Kaplan-Meier

+ Non-parametric estimate of survival function
+ Implicitly assumes that all lives are identical
– Hard to test hypotheses
– Does not easily handle explanatory variables taking lots of values, e.g. age

Cox model

+ Parametric model for differences in hazard function
+ Enables hypothesis testing
– Cannot easily estimate the shape of the hazard or survival (baseline hazard)
– Does not make full use of data, only the ranked order of the event times

Limitations of Kaplan-Meier and Cox models

- Kaplan-Meier and Cox models can only handle a single, non-recurring decrement, i.e. one mode of exit with no return
- What if there are multiple decrements?
  - An endowment policy can terminate by death or by surrendering the policy.
  - A critical-illness policy can terminate by death or diagnosis of critical illness.
- What if the decrement can recur?
  - A claim might occur under a PHI policy, then the life assured might recover.
  - A pensions policy might stop paying premiums, then restart.

Markov models

- Not dealing with T or T_x, but with *transition intensities between states*
- Handles multiple states
- Handles multiple decrements
- Handles recurring events
- Handles censored data
- Makes better use of data than Kaplan-Meier or Cox models

2-state model: fundamental assumptions

1. **Markov assumption**. The probability that a life aged x will be in either state at any future time t depends only on the age x and the state currently occupied.
2. The probability ${}_{dt}q_{x+t} = \mu_{x+t} + o(dt), \quad t \geq 0$

First assumption means that the future depends only on the present, not on the past history. Second assumption is implied by the random-variable approach to future lifetime.

Derivation of second assumption
Recall Mean Value Theorem (MVT):

Let f be a continuous function on $[a, b]$ that is differentiable on (a, b). Then there exists at least one $c \in (a, b)$ such that:

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$
$$\Rightarrow f(b) - f(a) = (b - a)f'(c)$$

So, taking the random-variable assumption:

$$\begin{aligned}
{}_{dt}q_{x+t} &= F_{x+t}(dt) \\
&= F_{x+t}(dt) - F_{x+t}(0), \qquad \text{as } F_{x+t}(0) = 0 \\
&= (dt - 0) f_{x+t}(c), \qquad \text{where } 0 < c < dt \text{ by MVT and } F'_{x+t} = f_{x+t} \\
&= dt. \left[f_{x+t}(0) + c.f'_{x+t}(0) + \frac{c^2}{2} f''_{x+t}(0) + \dots \right], \qquad \text{as Maclaurin Series} \\
&= dt.S_{x+t}(0)\mu_{x+t} + dt.c.[\dots] \\
&= dt\mu_{x+t} + o(dt), \qquad \text{as } dt.c < (dt)^2 \text{ and } S_{x+t}(0) = 1
\end{aligned}$$