned}$$

and marginals for t_i are

$$\begin{aligned} b_i(t_i) \;\propto\; & \sum_{r_i} p_{O,i}(\mathcal{O}_i|t_i, r_i)\, R_i(r_i - t_i)\, S_{ij}(s_{ij}|t_i, r_i) \times \qquad (7)\\ & \times \sum_{\{s_{ki}, s_{ik}\}} A_i(\boldsymbol{s}_{i^*})\, \delta\big(t_i, \min_{k\in\partial^* i} s_{ki}\big) \prod_{k\in\partial^* i} S_{ik}(s_{ik}|t_i, r_i)\, m_{ki}(s_{ki}, s_{ik}) \qquad (8) \end{aligned}$$

and similarly for r_i. A more efficient computation of the equations can be achieved by defining:

$$G_k^0(t_i, r_i) = \sum_{\substack{s_{ki} \geq t_i \\ s_{ik} > t_i}} S_{ik}(s_{ik}|t_i, r_i)\, m_{ki}(s_{ki}, s_{ik})$$

$$G_k^1(t_i, r_i) = \sum_{\substack{s_{ki} > t_i \\ s_{ik} > t_i}} S_{ik}(s_{ik}|t_i, r_i)\, m_{ki}(s_{ki}, s_{ik})$$

and substituting the extra neighbor message

$$m_{i^{*t}i}(s_{i^{*t}i}, s_{ii^{*t}}) = \begin{cases} \gamma_i^t & s_{i^{*t}i} = t, s_{ii^{*t}} = \infty \\ 1 - \gamma_i^t & s_{i^{*t}i} = \infty, s_{ii^{*t}} = \infty \end{cases}$$

Equations (6), (7) can be then rewritten as

$$\begin{aligned} m_{ij}(s_{ij}, s_{ji}) \propto & \sum_{t_i} \sum_{r_i} p_{O,i}(\mathcal{O}_i|t_i, r_i)\, R_i(r_i - t_i)\, S_{ij}(s_{ij}|t_i, r_i) \times \qquad (9) \\ & \times \sum_{\{s_{ki}, s_{ik}\}} A_i(\boldsymbol{s}_{i^*}) \left(\prod_{j \in \partial^* i} \mathbb{I}[s_{ki} \geq t_i] - \prod_{j \in \partial^* i} \mathbb{I}[s_{ki} > t_i] \right) \prod_{k \in \partial^* i \setminus j} S_{ik}(s_{ik}|t_i, r_i)\, m_{ki}(s_{ki}, s_{ik}) \\ \propto & \sum_{t_i < s_{ji}} \sum_{r_i \geq t_i} p_{O,i}(\mathcal{O}_i|t_i, r_i)\, R_i(r_i - t_i)\, S_{ij}(s_{ij}|t_i, r_i) \times \qquad (10) \\ & \times \prod_{t < t_i} (1 - \gamma_i^t) \left\{ \prod_{k \in \partial i \setminus j} G_k^0(t_i, r_i) - (1 - \gamma_i^{t_i}) \prod_{k \in \partial i \setminus j} G_k^1(t_i, r_i) \right\} \\ & + \sum_{r_i \geq s_{ji}} p_{O,i}(\mathcal{O}_i|s_{ji}, r_i)\, R_i(r_i - s_{ji}) \prod_{k \in \partial i \setminus j} G_k^0(s_{ji}, r_i) \\ b_i(t_i) \propto & \sum_{r_i} p_{R,i}(r_i - t_i)\, p_{O,i}(\mathcal{O}_i|t_i, r_i) \times \qquad (11) \\ & \prod_{t < t_i} (1 - \gamma_i^t) \left\{ \prod_{k \in \partial i} G_k^0(t_i, r_i) - (1 - \gamma_i^{t_i}) \prod_{k \in \partial i} G_k^1(t_i, r_i) \right\} \\ b_i(r_i) \propto & \sum_{t_i} p_{R,i}(r_i - t_i)\, p_{O,i}(\mathcal{O}_i|t_i, r_i) \times \qquad (12) \\ & \prod_{t < t_i} (1 - \gamma_i^t) \left\{ \prod_{k \in \partial i} G_k^0(t_i, r_i) - (1 - \gamma_i^{t_i}) \prod_{k \in \partial i} G_k^1(t_i, r_i) \right\} \end{aligned}$$

Note that products $\prod_{k \in \partial i \setminus j} G_k$ can be computed simultaneously for $j \in \partial i$ in time $O(|\partial i|)$ (either by computing it as the fraction $G_j^{-1} \prod_{k \in \partial i} G_k$, or by first recursively computing $\prod_{\ell=1}^{\ell'} G_{k_\ell}$ and $\prod_{\ell=\ell'}^{|\partial i|} G_{k_\ell}$ for $\ell' = 1, \ldots, |\partial i|$ and then $\prod_{\ell \neq \ell'} G_{k_\ell} =$

$\prod_{\ell=1}^{\ell'-1} G_{k\ell} \prod_{\ell=\ell'+1}^{|\partial i|} G_{k\ell}$, a method that does not involve divisions and is thus numerically more stable). The resulting implementation of the update of all messages in factor Ψ_i has complexity $O\left(|X_i| \sum_{j\in\partial i} \left(|X_i| + |X_{ij}|^2\right)\right)$.

IV. LOG-LIKELIHOOD AND BETHE FREE ENTROPY

The Bethe Free entropy f_{Bethe} is an approximation of the log-likelihood $f = \log Z$ that can be derived as its exact expression on acyclic graphs. For a factor graph such (5), one expression of f_{Bethe} is

$$f_{Bethe} = \sum_i \log z_i - \sum_{(ij)} \log z_{ij} \tag{13}$$

$$= \sum_i \left\{ \log z_i - \frac{1}{2} \sum_{j\in\partial i} \log z_{ij} \right\} \tag{14}$$

where

$$z_{ij} = \sum_{s_{ij}, s_{ji}} m_{ij}\left(s_{ij}, s_{ji}\right) m_{ji}\left(s_{ji}, s_{ij}\right) \tag{15}$$

$$z_i = \sum_{t_i, t_i, \{s_{ik}, s_{ki}\}} \Psi_i\left(t_i, r_i, \{s_{ki}, s_{ik}\}_{k\in\partial i}\right) \tag{16}$$

$$= \sum_{t_i, r_i} p_{R,i}\left(r_i - t_i\right) p_{O,i}\left(\mathcal{O}_i | t_i, r_i\right) \times \tag{17}$$

$$\times \prod_{t<t_i} \left(1 - \gamma_i^t\right) \left\{ \prod_{k\in\partial i} G_k^0\left(t_i, r_i\right) - \left(1 - \gamma_i^{t_i}\right) \prod_{k\in\partial i} G_k^1\left(t_i, r_i\right) \right\}$$

V. APPROXIMATE ML FOR PARAMETER LEARNING

To learn parameters, one possibility is maximizing the log likelihood $\log Z = \log p\left(\boldsymbol{t}, \boldsymbol{r}, \boldsymbol{s} | \mathcal{O}, \theta\right)$. A standard approach is Expectation Maximization (EM). Consider the standard EM, for variables $\boldsymbol{x} = (\boldsymbol{t}, \boldsymbol{r}, \boldsymbol{s})$ and parameters θ, as an iterative

procedure for the parameter vector θ_k

$$\theta_{k+1} = \arg\max_{\theta} \sum_{\boldsymbol{x}} p\left(\boldsymbol{x}|\mathcal{O}, \theta_k\right) \log p\left(\boldsymbol{x}, \mathcal{O}|\theta\right) \tag{18}$$

$$= \arg\max_{\theta} E_{\theta_k}\left[\log p\left(\boldsymbol{x}, \mathcal{O}|\theta\right)\right] \tag{19}$$

where E_{θ_k} is expectation with respect to $p\left(\boldsymbol{x}|\mathcal{O}, \theta_k\right)$. We can replace it with a gradient-climbing iteration, i.e. an approximate SGD (ρ is small and positive):

$$\theta_{\tau+1} = \theta_\tau + \rho \sum_{X} p\left(\boldsymbol{x}|\mathcal{O}, \theta_\tau\right) \nabla_\theta \log p\left(\boldsymbol{x}, \mathcal{O}|\theta_\tau\right) \tag{20}$$

$$= \theta_\tau + \rho E_{\theta_\tau}\left[\nabla_\theta \log p\left(\boldsymbol{x}, \mathcal{O}|\theta_\tau\right)\right] \tag{21}$$

For our specific graphical model with factors $\{\Psi_i\}$, $\log p\left(\boldsymbol{x}, \mathcal{O}|\theta\right)$ can be written as a sum over individuals, $\log p\left(\boldsymbol{x}, O|\theta\right) = \sum_i \log \Psi_i\left(x_i, x_i, \{\boldsymbol{s}_i\}, \mathcal{O}_i|\theta, \boldsymbol{x}_{\partial i}\right)$ so

$$E_{\theta_\tau}\left[\nabla_\theta \log p\left(\boldsymbol{x}, O|\theta\right)\right] = \sum_i E_{\theta_\tau}\left[\nabla_\theta \log \Psi_i\left(x_i, O_i|\theta, \boldsymbol{x}_{\partial i}\right)\right] \tag{22}$$

now $f_i = E_{\theta_\tau}\left[\nabla_\theta \log \Psi_i\left(x_i, O_i|\theta, \boldsymbol{x}_{\partial i}\right)\right]$ can be easily computed locally when computing (7) (it reuses many of the computations). We point out that on a fixed point of the BP equations (9), the approximate SGD step is actually an *exact* gradient ascent step on the Bethe Free Entropy 13, as z_{ij} terms do not depend on θ and the total derivative of f_{Bethe} with respect to θ equals its partial derivative, because a fixed point of BP is a stationary point of f, i.e. $\nabla_m f = 0$. In principle, one could wait for convergence of BP for an approximate "E" step, then perform one gradient ascent update for the "M" step, etc, so as to reproduce an (approximate) EM. But in practice, if ρ is small enough, both iterations can be interleaved to get a single iterative procedure that simultaneously computes marginals and adjusts parameters. The equation for the parameters' update for the model in (5) with respect to θ are then

$$\theta_{\tau+1} = \theta_\tau + \rho \nabla_\theta f_{Bethe} \tag{23}$$

with $\nabla_\theta f_{Bethe} = \sum_i \nabla_\theta f_i$ and

$$\frac{\partial f_i}{\partial \theta'} = \frac{1}{z_i} \sum_{t_i, r_i} \frac{dp_{R,i}}{d\theta'} \left(r_i - t_i, \theta'\right) p\left(\mathcal{O}_i | t_i, r_i, \theta''\right) \times \tag{24}$$

$$\times \prod_{t<t_i} \left(1 - \gamma_i^t\right) \left\{ \prod_{k \in \partial i} G_k^0\left(t_i, t_i\right) - \left(1 - \gamma_i^{t_i}\right) \prod_{k \in \partial i} G_k^1\left(t_i, t_i\right) \right\}$$

$$\frac{\partial f_i}{\partial \theta''} = \frac{1}{z_i} \sum_{t_i, r_i} p_{R,i}\left(r_i - t_i, \theta'\right) \frac{d}{d\theta''} p\left(\mathcal{O}_i | t_i, r_i, \theta''\right) \times \tag{25}$$

$$\times \prod_{t<t_i} \left(1 - \gamma_i^t\right) \left\{ \prod_{k \in \partial i} G_k^0\left(t_i, t_i\right) - \left(1 - \gamma_i^{t_i}\right) \prod_{k \in \partial i} G_k^1\left(t_i, t_i\right) \right\}$$

$$\frac{\partial f_i}{\partial \theta'''} = \frac{1}{z_i} \sum_{t_i, r_i} p_{R,i}\left(r_i - t_i, \theta'\right) p\left(\mathcal{O}_i | t_i, r_i, \theta''\right) \prod_{t<t_i} \left(1 - \gamma_i^t\right) \times \tag{26}$$

$$\times \sum_{j \in \partial i} \left\{ \sum_{\substack{s_{ki} \geq t_i \\ s_{ik} > t_i}} \frac{dS_{ij}}{d\theta'''} \left(s_{ij} | t_i, r_i, \theta'''\right) m_{ji}\left(s_{ji}, s_{ij}\right) \prod_{k \in \partial i \setminus j} G_k^0\left(t_i, t_i\right) \right\} \tag{27}$$

A simultaneous fixed point of (23) and (9) is both a BP fixed point and a stationary point (a local maximum when following gradient updates) of the approximate log-likelihood. The computation of the gradients is of course distributed, but the update and distribution of the updated parameters needs to be done centrally or through some form of parallel reduction.

---

[1] A particularly interesting case is with $\gamma_i^0 = \gamma \to 0$ and $\gamma_i^t = 0$ for $t > 0$: in this case individuals can be self-infected only at time 0, representing a closed system with a single unknown seed at time $t = 0$.