# Robustness properties of the (landmark) Aalen-Johansen estimator

Hein Putter

Department of Medical Statistics and Bioinformatics
Leiden University Medical Center

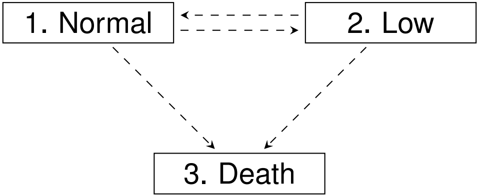
ISCB Vigo
July 11, 2017

# Two parts

1. Landmark Aalen-Johansen estimator (LMAJ)
   - Online at Stat Methods Med Res (2016+++)
   - Joint work with Cristian Spitoni
2. Alternative representations of the Aalen-Johansen (AJ) estimator of the state occupation probabilities
   - Ongoing work with Cristian Spitoni and Ronald Geskus

# CSL1 trial

- Placebo-controlled randomized clinical trial
- To evaluate the efficacy of prednisone in patients with histologically verified liver cirrhosis
- Prothrombin index, indicator of liver functioning
  - Dichotomized into normal and low (less than 70% of normal values)
  - Changes between normal and low assumed to take place at measurement
- Studied in the context of a multi-state model (Andersen et al. 1993)



# Notation

- State space $\mathcal{K} = \{1, \ldots, K\}$, finite
- Random process $\{X(t); t \geq 0\}$ taking values in $\mathcal{K}$
- History: The sigma-algebra generated by the observed trajectories $\mathcal{H}_t$

# Important concepts

- Transition intensities:

$$\lambda_{jk}(t \mid \mathcal{H}_{t-}) = \lim_{\Delta t \downarrow 0} P(X(t+\Delta t) = k \mid X(t) = j, \mathcal{H}_{t-})/\Delta t.$$

# Important concepts

- Transition intensities:

$$\lambda_{jk}(t \mid \mathcal{H}_{t-}) = \lim_{\Delta t \downarrow 0} P(X(t+\Delta t) = k \mid X(t) = j, \mathcal{H}_{t-})/\Delta t.$$

- Transition probabilities:

$$P_{\ell m}(s, t \mid \mathcal{H}_{s-}) = P(X(t) = m \mid X(s) = \ell, \mathcal{H}_{s-}).$$

  - This is the probability that the process is in state $m$ at time $t$, given that it was in state $\ell$ at time $s$ and conditionally on the past trajectory until time $s$
- Note: the transition intensities only concern direct transitions, while transition probabilities might go over several transitions

# Important concepts

- Transition intensities:

$$\lambda_{jk}(t \mid \mathcal{H}_{t-}) = \lim_{\Delta t \downarrow 0} P(X(t+\Delta t) = k \mid X(t) = j, \mathcal{H}_{t-})/\Delta t.$$

- Transition probabilities:

$$P_{\ell m}(s,t \mid \mathcal{H}_{s-}) = P(X(t) = m \mid X(s) = \ell, \mathcal{H}_{s-}).$$

  - This is the probability that the process is in state $m$ at time $t$, given that it was in state $\ell$ at time $s$ and conditionally on the past trajectory until time $s$
- Note: the transition intensities only concern direct transitions, while transition probabilities might go over several transitions
- State occupation probabilities:

$$P_m(t) = P(X(t) = m)$$

# Markov property

- The multi-state process is a Markov process if it fulfills the Markov property:

$$P(X(t) = m \,|\, X(s) = \ell, \mathcal{H}_{s-}) = P(X(t) = m \,|\, X(s) = \ell)$$

  - Future behavior of the process depends on the history only through the present

## If the Markov property holds

- The transition intensities/hazards and the transition probabilities do not depend on the further past anymore, so

$$\begin{aligned} \lambda_{jk}(t) &= \lim_{\Delta t \downarrow 0} P(X(t + \Delta t) = k \,|\, X(t) = j)/\Delta t \\ P_{\ell m}(s, t) &= P(X(t) = m \,|\, X(s) = \ell) \end{aligned}$$

# Relation intensities and probabilities

- The Markov assumption is very convenient, mathematically, because:
- If the multi-state model is Markov, then

$$\mathbf{P}(s,t) = \prod_{s<u\leq t} \left(\mathbf{I} + d\mathbf{\Lambda}(u)\right)$$

- $\mathbf{P}(s,t)$, $K \times K$ matrix with $(\ell, m)$th element $P_{\ell m}(s,t)$
- $\mathbf{\Lambda}(t)$ also $K \times K$ matrix
  - $(j,k)$th element $\Lambda_{jk}(t) = \int_0^t \lambda_{jk}(s)ds$
  - $(j,j)$th diagonal element $\Lambda_{jj}(t) = -\sum_{k\neq j} \Lambda_{jk}(t)$

# State occupation probabilities

- Vector $\mathbf{P}(t)$ of state occupation probabilities
  - $m$th element $P_m(t) = P(X(t) = m)$
  - Can be retrieved through $\mathbf{P}(t) = \pi(0)\mathbf{P}(0, t)$
  - $\pi(0)$ a $1 \times K$ vector with $k$th element $\pi_k(0) = P(X(0) = k)$
- Together:

$$\mathbf{P}(t) = \pi(0) \prod_{0<u\leq t} (\mathbf{I} + d\mathbf{\Lambda}(u))$$

- Product integral relation similar to

$$S(t) = \prod_{0<u\leq t} (1 - d\Lambda(u))$$

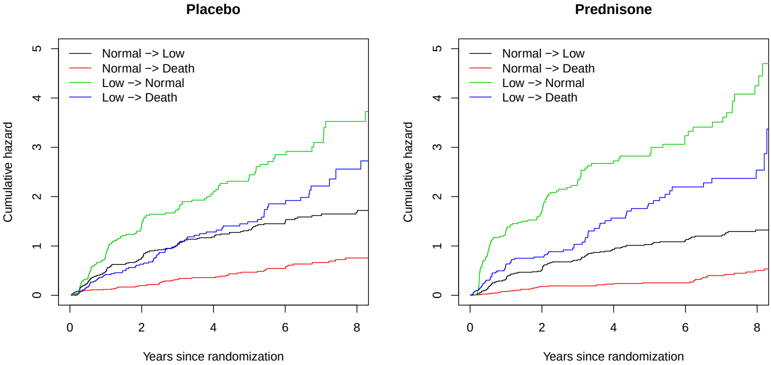
# Estimation of transition intensities

- Observe right censored versions of the multi-state process
- Non-parametric estimate of transition intensity $j \to k$:

$$\widehat{\Lambda}_{jk}(t) = \int_0^t Y_j^{-1}(u) dN_{jk}(u)$$

- Extension of the Nelson-Aalen estimator in standard survival analysis

# Transition intensities



# The Aalen-Johansen estimator

- Gather estimated transition intensities in matrices $\widehat{\mathbf{\Lambda}}$
- Aalen-Johansen estimator uses relationship between transition intensities and probabilities
  - Replacing $\mathbf{\Lambda}$ by estimate $\widehat{\mathbf{\Lambda}}$

$$\widehat{\mathbf{P}}(s,t) = \prod_{(s,t]} \left( \mathbf{I} + \Delta \widehat{\mathbf{\Lambda}}(u) \right)$$

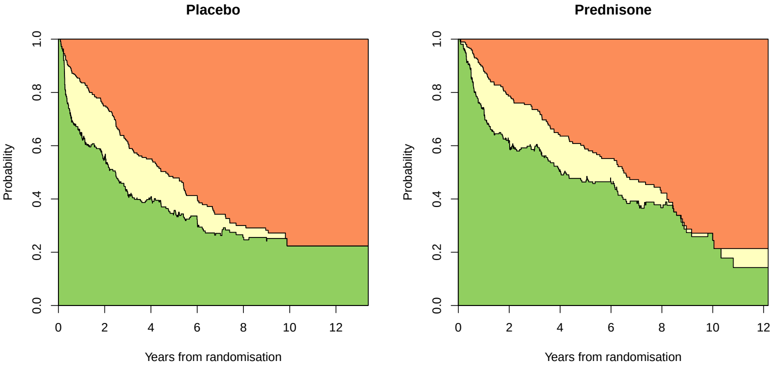
- Estimate of state occupation probabilities

$$\widehat{\mathbf{P}}(t) = \hat{\pi}(0) \prod_{0<u\leq t} \left( \mathbf{I} + \Delta \widehat{\mathbf{\Lambda}}(u) \right)$$

with $\hat{\pi}_k(0) = n^{-1} \sum_{i=1}^{n} 1\{X_i(0) = k\}$

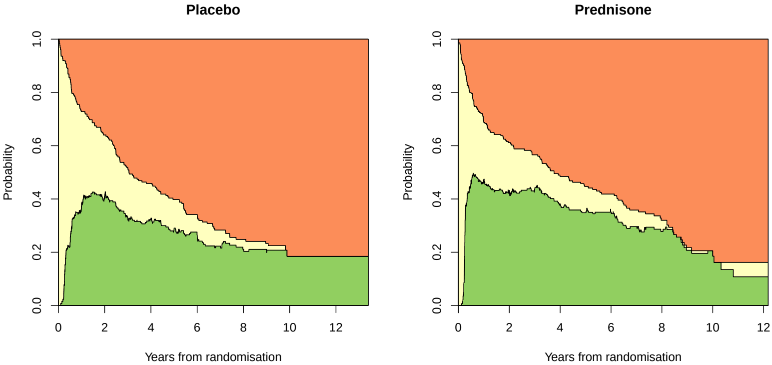
# Transition probabilities

## Given normal prothrombin level at baseline



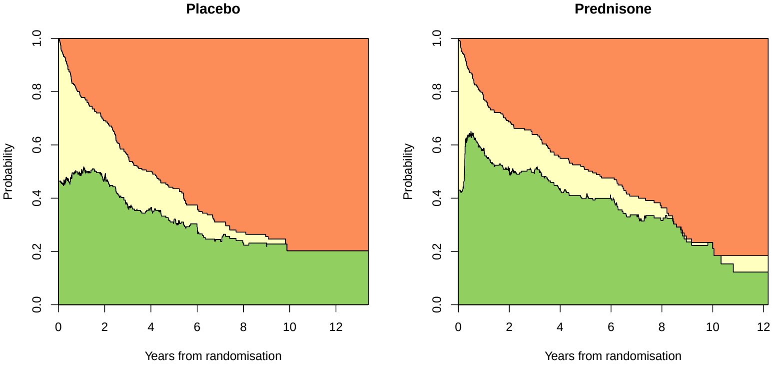
# Transition probabilities

## Given low prothrombin level at baseline



# State occupation probabilities

- At baseline 46.4% of placebo and 43.0% of prednisone in normal state



# Back to the Markov property

- The Aalen-Johansen is valid for arbitrary multi-state models
- As long as they are Markovian
- The Markov property facilitates elegant mathematical theory connecting transition intensities and transition probabilities
- It is therefore often assumed (taken for granted)
- But not at all obvious
- Violations
  - Duration dependence (semi-Markov)
  - “State visit” dependence
- Unfortunately, in our application, it turns out that normal → low hazard increases with number of previous “lows”

# Transition probabilities for non-Markovian multi-state models

- When the multi-state model is not Markov, $P_{\ell m}(s, t \,|\, \mathcal{H}_{s-})$ will depend on the past before time $s$
  - Markov renewal process: $P_{\ell m}(s, t \,|\, \mathcal{H}_{s-})$ will crucially depend on the time at which state $\ell$ was entered before $s$
- When nature of violation of Markov is known, it is of course better to take that into account
- Transition probability $P_{\ell m}(s, t)$ still of interest when
  1. Extra relevant information in $\mathcal{H}_{s-}$ is not available
  2. When it is unknown how $\lambda_{jk}(t; \mathcal{H}_{t-})$ depends on the history
     - In practice, often uncertain whether or not the multi-state model is Markov
     - Useful to have estimator that is robust against possible non-Markovianity

# Recent work

- After earlier work by Meira-Machado et al. (2006) and Allignol et al. (2014)
- Two recent papers, both Biometrics in 2015
- Both based on subsampling approach: selection is made of the data consisting of subjects occupying a given state at a particular time
- de Uña Álvarez and Meira-Machado (2015)
  - Irreversible illness-death model, extension to general forward-going multi-state models
  - Works with (differences between) Kaplan-Meiers
- Titman (2015)
  - Valid for any multi-state model
  - Based on derived competing risks processes on a "per $\ell m$ basis"
    - As a result, the estimators of $\widehat{P}_{lm}(s,t)$ are not guaranteed to add up to 1

# Datta & Satten (2001)

- Datta & Satten (2001) showed that

$$\widehat{\mathbf{P}}(t) = \hat{\pi}(0) \prod_{0<u\leq t} \left(\mathbf{I} + \Delta\widehat{\boldsymbol{\Lambda}}(u)\right)$$

  provides a consistent estimator of the state occupation probabilities $P(X(t) = m)$, also if the Markov assumption is violated
- Note: the estimators $\widehat{\boldsymbol{\Lambda}}(u)$ are exactly as used before in the Markov model, but now have an interpretation as *partly conditional rates*
- Can we exploit this result to robustly estimate (partly conditional) transition probabilities?

# The landmark Aalen-Johansen estimator

### For fixed $s$ and $\ell$

- Interested in $P_{\ell m}(s, t) = P(X(t) = m \,|\, X(s) = \ell)$
- Can also be seen as a state occupation probability $P(X(t) = m)$ in a subset
- Use same subsetting/subsampling idea as used in de Uña Álvarez and Meira-Machado (2015) and Titman (2015)
- Superscript $(\mathrm{LM})$ denotes counting and at risk processes and estimators based on a landmark data set which selects subjects who are at state $\ell$ at time $s$
- Dependence on $s$ and $\ell$ suppressed in notation

# The landmark Aalen-Johansen estimator

- Landmark based version of $\widehat{\mathbf{\Lambda}}(t)$:

$$\widehat{\Lambda}_{jk}^{(\mathrm{LM})}(t) = \int_0^t \overline{Y}_j^{(\mathrm{LM})^{-1}}(u) d\overline{N}_{jk}^{(\mathrm{LM})}(u)$$

- LMAJ estimator given by

$$\widehat{P}_{\ell m}^{\mathrm{LMAJ}}(s,t) = \hat{\pi}^{(\mathrm{LM})}(s) \prod_{s<u\leq t} \left( \mathbf{I} + \Delta \widehat{\mathbf{\Lambda}}^{(\mathrm{LM})}(u) \right),$$

with $\hat{\pi}_\ell^{(\mathrm{LM})}(s) = 1$, and 0 otherwise

# Comments

$$\widehat{P}^{\mathrm{LMAJ}}_{\ell m}(s,t) = \hat{\pi}^{(\mathrm{LM})}(s) \prod_{s<u\leq t} \left(\mathbf{I} + \Delta \widehat{\mathbf{\Lambda}}^{(\mathrm{LM})}(u)\right)$$

- It is the Aalen-Johansen estimator of the state occupation probabilities based on the subset of subjects in state $\ell$ at time $s$
- Conditioning is simply accomplished by taking relevant subset
- This is also at the core of the landmarking method for dynamic prediction
- Hence the name landmark Aalen-Johansen, henceforth LMAJ

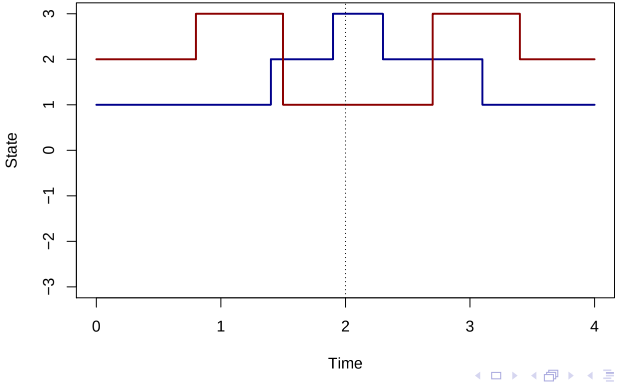
# Consistency

## Property

- The LMAJ estimator $\widehat{P}_{\ell m}^{\mathrm{LMAJ}}(s, t)$ is a consistent estimator of $P_{\ell m}(s, t)$, provided
  - The conditions of Datta & Satten (2001) hold for consistency of state occupation probabilities
  - $P(X(s) = \ell) > 0$

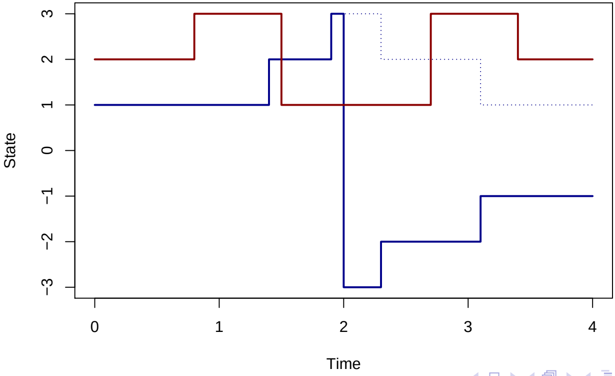
# Idea of the proof

- For instance interested in $P_{\ell m}(s,t)$, for $s = 2$, $\ell = 1$
- Define coupled multi-state model $X^{\star}(t)$ derived from original $X(t)$



# Idea of the proof

- For instance interested in $P_{\ell m}(s, t)$, for $s = 2$, $\ell = 1$
- Define coupled multi-state model $X^{\star}(t)$ derived from original $X(t)$



# Idea of the proof

- Relation between (transition) probabilities of the two multi-state models

$$P^\star_m(t) = P(X^\star(t) = m) = P(X(t) = m, X(s) = \ell),$$

- As a result, the transition probability of interest can be written as

$$P_{\ell m}(s,t) = P(X(t) = m \,|\, X(s) = \ell) = \frac{P^\star_m(t)}{P^\star_\ell(s)}$$

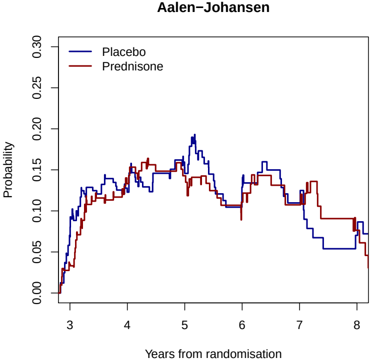
- A ratio of state occupation probabilities
  - For which we use the Datta & Satten (2001) result
- It turns out that the ratio of these state occupation probability estimates is exactly the LMAJ!

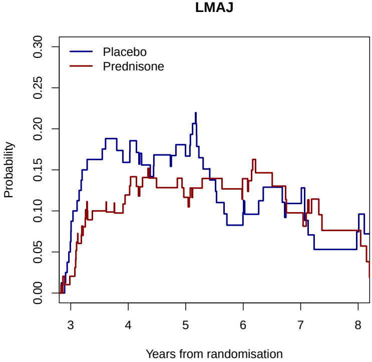
# Remarks

- Standard errors
  - Greenwood estimators of pointwise standard errors of the Aalen-Johansen estimator of the state occupation probabilities remain valid also if the Markov assumption is violated (Glidden 2002)
  - Same is true for the LMAJ estimator
  - For simultaneous confidence bands more elaborate methods need to be used
- Generalized conditional probabilities
  - As in Titman (2015), LMAJ can also be generalized to estimates of $P_{\mathcal{L}\mathcal{M}}(s,t) = P(X(t) \in \mathcal{M} \mid X(s) \in \mathcal{L})$
  - Replace $1\{X_i(s) = \ell\}$ by $1\{X_i(s) \in \mathcal{L}\}$
  - Sum over all $m \in \mathcal{M}$

# Transition probabilities

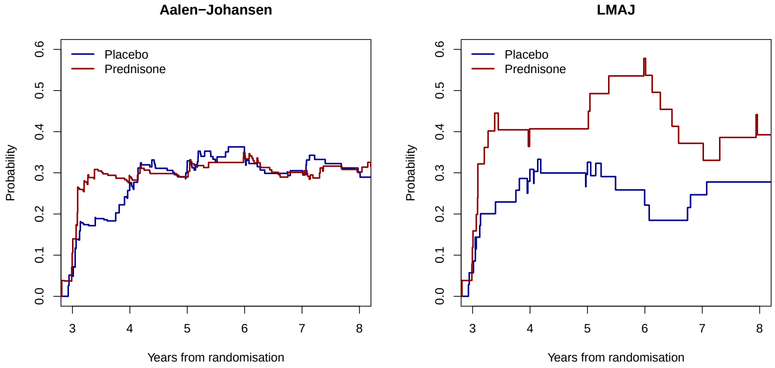
## Probability of low, given normal at 1000 days



# Transition probabilities

## Probability of normal, given low at 1000 days



# "Half-way" summary

## The landmark Aalen-Johansen estimator

- Simple and intuitive
- Consistent under same conditions as Datta & Satten (2001)
- Valid for arbitrary multi-state models
- Slightly more efficient in simulations than Titman (2015)
- Probabilities for different target states add up to 1

- Implemented (function `LMAJ`) in **mstate** package for R (version 0.2.9 and higher)

# Datta & Satten's result

- LMAJ is consistent under same conditions as Datta & Satten (2001)
- Uses Datta & Satten's consistency result
- Why does it work?

# Why does this work?

## In the absence of censoring

- Suppose that initial distribution is

$$V = \begin{pmatrix} n_1 & n_2 & n_3 & \dots & n_K \end{pmatrix}$$

- And suppose that we observe a transition from state 2 to 3

$$U \ \left(= \left(\mathbf{I} + \Delta\widehat{\boldsymbol{\Lambda}}(u)\right)\right) \ = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 - 1/n_2 & 1/n_2 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

- Then new distribution is

$$VU = \begin{pmatrix} n_1 & n_2 - 1 & n_3 + 1 & \dots & n_K \end{pmatrix}$$

# Robustness of Aalen-Johansen (AJ)

- Multiplication of AJ just corresponds to moving an individual from one state to another
- No Markov assumption needed as long as one focuses on state distribution only
- In the absence of censoring:

$$\widehat{P}_m^{\mathrm{AJ}}(t) = \frac{1}{n} \sum_{i=1}^{n} 1\{X_i(t) = m\}$$

- Clear that $\widehat{P}_m^{\mathrm{AJ}}(t)$ is consistent and that the Markov condition is not needed
- If there is censoring, one needs to estimate the $U$ matrices
  - This is exactly what the Nelson-Aalen estimates do

# Equivalent representations of AJ

- Two alternative, equivalent representations of AJ shed light on the robustness property
- Redistribution to the right (RTR)
- Inverse probability of censoring weighting (IPCW)
- Equivalence well known for standard survival analysis (Kaplan-Meier) and for competing risks

# RTR for multi-state models

## More notation

- $\tau_0 = 0; \tau_1 < \ldots < \tau_J$: distinct time points at which transitions (of any kind) occur
- In case of simultaneous transitions and censorings, the transitions occur first
- For each time point $\tau_j$ and each state $\ell$, define
  - $\mathcal{R}_\ell(\tau_j)$: risk set; those that are in state $\ell$ at time $\tau_j-$
  - $Y_\ell(\tau_j)$: size of the risk set $\mathcal{R}_\ell(\tau_j)$

# RTR algorithm

**0.** At $\tau_0 = 0$, divide mass equally among all at risk, define the weight of subject $i$ as $w_i(\tau_0)$. Let $\hat{p}_\ell(\tau_0)$ be the sum of weights over all subjects in state $\ell$:
$\hat{p}_\ell(\tau_0) = \sum_i w_i(\tau_0)\mathbf{1}\{X_i(\tau_0) = \ell\}$

**1.** For all $j = 1, \ldots, J$

- **1.1** Construct $\mathcal{R}_\ell(\tau_j)$, calculate $Y_\ell(\tau_j)$
- **1.2** For each state $\ell$, redivide previous total weight $\hat{p}_\ell(\tau_{j-1})$ equally among all $i \in \mathcal{R}_\ell(\tau_j)$, leading to new weights $w_i(\tau_j)$
- **1.3** Transport asigned mass among all observed transitions
- **1.4** Calculate $\hat{p}_\ell(\tau_j) = \sum_i w_i(\tau_j)\mathbf{1}\{X_i(\tau_j) = \ell\}$

# Equivalence

- Claim is that for every $\tau_j$, $\hat{p}_\ell(\tau_j)$ equals the Aalen-Johansen estimator of the state occupation probabilities
- Proof by showing that

$$\hat{p}_\ell^{\mathrm{AJ}}(\tau_j) - \hat{p}_\ell^{\mathrm{AJ}}(\tau_{j-1}) = \hat{p}_\ell^{\mathrm{RTR}}(\tau_j) - \hat{p}_\ell^{\mathrm{RTR}}(\tau_{j-1})$$

# Characterizing the weights

- If we write
$$\hat{p}_\ell(t) = \sum_i w_i(t) \mathbf{1}\{X_i(t) = \ell\},$$
who contributes to this sum?
- The "usable" observations are the subjects for whom we know in which state they were at time $t$
- Define
$$K_i(t) = \begin{cases} 1, & \text{if it is known in which state subject } i \text{ is at time } t; \\ 0, & \text{otherwise.} \end{cases}$$
- Consider $P(K(t-) = 1 \mid X(t-) = \ell)$
- Using a weight like $(P(K(t-) = 1 \mid X(t-) = \ell))^{-1}$ for the "usable" subjects would compensate for subjects that are missing from the sum

# IPCW representation

- If we define $\hat{w}_i(\tau_j) = (\hat{P}(K(t-) = 1 \mid X(t-) = \ell))^{-1}$
  - Using Bayes' rule and using empirical (non-parametric) estimates
- And define $\hat{p}_\ell^{\mathrm{IPCW}}(\tau_j) = \sum_i \hat{w}_i(\tau_j) \mathbf{1}\{X_i(\tau_j) = \ell\}$
- Then
$$\hat{p}_\ell^{\mathrm{AJ}}(\tau_j) = \hat{p}_\ell^{\mathrm{RTR}}(\tau_j) = \hat{p}_\ell^{\mathrm{IPCW}}(\tau_j)$$
- For standard survival analysis, this simplifies to the usual IPCW weights

# Discussion

- Easy to show that $\hat{p}_\ell^{\mathrm{IPCW}}(\tau_j)$ is consistent, irrespective of the Markov assumption
- Testing the Markov assumption
  - Andrew Titman, yesterday
  - Log-rank test between subjects in state $\ell$ at time $s$ and subjects not in state $\ell$ at time $s$
- Equivalence of AJ and RTR (IPCW) representations extends to presence of covariates

# References

Aalen OO and Johansen S.
An empirical transition matrix for nonhomogeneous Markov chains based on censored observations.
*Scand Stat Theory Appl* 1978; 5: 141–150.

Datta S and Satten GA.
Validity of the Aalen–Johansen estimators of stage occupation probabilities and Nelson–Aalen estimators of integrated transition hazards for non-Markov models.
*Stat Probab Lett* 2001; 55: 403–411.

de Uña Álvarez J and Meira-Machado L.
Nonparametric estimation of transition probabilities in the non-Markov illness-death model: A comparative study.
*Biometrics* 2015; 71: 364–375.

Titman AC.
Transition probability estimates for non-Markov multi-state models.
*Biometrics* 2015; 71: 1034–1041.

Putter H and Spitoni C.
Non-parametric estimation of transition probabilities in non-Markov multi-state models: The landmark Aalen-Johansen estimator.
*Stat Meth Med Res* 2016+++.