Semi-Markov multi-state models for panel data

Making them accessible and stable

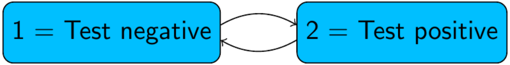
Christopher Jackson

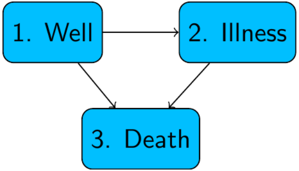
Royal Statistical Society Conference, Edinburgh, Sep 2025

MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

Multi-state models





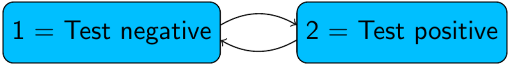
... or any other state and transition structure

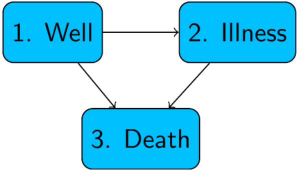
Parameters: continuous-time models with transition intensities / rates / hazards
$q_{rs} = \exp(\beta_{rs}\mathbf{x})$

Estimate:, e.g.,

- expected time spent in a state (e.g. duration of an infection)
- probabilities of transition between states, over periods of time ...

Multi-state models





. . . or any other state and transition structure

Parameters: continuous-time models with transition intensities / rates / hazards

$q_{rs} = \exp(\beta_{rs}\mathbf{x})$

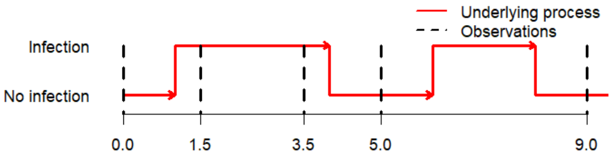
Estimate:, e.g.,

- expected time spent in a state (e.g. duration of an infection)
- probabilities of transition between states, over periods of time . . .

Data

Multi-state models get applied to a wide range of data structures

Continuous-time models, but intermittent observations: In our applications, we only know the state at a finite set of times — e.g. when person is tested for infection



Don't know transition times between states:

- e.g. when someone got the infection, when it cleared

Some infections may be completely unobserved for people in the data

Model estimation and challenges

Standard framework based on maximum likelihood estimation (Kalbfleisch and Lawless, JASA 1985) `msm` R package (CRAN, Jackson 2011 J. Stat. Soft.) widely used.

Consequences of intermittent observation:

Strong model assumptions

- Markov assumption: exponentially-distributed staying time in state
- Constant or piecewise-constant hazards

Estimation challenges:

- lots of parameters, hard to tell which are informed by data →
- model fitting prone to non-convergence

Model estimation and challenges

Standard framework based on maximum likelihood estimation (Kalbfleisch and Lawless, JASA 1985) `msm` R package (CRAN, Jackson 2011 J. Stat. Soft.) widely used.

Consequences of intermittent observation:

Strong model assumptions

- Markov assumption: exponentially-distributed staying time in state
- Constant or piecewise-constant hazards

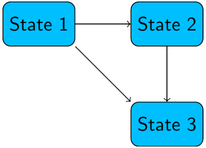
Estimation challenges:

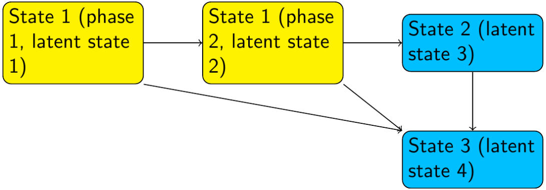
- lots of parameters, hard to tell which are informed by data $\rightarrow$
- model fitting prone to non-convergence

Semi-Markov “phase-type” model

(Titman and Sharples Biometrics 2012).

Allows the rate of transition out of some state to change with the time spent in that state. Example:





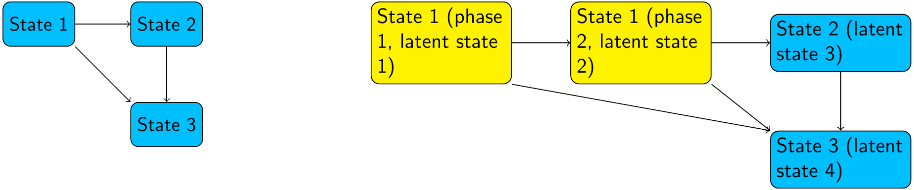
Replace an observable state (state 1 in this picture) with a set of latent states (“phases”). Latent states follow a Markov model.

Hidden Markov model: hence likelihood calculation tractable

Semi-Markov "phase-type" model

(Titman and Sharples Biometrics 2012).

Allows the rate of transition out of some state to change with the time spent in that state. Example:



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

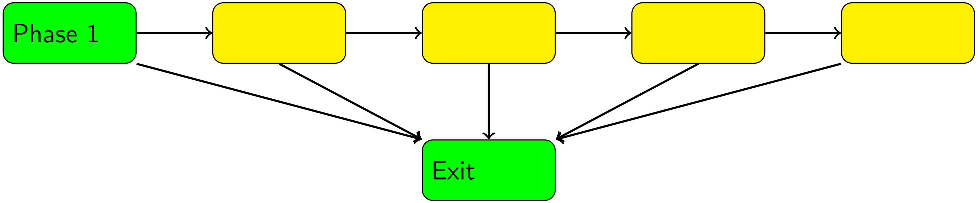
Hidden Markov model: hence likelihood calculation tractable

Phase-type sojourn distributions

In a Markov model, the sojourn time in every state is exponentially distributed

In this semi-Markov model, the sojourn time in some state follows a "phase-type" sojourn distribution

Time from entering state 1 to reaching the "Exit" state in a continuous-time Markov model structure like this:



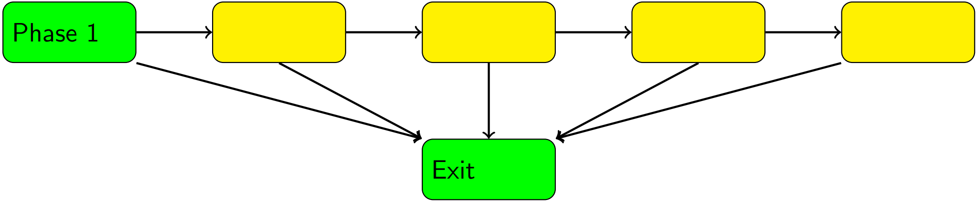
- More latent phases and transitions → more flexible sojourn distribution . . . more estimation challenges, particularly with coarse data

Phase-type sojourn distributions

In a Markov model, the sojourn time in every state is exponentially distributed

In this semi-Markov model, the sojourn time in some state follows a "phase-type" sojourn distribution

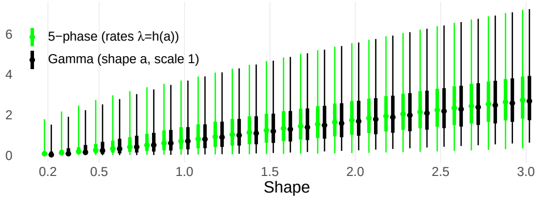
Time from entering state 1 to reaching the "Exit" state in a continuous-time Markov model structure like this:



- More latent phases and transitions → more flexible sojourn distribution ... more estimation challenges, particularly with coarse data

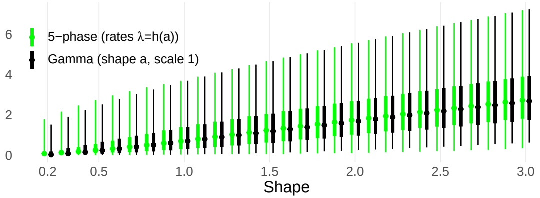
Phase-type approximations to common distributions

Approximate the Weibull or Gamma (shape a, scale b) with a phase-type family with rate $\boldsymbol{\lambda} = b\mathbf{h}(a)$ (Titman, Stat Comp 2014)



1. Find a function $h()$ that maps shape a to parameters $\boldsymbol{\lambda}$ of best-matching phase type — store $h()$ in software
2. Observe data. Define multi-state model where some state has this sojourn distribution, parameterised by a, b.
3. Fit model using (tractable!) hidden Markov model likelihood, based on the matching phase-type family

Approximate the Weibull or Gamma (shape a, scale b) with a phase-type family with rate $\boldsymbol{\lambda} = b\mathbf{h}(a)$ (Titman, Stat Comp 2014)

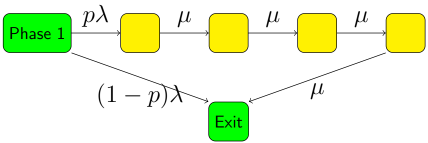


1. Find a function $h()$ that maps shape a to parameters $\boldsymbol{\lambda}$ of best-matching phase type — store $h()$ in software
2. Observe data. Define multi-state model where some state has this sojourn distribution, parameterised by a, b.
3. Fit model using (tractable!) hidden Markov model likelihood, based on the matching phase-type family

Phase-type approximations by moment matching

Titman (2014) did a complicated spline fitting to find the mapping $h()$

Easier way: for phase-type distributions of this form, there is an analytic formula for λ, μ, p that give a particular mean, variance and skewness



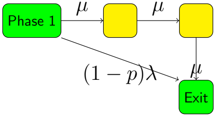
(Bobbio et al. Stochastic Models 2005)

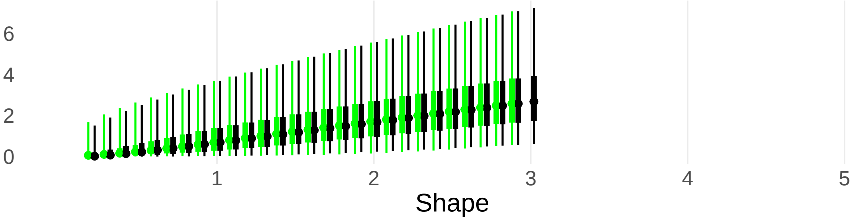
So to approximate Gamma (or Weibull), given shape and scale,

- Calculate first three moments of the Gamma (or Weibull)
- $\rightarrow$ determine matching phase parameters λ, p, μ.

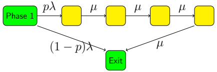
With more phases n, can match a wider range of the target distribution. Example...

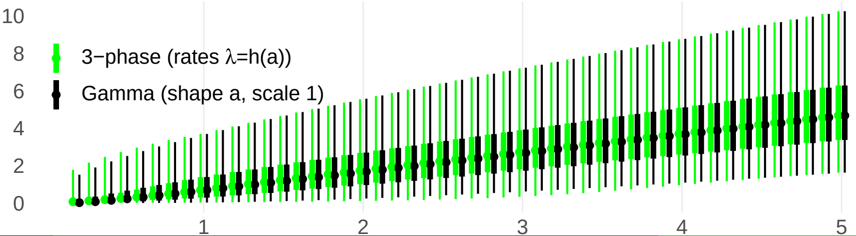
Gamma with shapes 0 to 3 can be moment-matched to 3-phase distributions



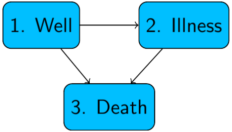


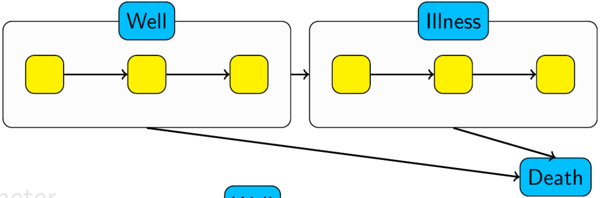
Gamma with shapes up to 5 can be moment-matched to 5-phase distributions



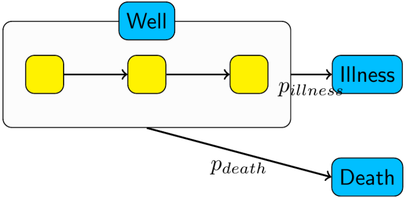


One or more states r can have a “phase-type approximation” sojourn distribution with shape a, scale b





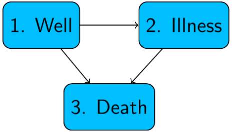
Covariates: can modify the scale parameter b

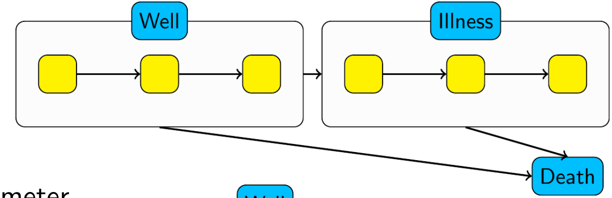


Or if 2+ “competing risks” on leaving a “phase-type” state: covariates can affect probability of competing state

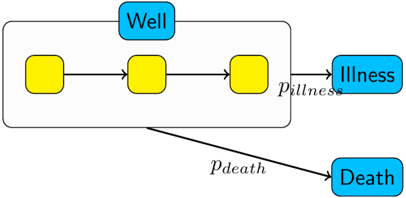
Building multi-state models with “phase-type” states

One or more states r can have a “phase-type approximation” sojourn distribution with shape a, scale b





Covariates: can modify the scale parameter b



Or if 2+ “competing risks” on leaving a “phase-type” state: covariates can affect probability of competing state

Computation

Bayesian inference, maximum likelihood or approximate Bayes (Laplace around posterior mode), using "off-the-shelf" algorithms in the **Stan** software

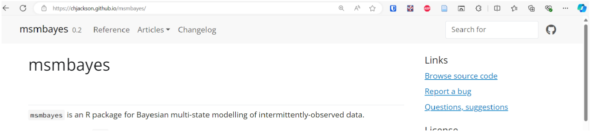
Identifiability/stability

- Weakly informative prior/penalty from background information recommended
- Pure MLE often fails with infrequently-observed data
- What if no information in data? Get posterior where all information comes from the prior. More useful than convergence failure

Scalability: With number of distinct covariate values / observation times, size of the latent state space. . .

msmbayes R package

Extends the msm package for Markov models to do Bayesian inference and phase-type semi-Markov models

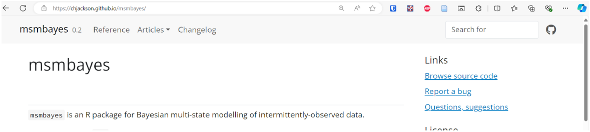


Familiar interface, like common R modelling packages

```
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)
)
mod <- msmbayes(data = infsim, state="state", time="months",
                subject="subject", qmatrix=q, priors=priors,...)
summary(mod)
```

msmbayes R package

Extends the msm package for Markov models to do Bayesian inference and phase-type semi-Markov models



Familiar interface, like common R modelling packages

```
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)
)
mod <- msmbayes(data = infsim, state="state", time="months",
                subject="subject", qmatrix=q, priors=priors,...)
summary(mod)
```

Simulation-based calibration study (as in Talts et al. 2018)

Assess correctness of Bayesian computation procedure:

- Simulate many datasets from prior predictive distribution
- Fit models to them: average of resulting posteriors should match the prior

Designed here around infection duration example (2 or 3-state).

Results:

- MCMC estimation accurate under a range of model structures, but slow
- Laplace approximation is mildly biased / underestimates uncertainty, but allows scalability

Simulation-based calibration study (as in Talts et al. 2018)

Assess correctness of Bayesian computation procedure:

- Simulate many datasets from prior predictive distribution
- Fit models to them: average of resulting posteriors should match the prior

Designed here around infection duration example (2 or 3-state).

Results:

- MCMC estimation accurate under a range of model structures, but slow
- Laplace approximation is mildly biased / underestimates uncertainty, but allows scalability

Applications

(a) Estimating infection duration and incidence

Cohort of people tested intermittently.
Start / end times of person's infection unknown.

(b) Cancer screening

No cancer → detectable precursor → clinical cancer . . ..
Incidence, time to progression not exponentially distributed.
Choose optimal screening interval. (e.g. Akwiwu et al. BMC Med Res Meth 2022)

(c) Cognitive function

Longitudinal studies of ageing, observations every 2 years.
Dynamics of decline in cognitive function, and mortality.

Applications

(a) Estimating infection duration and incidence

Cohort of people tested intermittently.
Start / end times of person's infection unknown.

(b) Cancer screening

No cancer → detectable precursor → clinical cancer ....
Incidence, time to progression not exponentially distributed.
Choose optimal screening interval. (e.g. Akwiwu et al. BMC Med Res Meth 2022)

(c) Cognitive function

Longitudinal studies of ageing, observations every 2 years.
Dynamics of decline in cognitive function, and mortality.

Applications

(a) Estimating infection duration and incidence

Cohort of people tested intermittently.
Start / end times of person's infection unknown.

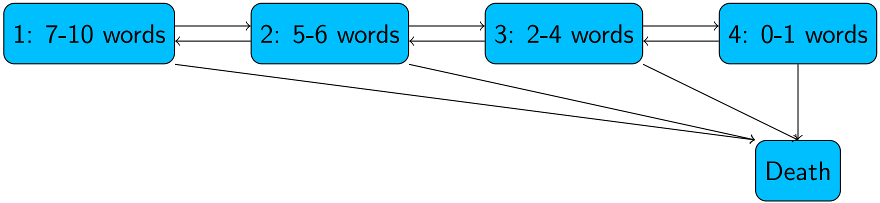
(b) Cancer screening

No cancer $\rightarrow$ detectable precursor $\rightarrow$ clinical cancer ....
Incidence, time to progression not exponentially distributed.
Choose optimal screening interval. (e.g. Akwiwu et al. BMC Med Res Meth 2022)

(c) Cognitive function

Longitudinal studies of ageing, observations every 2 years.
Dynamics of decline in cognitive function, and mortality.

Realistically-complex illustrative application:



English Longitudinal Study of Ageing. Cognitive function test: how many words from a list of 10 recalled after a few minutes. $\approx$ 5000 observations from people aged 50+, every 2 years.

Semi-Markov model on all four states: 21 latent "phases"

Predictors of transitions: age, gender, education

Strong priors on mortality rates from national statistics

MCMC not feasible, use Laplace approximation to posterior

Results

"Covariate effect parameters" hard to interpret. Instead:

Calculate expected total amount of time spent with no/mild cognitive impairment over next 10 years

Compare this between categories of one covariate (standardised over others)

Discussion

Made semi-Markov models for intermittently-observed data practicable

Software to make Bayesian inference in general Markov and semi-Markov models accessible

Challenges: computational scalability, prior specification, more practical experience and training resources. . .

https://chjackson.github.io/msmbayes

Paper on ArXiV linked from there, with full details of the studies described here.