

Bayesian multi-state models and software

for intermittently-observed individual data, motivated by
estimating infection duration distributions.

Chris Jackson



MRC
Biostatistics
Unit



UNIVERSITY OF
CAMBRIDGE

Brief introduction to multi-state models for intermittently-observed data

Introducing a new R package `msmbayes`, which makes these models Bayesian

Work in progress using these models, made easier with this software:

- ▶ using Bayesian methods to facilitate relaxing the Markov assumption
- ▶ estimating the duration of infection from intermittent testing data

Brief introduction to multi-state models for intermittently-observed data

Introducing a new R package [msmbayes](#), which makes these models **Bayesian**

Work in progress using these models, **made easier** with this software:

- ▶ using Bayesian methods to facilitate relaxing the Markov assumption
- ▶ estimating the duration of infection from intermittent testing data

Brief introduction to multi-state models for intermittently-observed data

Introducing a new R package `msmbayes`, which makes these models Bayesian

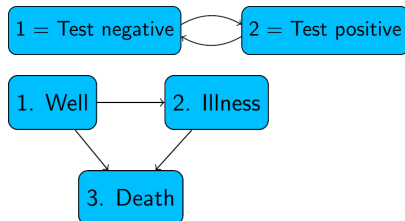
Work in progress using these models, made easier with this software:

- ▶ using Bayesian methods to facilitate relaxing the Markov assumption
- ▶ estimating the duration of infection from intermittent testing data

Multi-state process

Multi-state models represent how an individual can move between different states over time.

Typical examples:



They have been applied to many different kinds of data structures

- ▶ diverse modelling frameworks and literature

Intermittently observed, individual data

For a set of individuals, a state is observed, at a set of arbitrary times.

Person	Time	State
1	0	1
1	1.2	1
1	5	2
2	0.1	1
2	8	3
...

State in between the times is **not necessarily known** at any other times

“Intermittently-observed”, “panel”, “interval-censored” data.

Continuous-time multi-state models

Parameters: **transition intensities** $q_{rs}(t)$: rate of moving to s for people in r at time t .

With covariates \mathbf{x} , typically $q_{rs}(t; \mathbf{x}) = q_{rs}^{(0)}(t) \exp(\beta'_{rs} \mathbf{x})$

Given these, determine the **transition probabilities**

$$p_{rs}(u, t) = P(S(t+u) = s \mid S(t) = r, \dots)$$

for a state $S(t)$ over any time interval $t, t+u$

- ▶ hence compute the likelihood for intermittently-observed, individual data

Continuous-time multi-state models

Parameters: **transition intensities** $q_{rs}(t)$: rate of moving to s for people in r at time t .

With covariates \mathbf{x} , typically $q_{rs}(t; \mathbf{x}) = q_{rs}^{(0)}(t) \exp(\beta'_{rs} \mathbf{x})$

Given these, determine the **transition probabilities**

$$p_{rs}(u, t) = P(S(t+u) = s \mid S(t) = r, \dots)$$

for a state $S(t)$ over any time interval $t, t+u$

- ▶ hence compute the likelihood for intermittently-observed, individual data

Parameters: **transition intensities** $q_{rs}(t)$: rate of moving to s for people in r at time t .

With covariates \mathbf{x} , typically $q_{rs}(t; \mathbf{x}) = q_{rs}^{(0)}(t) \exp(\beta'_{rs} \mathbf{x})$

Given these, determine the **transition probabilities**

$$p_{rs}(u, t) = P(S(t+u) = s \mid S(t) = r, \dots)$$

for a state $S(t)$ over any time interval $t, t+u$

- hence compute the likelihood for intermittently-observed, individual data

Continuous-time multi-state models: properties

Transition probabilities can be calculated easily in practice (via solving Kolmogorov differential equation)

- ▶ if the intensities $q_{rs}(t)$ don't depend on previous states visited and transition times: **Markov assumption**
- ▶ however they can easily depend on the time t since the start of the process

Solution particularly simple if $q_{rs}(t)$ is **constant** with t

- ▶ **Sojourn time** (one period in state r) is $\sim \text{Exponential}(\sum_s q_{rs})$

(Simple generalisation if $q_{rs}(t)$ is **piecewise-constant**, different in different time periods)

Exponential sojourn times may be unrealistic. Hazards may depend on time spent in state

Continuous-time multi-state models: properties

Transition probabilities can be calculated easily in practice (via solving Kolmogorov differential equation)

- ▶ if the intensities $q_{rs}(t)$ don't depend on previous states visited and transition times: **Markov assumption**
- ▶ however they can easily depend on the time t since the start of the process

Solution particularly simple if $q_{rs}(t)$ is **constant** with t

- ▶ **Sojourn time** (one period in state r) is $\sim \text{Exponential}(\sum_s q_{rs})$

(Simple generalisation if $q_{rs}(t)$ is **piecewise-constant**, different in different time periods)

Exponential sojourn times may be unrealistic. Hazards may depend on time spent in state

Continuous-time multi-state models: properties

Transition probabilities can be calculated easily in practice (via solving Kolmogorov differential equation)

- ▶ if the intensities $q_{rs}(t)$ don't depend on previous states visited and transition times: **Markov assumption**
- ▶ however they can easily depend on the time t since the start of the process

Solution particularly simple if $q_{rs}(t)$ is **constant** with t

- ▶ **Sojourn time** (one period in state r) is $\sim \text{Exponential}(\sum_s q_{rs})$

(Simple generalisation if $q_{rs}(t)$ is **piecewise-constant**, different in different time periods)

Exponential sojourn times may be unrealistic. Hazards may depend on time spent in state

Continuous-time multi-state models: properties

Transition probabilities can be calculated easily in practice (via solving Kolmogorov differential equation)

- ▶ if the intensities $q_{rs}(t)$ don't depend on previous states visited and transition times: **Markov assumption**
- ▶ however they can easily depend on the time t since the start of the process

Solution particularly simple if $q_{rs}(t)$ is **constant** with t

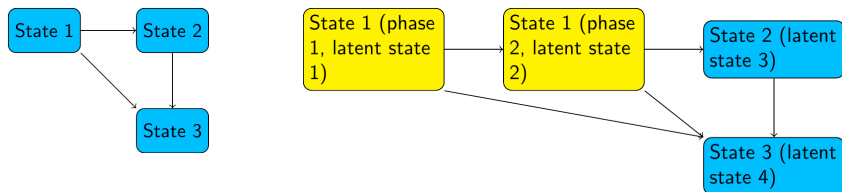
- ▶ **Sojourn time** (one period in state r) is $\sim \text{Exponential}(\sum_s q_{rs})$

(Simple generalisation if $q_{rs}(t)$ is **piecewise-constant**, different in different time periods)

Exponential sojourn times may be unrealistic. Hazards may depend on time spent in state

Phase-type models: relaxing the Markov assumption

Phase-type model (Titman and Sharples Biometrics 2012). Example:



Replace an observable state (state 1 in this picture) with a set of latent states (“phases”). Latent states follow a Markov model.

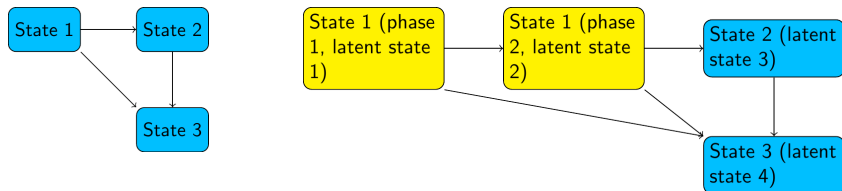
Hidden Markov model: hence likelihood calculation tractable

Semi-Markov model:

- ▶ Risk of transition out of state depends on time spent in it
- ▶ Observable state has a “phase-type” sojourn distribution, instead of the exponential.
- ▶ More latent phases and transitions → more flexible sojourn distribution (... more estimation challenges).

Phase-type models: relaxing the Markov assumption

Phase-type model (Titman and Sharples Biometrics 2012). Example:



Replace an observable state (state 1 in this picture) with a set of latent states (“phases”). Latent states follow a Markov model.

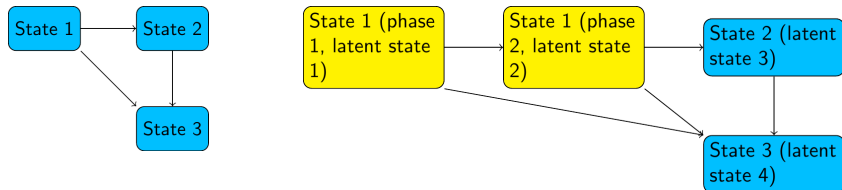
Hidden Markov model: hence likelihood calculation tractable

Semi-Markov model:

- ▶ Risk of transition out of state depends on time spent in it
- ▶ Observable state has a “phase-type” sojourn distribution, instead of the exponential.
- ▶ More latent phases and transitions → more flexible sojourn distribution (... more estimation challenges).

Phase-type models: relaxing the Markov assumption

Phase-type model (Titman and Sharples Biometrics 2012). Example:



Replace an observable state (state 1 in this picture) with a set of latent states (“phases”). Latent states follow a Markov model.

Hidden Markov model: hence likelihood calculation tractable

Semi-Markov model:

- ▶ Risk of transition out of state depends on time spent in it
- ▶ Observable state has a “phase-type” sojourn distribution, instead of the exponential.
- ▶ More latent phases and transitions → more flexible sojourn distribution (... more estimation challenges).

Software: msm R package

Fits multi-state models for intermittently-observed data, and variants, using maximum likelihood

- ▶ Stock R optimisation algorithms (“BFGS” by default)

Popular (>1000 citations) , but common problems reported:

- ▶ Convergence failed, or unreliable (“non positive-definite Hessian”).
- ▶ ... Evidence of weak identifiability, irregular likelihoods.
- ▶ Often due to users specifying over-parameterised models
 - ▶ Hard to relate the meaning of the “transition intensity” parameters to the (intermittently-observed) data
 - ▶ Confusion with discrete-time, where transition probabilities easily estimated by proportions transitioning in data
- ▶ Computational expense from evaluating likelihood

Software: msm R package

Fits multi-state models for intermittently-observed data, and variants, using maximum likelihood

- ▶ Stock R optimisation algorithms (“BFGS” by default)

Popular (>1000 citations) , but common problems reported:

- ▶ Convergence failed, or unreliable (“non positive-definite Hessian”).
- ▶ ... Evidence of weak identifiability, irregular likelihoods.
- ▶ Often due to users specifying over-parameterised models
 - ▶ Hard to relate the meaning of the “transition intensity” parameters to the (intermittently-observed) data
 - ▶ Confusion with discrete-time, where transition probabilities easily estimated by proportions transitioning in data
- ▶ Computational expense from evaluating likelihood

Software: msm R package

Fits multi-state models for intermittently-observed data, and variants, using maximum likelihood

- ▶ Stock R optimisation algorithms (“BFGS” by default)

Popular (>1000 citations) , but common problems reported:

- ▶ Convergence failed, or unreliable (“non positive-definite Hessian”).
- ▶ ... Evidence of weak identifiability, irregular likelihoods.
- ▶ Often due to users specifying over-parameterised models
 - ▶ Hard to relate the meaning of the “transition intensity” parameters to the (intermittently-observed) data
 - ▶ Confusion with discrete-time, where transition probabilities easily estimated by proportions transitioning in data
- ▶ Computational expense from evaluating likelihood

Software: msm R package

Fits multi-state models for intermittently-observed data, and variants, using maximum likelihood

- ▶ Stock R optimisation algorithms (“BFGS” by default)

Popular (>1000 citations) , but common problems reported:

- ▶ Convergence failed, or unreliable (“non positive-definite Hessian”).
- ▶ ... Evidence of weak identifiability, irregular likelihoods.
- ▶ Often due to users specifying over-parameterised models
 - ▶ Hard to relate the meaning of the “transition intensity” parameters to the (intermittently-observed) data
 - ▶ Confusion with discrete-time, where transition probabilities easily estimated by proportions transitioning in data
- ▶ Computational expense from evaluating likelihood

Software: msm R package

Fits multi-state models for intermittently-observed data, and variants, using maximum likelihood

- ▶ Stock R optimisation algorithms (“BFGS” by default)

Popular (>1000 citations) , but common problems reported:

- ▶ Convergence failed, or unreliable (“non positive-definite Hessian”).
- ▶ ... Evidence of weak identifiability, irregular likelihoods.
- ▶ Often due to users specifying over-parameterised models
 - ▶ Hard to relate the meaning of the “transition intensity” parameters to the (intermittently-observed) data
 - ▶ Confusion with discrete-time, where transition probabilities easily estimated by proportions transitioning in data
- ▶ Computational expense from evaluating likelihood

Software: `msm` R package

Fits multi-state models for intermittently-observed data, and variants, using maximum likelihood

- ▶ Stock R optimisation algorithms (“BFGS” by default)

Popular (>1000 citations) , but common problems reported:

- ▶ Convergence failed, or unreliable (“non positive-definite Hessian”).
- ▶ ... Evidence of weak identifiability, irregular likelihoods.
- ▶ Often due to users specifying over-parameterised models
 - ▶ Hard to relate the meaning of the “transition intensity” parameters to the (intermittently-observed) data
 - ▶ Confusion with discrete-time, where transition probabilities easily estimated by proportions transitioning in data
- ▶ Computational expense from evaluating likelihood

Aim: Develop a Bayesian equivalent to the msm package

Constrain parameters using prior information

Emphasise posterior and prior distributions, to communicate uncertainty

- ▶ making any weakness of information clearer to users

Exploit modern computational algorithms

- ▶ both correctness and speed desirable

A motivating application: infection positivity duration

Determining the **length of an infection** (defined as testing positive) based on individual, longitudinal, testing data



Times in infection (and non-infection) states typically **not exponentially-distributed**

Simulate data roughly inspired by respiratory infections (SIREN study of healthcare workers, COVID-19 Infection Survey)

- ▶ 3600 observations (36 monthly tests from 100 people), though will be more in practice...

msmbayes R package

<https://chjackson.github.io/msmbayes>.

- ▶ Uses Stan for Bayesian estimation, via the `cmdstanr` package.
- ▶ MCMC, variational inference or posterior mode optimisation / Laplace approximation
- ▶ `posterior` package and `rvar` data type were helpful

To fit a model, supply the data, a matrix indicating the allowed transitions, and the priors. Example syntax:

```
Q <- rbind(c(0, 1),  
           c(1, 0)) # 2-state transition structure  
priors <- list(  
  msmprior("time(1,2)", median=10, upper=30),  
  msmprior("time(2,1)", median=0.5, upper=1)  
)  
msmbayes(data = infsim, state="state", time="months",  
          subject="subject", qmatrix=q, priors=priors,...)
```

msmbayes R package

<https://chjackson.github.io/msmbayes>.

- ▶ Uses Stan for Bayesian estimation, via the `cmdstanr` package.
- ▶ MCMC, variational inference or posterior mode optimisation / Laplace approximation
- ▶ `posterior` package and `rvar` data type were helpful

To fit a model, supply the data, a matrix indicating the allowed transitions, and the priors. Example syntax:

```
Q <- rbind(c(0, 1),  
           c(1, 0)) # 2-state transition structure  
priors <- list(  
  msmprior("time(1,2)", median=10, upper=30),  
  msmprior("time(2,1)", median=0.5, upper=1)  
)  
msmbayes(data = infsim, state="state", time="months",  
          subject="subject", qmatrix=q, priors=priors,...)
```

Intuitive syntax to specify priors

Specify judgements for $1/q_{rs}$: the **mean time from state r to s** (given no competing risks)

Supply prior median and upper (or lower) 95% quantile, e.g.

- ▶ 10 months (up to 30 months) for mean time until next infection $1/q_{12}$
- ▶ 2 weeks (up to 1 month) for mean length of infection $1/q_{21}$
- ▶ log-normal prior for q_{rs} is deduced.

```
priors <- list(  
  msmprior("time(1,2)", median=10, upper=30),  
  msmprior("time(2,1)", median=0.5, upper=1)  
)
```

(**careful:** uncertainty about mean, not the between-individual distribution!)

Intuitive syntax to specify priors

Specify judgements for $1/q_{rs}$: the mean time from state r to s (given no competing risks)

Supply prior median and upper (or lower) 95% quantile, e.g.

- ▶ 10 months (up to 30 months) for mean time until next infection $1/q_{12}$
- ▶ 2 weeks (up to 1 month) for mean length of infection $1/q_{21}$
- ▶ log-normal prior for q_{rs} is deduced.

```
priors <- list(  
  msmprior("time(1,2)", median=10, upper=30),  
  msmprior("time(2,1)", median=0.5, upper=1)  
)
```

(careful: uncertainty about mean, not the between-individual distribution!)

Intuitive syntax to specify priors

Specify judgements for $1/q_{rs}$: the mean time from state r to s (given no competing risks)

Supply prior median and upper (or lower) 95% quantile, e.g.

- ▶ 10 months (up to 30 months) for mean time until next infection $1/q_{12}$
- ▶ 2 weeks (up to 1 month) for mean length of infection $1/q_{21}$
- ▶ log-normal prior for q_{rs} is deduced.

```
priors <- list(  
  msmprior("time(1,2)", median=10, upper=30),  
  msmprior("time(2,1)", median=0.5, upper=1)  
)
```

(careful: uncertainty about mean, not the between-individual distribution!)

Intuitive syntax to specify priors

Specify judgements for $1/q_{rs}$: the mean time from state r to s (given no competing risks)

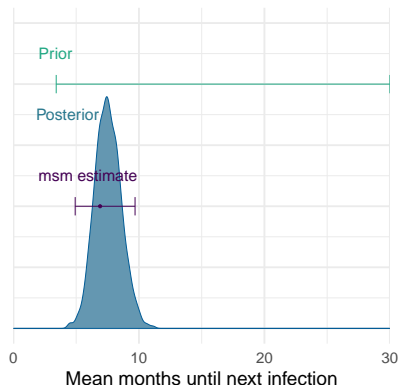
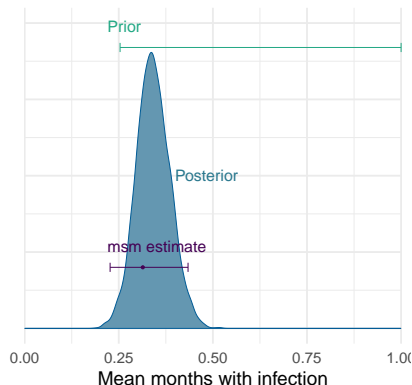
Supply prior median and upper (or lower) 95% quantile, e.g.

- ▶ 10 months (up to 30 months) for mean time until next infection $1/q_{12}$
- ▶ 2 weeks (up to 1 month) for mean length of infection $1/q_{21}$
- ▶ log-normal prior for q_{rs} is deduced.

```
priors <- list(  
  msmprior("time(1,2)", median=10, upper=30),  
  msmprior("time(2,1)", median=0.5, upper=1)  
)
```

(careful: uncertainty about mean, not the between-individual distribution!)

Outputs: two transition rates, no covariates.

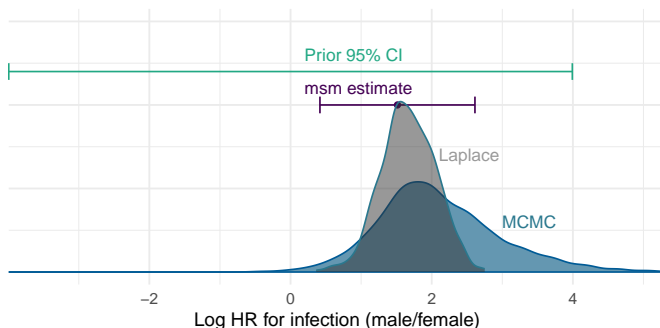


Posterior distribution from `msmbayes` agrees with maximum likelihood estimates from `msm`

Influence of prior apparent from moved posterior mode

Outputs: two transition rates, two covariates on each.

8 parameters, 3600 observations. Example parameter estimate:



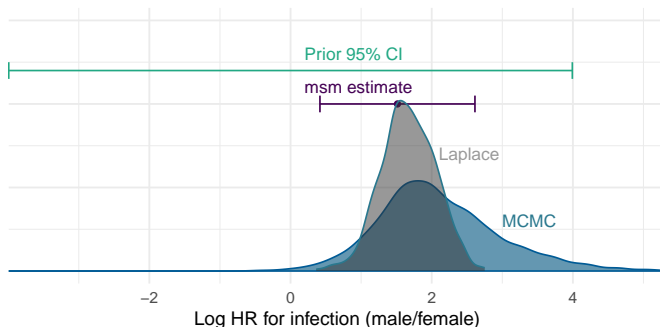
Speed / accuracy trade-off

1 hour for MCMC, which reveals skewed posterior

A few seconds to estimate posterior mode by optimisation (similar speed to msm). 1 min for Laplace approx around mode, which underestimates uncertainty

Outputs: two transition rates, two covariates on each.

8 parameters, 3600 observations. Example parameter estimate:



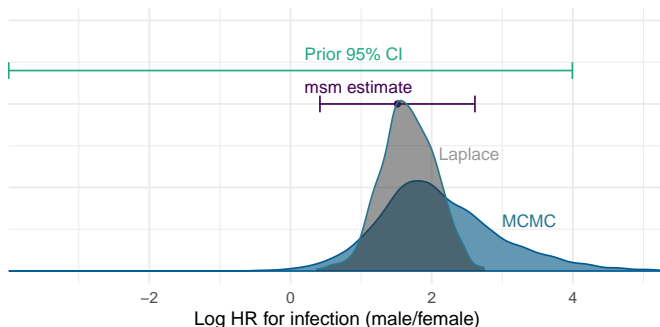
Speed / accuracy trade-off

1 hour for MCMC, which reveals skewed posterior

A few seconds to estimate posterior mode by optimisation (similar speed to msm). 1 min for Laplace approx around mode, which underestimates uncertainty

Outputs: two transition rates, two covariates on each.

8 parameters, 3600 observations. Example parameter estimate:

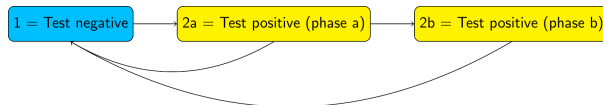


Speed / accuracy trade-off

1 hour for MCMC, which reveals skewed posterior

A few seconds to estimate posterior mode by optimisation (similar speed to msm). 1 min for Laplace approx around mode, which underestimates uncertainty

Fitting phase-type models



Simulate data with phase-type model for infection state:

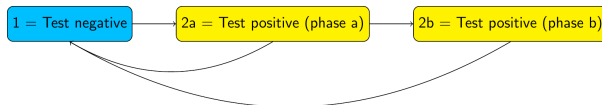
... Sojourn distribution is a 50/50 mixture of latent “short stayers” (exponential mean 4 days) and “long stayers” (sum of exponentials mean 4 + 30)

Maximum likelihood estimation of this model in `msm` **does not converge!**

Instead, fit model in `msmbayes`. Simply specify `nphase`: number of hidden phases per state.

```
draws <- msmbayes(infsim2, state="state", time="months",  
                  subject="subject", Q=Q, nphase=c(1, 2))
```

Fitting phase-type models



Simulate data with phase-type model for infection state:

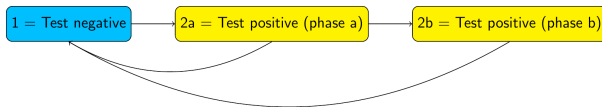
... Sojourn distribution is a 50/50 mixture of latent “short stayers” (exponential mean 4 days) and “long stayers” (sum of exponentials mean 4 + 30)

Maximum likelihood estimation of this model in `msm` **does not converge!**

Instead, fit model in `msmbayes`. Simply specify `nphase`: number of hidden phases per state.

```
draws <- msmbayes(infsim2, state="state", time="months",  
                  subject="subject", Q=Q, nphase=c(1, 2))
```

Fitting phase-type models



Simulate data with phase-type model for infection state:

... Sojourn distribution is a 50/50 mixture of latent “short stayers” (exponential mean 4 days) and “long stayers” (sum of exponentials mean 4 + 30)

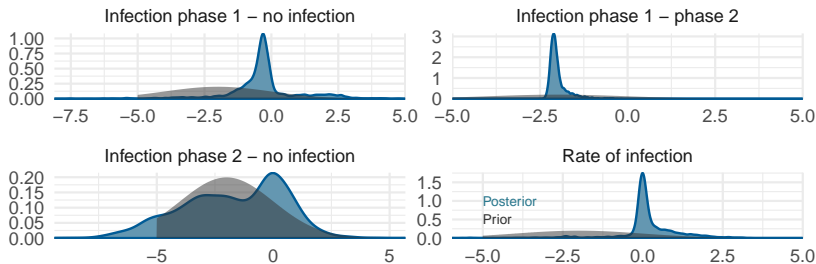
Maximum likelihood estimation of this model in `msm` **does not converge!**

Instead, fit model in `msmbayes`. Simply specify `nphase`: number of hidden phases per state.

```
draws <- msmbayes(infsim2, state="state", time="months",  
                  subject="subject", Q=Q, nphase=c(1, 2))
```

Priors + posteriors from fitted phase-type models

Log transition rates



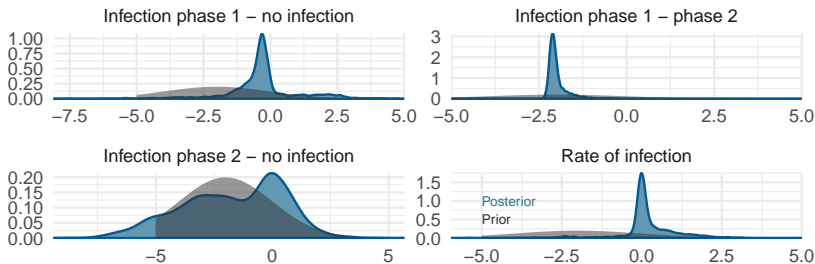
Transition rate from latent phase 2 is not identifiable

Mean sojourn time in infection state is identifiable, and more uncertain than under a basic Markov model

Phase-type model	Markov model
30 days (12, 37)	31 (27, 37)
Phase 1 21 (2, 34), Phase 2 43 (4, 254)	

Priors + posteriors from fitted phase-type models

Log transition rates

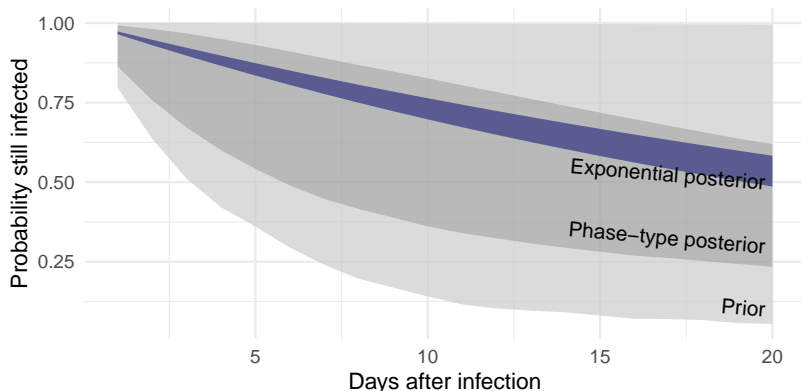


Transition rate from latent phase 2 is not identifiable

Mean sojourn time in infection state is identifiable, and more uncertain than under a basic Markov model

Phase-type model	Markov model
30 days (12, 37)	31 (27, 37)
Phase 1 21 (2, 34), Phase 2 43 (4, 254)	

Posterior of sojourn distribution from phase-type model

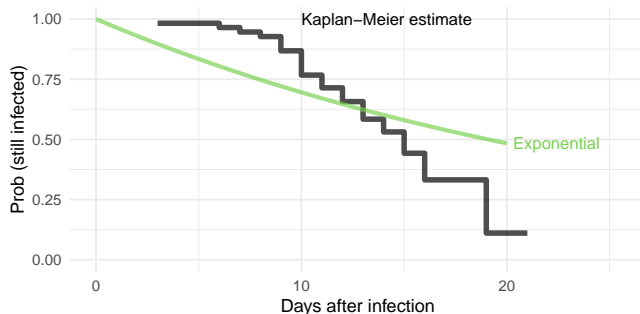


Very uncertain, but influenced by the weak prior for the unidentifiable transition rate

Further research needed on constructing meaningful priors in phase-type models

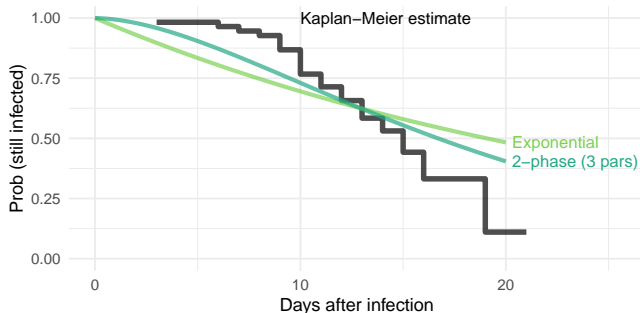
Do phase-type models fit infection duration data?

From ATACCC “Assessment of Transmission and Contagiousness of COVID-19 in Contacts” study, via daily PCR tests from 57 people



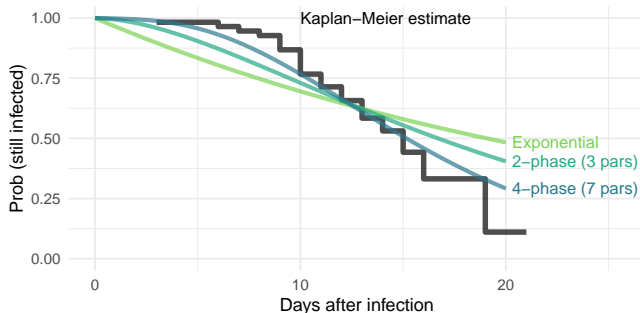
Do phase-type models fit infection duration data?

From ATACCC “Assessment of Transmission and Contagiousness of COVID-19 in Contacts” study, via daily PCR tests from 57 people



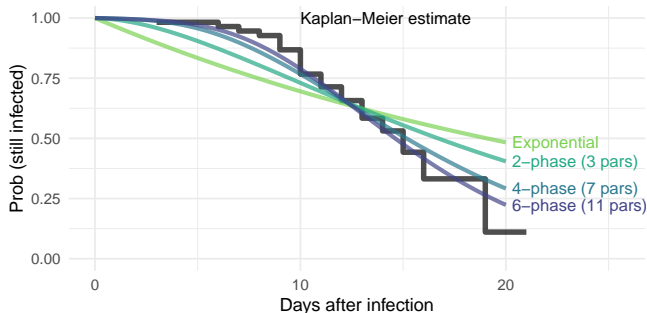
Do phase-type models fit infection duration data?

From ATACCC “Assessment of Transmission and Contagiousness of COVID-19 in Contacts” study, via daily PCR tests from 57 people



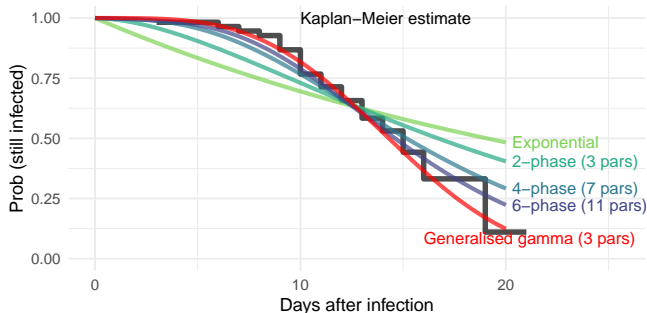
Do phase-type models fit infection duration data?

From ATACCC “Assessment of Transmission and Contagiousness of COVID-19 in Contacts” study, via daily PCR tests from 57 people



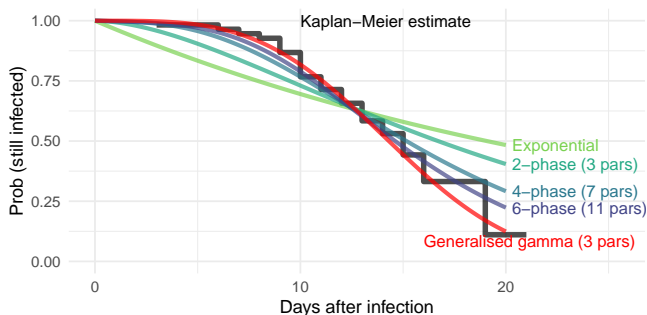
Do phase-type models fit infection duration data?

From ATACCC “Assessment of Transmission and Contagiousness of COVID-19 in Contacts” study, via daily PCR tests from 57 people



Do phase-type models fit infection duration data?

From ATACCC “Assessment of Transmission and Contagiousness of COVID-19 in Contacts” study, via daily PCR tests from 57 people



Phase-type model can be made to fit arbitrarily well, but may not be "parameter-efficient".

However its structure as a latent Markov model allows easy computation for intermittently-observed data, where generic time-to-event models (e.g. Weibull, gamma...) are difficult

`msmbayes` package allows the multi-state models from `msm` to be enhanced with priors

- ▶ advantages: use of background knowledge and judgement, stabler computation, clearer uncertainty presentation

Facilitates `phase-type models` which relax the (strong) Markov assumption

- ▶ Open questions about choice of structure / priors for optimal fit to data and plausible uncertainty

Application to infection duration distribution estimation?

- ▶ Compare with other assumption-light approaches (Joshua Blake's thesis: integrating over latent event times)
- ▶ More complex problem than suggested here (more predictors of incidence and duration, states representing immunity... scalability issues)

`msmbayes` package allows the multi-state models from `msm` to be enhanced with priors

- ▶ advantages: use of background knowledge and judgement, stabler computation, clearer uncertainty presentation

Facilitates `phase-type models` which relax the (strong) Markov assumption

- ▶ Open questions about choice of structure / priors for optimal fit to data and plausible uncertainty

Application to infection duration distribution estimation?

- ▶ Compare with other assumption-light approaches (Joshua Blake's thesis: integrating over latent event times)
- ▶ More complex problem than suggested here (more predictors of incidence and duration, states representing immunity... scalability issues)

`msmbayes` package allows the multi-state models from `msm` to be enhanced with priors

- ▶ advantages: use of background knowledge and judgement, stabler computation, clearer uncertainty presentation

Facilitates `phase-type models` which relax the (strong) Markov assumption

- ▶ Open questions about choice of structure / priors for optimal fit to data and plausible uncertainty

Application to infection duration distribution estimation?

- ▶ Compare with other assumption-light approaches (Joshua Blake's thesis: integrating over latent event times)
- ▶ More complex problem than suggested here (more predictors of incidence and duration, states representing immunity... scalability issues)