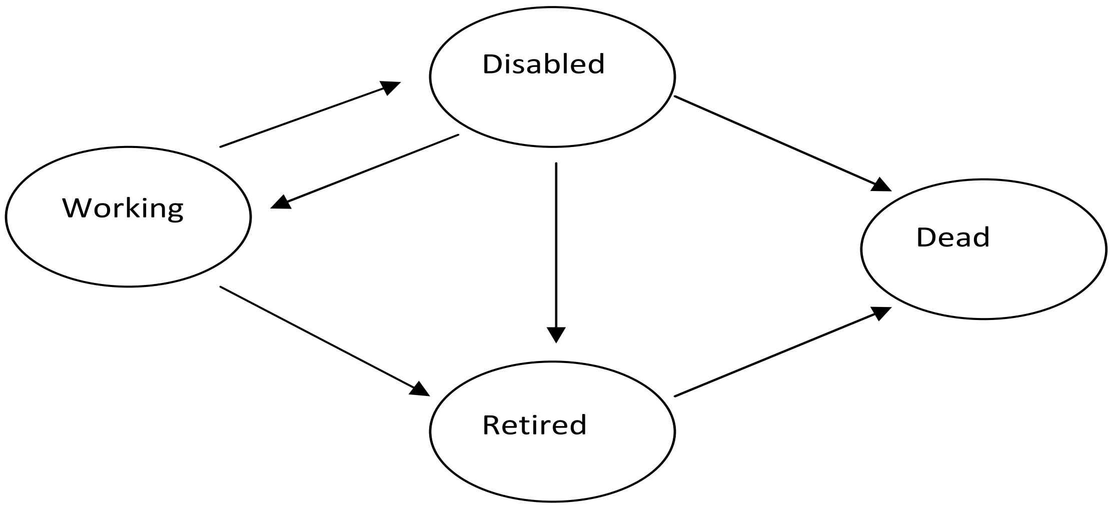
Survival Models: Week 7

Markov models - Introduction

Goal is to model the movement of individuals between various states. For example an insurer or superannuation company might be interested in the following model:

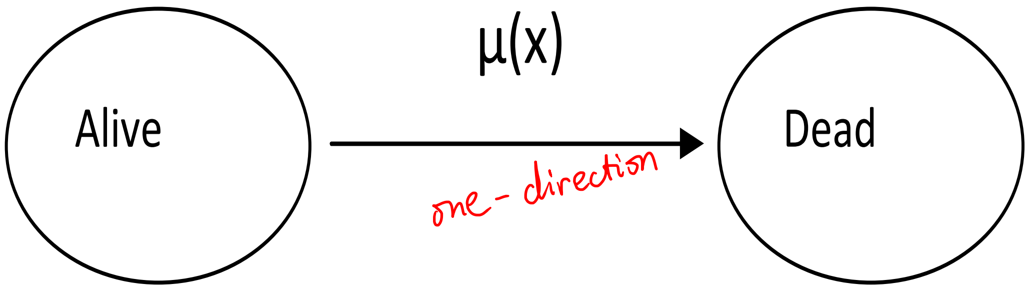


Markov models - Introduction

- The goal is to estimate the probability of transitioning between states.
- The term "markov" is used because we will assume that the probability of moving from one state to another depends only on the current state.
- The multi-state model approach is different to the approaches that we have seen before. What are we estimating?

Markov models - Two-state

First we will focus on the following two-state markov model:



$\mu(x) = \mu_x$ is the hazard (also known as transition intensity) at age x. Alive state and dead state with transitions in only one direction. The probability a life alive at a given age will be dead at any subsequent age is determined by the transition intensity $\mu_{x+t}, t \geq 0$.

Markov models - Model Assumptions

- The probabilities that a life will be found in either state at any subsequent age depend only on the ages involved and on the state currently occupied. This is the Markov assumption. Past events do not affect the probability of a future event.

- For a short interval of length dt, $t \geq 0$. Where ${}_{dt}q_{x+t} \approx \mu_{x+t}dt + o(dt)$, $o(dt)$ goes to zero faster than dt itself. This assumption means the probability of dying in a short period of time is approximately the transition intensity multiplied by the length of the interval dt of time.

- The transition intensity μ_{x+t} is a constant, for $0 \leq t \leq 1$. This treats the intensity as constant for all individuals aged x.

Model Assumptions

Based on the second assumption as listed in the previous slide, we can show that

$$\frac{d}{dt}{}_tp_x = -{}_tp_x\mu_{x+t}$$

proof skipped

introduced before

and

$${}_tp_x = \exp\left(-\int_0^t \mu_{x+s}ds\right)$$

Markov models - Definitions and Data

We observe N individuals (assumed independent and having the same hazard) during some finite period of observation, between ages of $[x, x+1)$. Using information obtained on these individuals we will estimate the hazard $\mu(x)$. To do this we will define the following: new model

- $x + a_i$ is the age that individual i comes under observation.
- $x + b_i$ is the age that individual i will cease being observed if life survives to that age: $x + b_i$.

  $0 < a_i < b_i < 1$
- T_i is defined such that $x + T_i$ is the age at which observation of life i ends.
- V_i is a random variable representing the length of time that individual i is under observation, which is called the waiting time.
- δ_i is an indicator variable that equals to 1 if individual i dies and 0 otherwise.

individual dies ~~in~~ under the observation

individual is censored

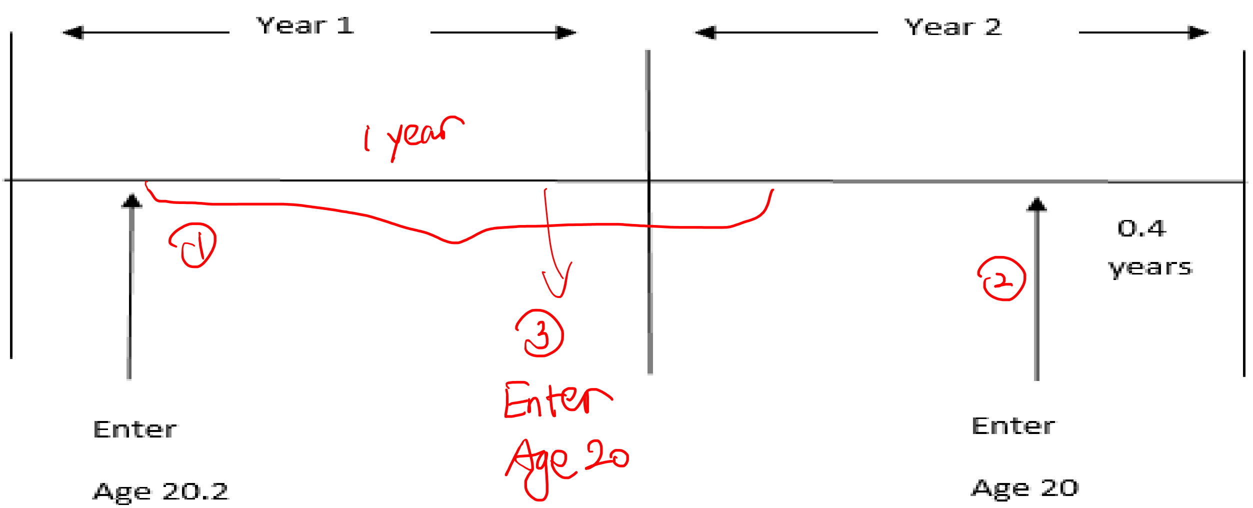
Note:

- If $\delta_i = 0$ then the random variable V_i will take the value $b_i - a_i$ and $T_i = b_i$.
- If $\delta_i = 1$ then the random variable V_i will take the value $T_i - a_i$.

study individual aged 20
aim: estimate μ_{20} (20, 21)
2-year investigation

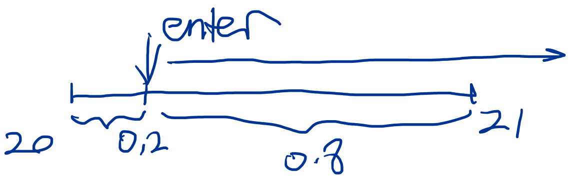
Example

N individuals, some individual will turn to 20 during this 2-year period, some individuals enter the study after 20 due to some reasons



Person 1: $a = 0.2$, $b = 1$. There is enough time remaining in the study for this person to be observed until age $x + 1$.
Person 2: $a = 0$, $b = 0.4$. This person will be aged $x + 0.4$ when the study ends.
Details see handwritten notes.

Person ① : enter the study at age 20.2, survives to age 21.



enough time remaining in this study to observe until 21

$a=0.2$

$b=1$

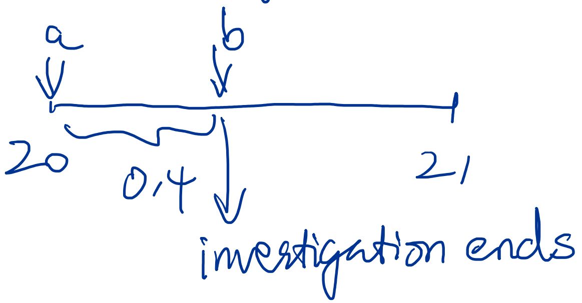
$V=b-a=0.8$

$\delta=0$

$T=1$ (time lived)

Person ② : enter the study at age 20, but investigation ends at age 20.4

not enough time remaining to observe until 21.

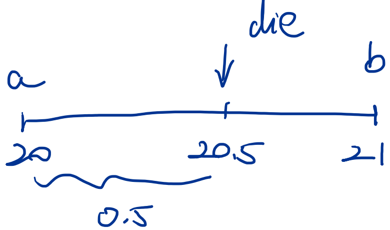


$\longrightarrow$ censored

$a=0$ $b=0.4$ $T=0.4$ $V=0.4$

$\delta=0$

Person ③ : enter the study at age 20 dies at age 20.5



$a=0$ $T=0.5$ $\delta=1$

$b=1$ $V=0.5$

MLE - censoring

$L = \prod f(t) \prod S(c_i)$

$\downarrow$ complete $\qquad$ $\downarrow$ censoring $\qquad$ $P(T_i \geq c_i)$

$\delta_i = 1$ $\qquad$ $\delta_i = 0$ $\qquad$ ${}_{v_i}P_{x+a_i}$

$\downarrow$

${}_{v_i}P_{x+a_i} \, \mu(x+a_i+v_i)$

conclude : i^{th} contribution to the likelihood

${}_{v_i}P_{x+a_i} \, \mu(x+a_i+v_i)^{\delta_i}$

Likelihood

We will now write the likelihood for the observed data (V_i, δ_i) $i = 1, \ldots, N$. This will allow us to estimate the hazard $\mu(x)$. We will consider the cases where $\delta_i = 0$ and $\delta_i = 1$ separately.

- $\delta_i = 0$: the individual lives from age $x + a_i$ to $x + a_i + v_i$. The contribution of this event to the likelihood is ${}_{v_i}p_{x+a_i}$.

- $\delta = 1$: the individual lives from age $x + a_i$ to $x + a_i + v_i$ and then dies aged $x + a_i + v_i$. The contribution of this event to the likelihood is ${}_{v_i}p_{x+a_i}\mu(x + a_i + v_i)$.

This means that the individual's contribution to the likelihood is ${}_{v_i}p_{x+a_i}\mu(x + a_i + v_i)^{\delta_i}$.

Likelihood

To write the likelihood we use the assumption of independence and the fact that ${}_v p_{x+a} = \exp(-\int_{x+a}^{x+a+v} \mu(s)ds)$. As mentioned before, we have made the assumption that the hazard $\mu(s) = \mu$ is constant over the year of age $[x, x+1)$. This gives the following:

$$
\begin{aligned}
L &= \prod_{i=1}^{N} \exp(-v_i\mu)\mu^{\delta_i} \\
&= \exp(-\mu \sum_{i=1}^{N} v_i)\mu^{\sum_{i=1}^{N} \delta_i} \\
\Longrightarrow \; l &= -\mu \sum_{i=1}^{N} v_i + log(\mu) \sum_{i=1}^{N} \delta_i \\
&= -\mu v + log(\mu)\delta,
\end{aligned}
$$

where:

- $\sum_{i=1}^{N} v_i = v$ is the total time under observation of all N individuals
- $\sum_{i=1}^{N} \delta_i = \delta$ is the total number of deaths.

We can now maximize the likelihood to find our estimate of μ.

$$l = -\mu v + log(\mu)\delta$$

$$l^{'} = -v + \frac{\delta}{\mu}$$

$$\implies \hat{\mu} = \frac{\delta}{v}.$$

So, our estimate of μ is the total number of deaths divided by the total time under observation.

Likelihood

Using the results for MLE estimators that were discussed earlier in the course we have that:

$$l'' = -\frac{\delta}{\mu^2},$$

and hence that asymptotically $\hat{\mu} \approx N(\mu, \frac{\mu^2}{E(\delta)})$. To use this result we replace μ with $\hat{\mu}$ and $E(\delta)$ with δ.

Example

From O'Neill Notes Example 5.1: We observe 10 individuals at some stage over the first 3 years of their life. Estimate the hazard in the second year of life and the estimated standard error. The a and b are the ages of entry and censoring in the first 3 years of life. Here T is the age at which observation stops, end event occurs.

Ind	T	a	b
1	2.3	1.7	2.3
2	1.2	0	3
3	1.5	1.1	1.5
4	.5	0	3
5	1.6	0	3
6	2.1	0	3
7	.6	0	2
8	3	0	3

```
9   2.4 1.5 2.4
10   .6  0  1
```

In the solution below, the definition of T is consistent with the definition in the notes, observation stops at age $x + T$.

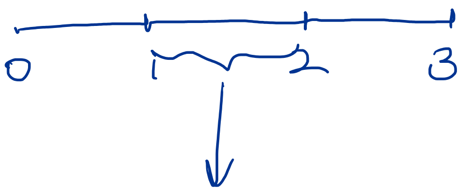
```
Ind  T  a  b  v  \delta
1    1 .7  1 .3  0
2   .2  0  1 .2  1
3   .5 .1 .5 .4  0
5   .6  0  1 .6  1
6    1  0  1  1  0
8    1  0  1  1  0
9    1 .5  1 .5  0
```

Explanation: individual one begins observation at 1.7 years, or during the second year of life. This is why a is .7 for this individual. Person observed until age 2.3, so relative to second year exposure they were exposed to $v = .3$ time units. This means year 2 exposure ended at second birthday, $x + b = 1 + 1 = 2$, the censoring time. No event was observed so $\delta = 0$ for this subject. Second year observation was fully observed so $x + T = 1 + 1 = 2$ for this observation makes time beyond $x = 1$ as $T = 1$ time unit.

The second individual begins second year observation at first birthday or $a = 0$, and is followed until the event at 1.2 units. This subject is under observation for $v = .2$ units, and would have been followed till second birthday if possible, thus $b = 1$. Subject is under observation till $x + T = 1 + .2 = 1.2$ time units. There was an ending event so $\delta_2 = 1$ for this subject.

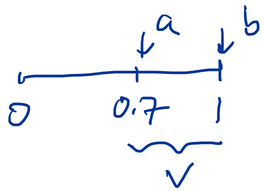
We have $\delta = 2$, $v = 4$, so $\hat{\mu} = \delta/v = 2/4 = .5$, and $se(\hat{\mu}) = \sqrt{\delta/v^2} = \sqrt{2/4^2}$.

Example: 3 year investigation



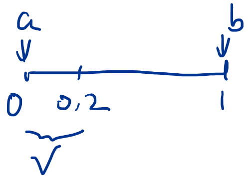
only interested in 2nd year of life

P1:



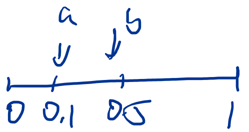
$a=0.7,\ b=1,\ T=1,\ \delta=0,\ v=0.3$

P2:



$a=0,\ b=1,\ T=0.2,\ \delta=1,\ v=0.2$

P3:



$a=0.1,\ b=0.5,\ T=0.5,\ \delta=0,\ v=0.4$

P4

⊗ omit: dies before second year

Nothing observed

Markov models - Multi-state

We will now consider markov models with more than two-states. To allow for the fact that there are multiple states we need to introduce the following notation for individuals aged $[x, x+1)$.

$g \to h$

- ${}_tp_x^{gh}$ is the probability a person in state g at age x is in state h at age $x+t$.

$g \to g$

- ${}_tp_x^{\overline{gg}}$ is the probability a person in state g at age x remains in that state continuously until age $x+t$.

note: ${}_tP_x^{gg} \neq {}_tP_x^{\overline{gg}}$

gg can be $g \to ? \to g$

$\overline{gg}$ is $g \to g \to g \to \cdots \to g$.

- μ^{gh} is the transition intensity (or hazard) between states g and h. We are still assuming that the hazard is constant over each year of age.

- v_i^g is the amount of time spent by individual i in state g.

- δ_i^{gh} is the number of transitions by individual i from state g into state h.

Kolmogorov Forward equations

The Kolmogorov forward equations will allow us to compute probabilities of interest from our estimated harzards or transition intensities.

$$\frac{d}{dt}{}_tp_x^{gh} = \sum_{r \neq h} ({}_tp_x^{gr} \mu(x+t)^{rh} - {}_tp_x^{gh} \mu(x+t)^{hr})$$

and

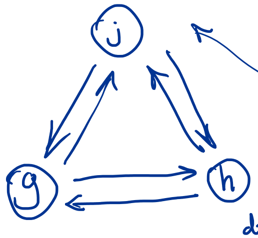
$$\frac{d}{dt}{}_tp_x^{\overline{gg}} = -{}_tp_x^{\overline{gg}} \sum_{r \neq g} \mu(x+t)^{gr}.$$

Note: The last equation gives us that ${}_tp_x^{\overline{gg}} = \exp(\int_0^t - \sum_{r \neq g} \mu(s)^{gr} ds)$. The results can further be simplified by remembering that the transition intensity is assumed constant.

Proof of Kolmogorov Forward equation

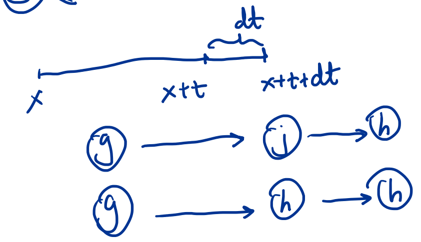
By law of total Probability:

$$ {}_{t+dt}P_x^{gh} = \sum_{j\neq h} {}_tP_x^{gj}\; {}_{dt}P_{x+t}^{jh} + {}_tP_x^{gh}\; {}_{dt}P_{x+t}^{hh} $$



other state than h, include g

j includes g



We also have $\quad {}_{dt}P_{x+t}^{hh} = 1 - \sum_{j\neq h} {}_{dt}P_{x+t}^{hj}$

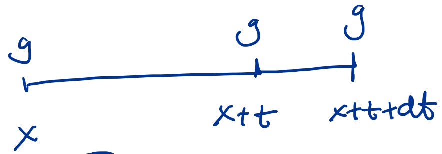
$$ \Rightarrow {}_{t+dt}P_x^{gh} = \sum_{j\neq h} {}_tP_x^{gj}\; {}_{dt}P_{x+t}^{jh} + {}_tP_x^{gh}\left(1 - \sum_{j\neq h} {}_{dt}P_{x+t}^{hj}\right) $$

Using assumption $\quad {}_{dt}P_{x+t}^{jh} \approx \mu_{x+t}^{jh}\cdot dt$

$$ {}_{t+dt}P_x^{gh} - {}_tP_x^{gh} \approx \sum_{j\neq h} {}_tP_x^{gj}\,\mu^{jh}\,dt - \sum_{j\neq h} {}_tP_x^{gh}\,\mu^{hj}\,dt $$

$$ \Rightarrow \lim_{dt\to 0} \frac{{}_{t+dt}P_x^{gh} - {}_tP_x^{gh}}{dt} \approx \sum_{j\neq h}\left({}_tP_x^{gj}\,\mu^{jh} - {}_tP_x^{gh}\,\mu^{hj}\right) \approx \frac{d\,{}_tP_x^{gh}}{dt} $$

Similarly



$$ {}_{t+dt}P_x^{\overline{gg}} = {}_tP_x^{\overline{gg}}\; {}_{dt}P_{x+t}^{\overline{gg}} $$

$$ \approx {}_tP_x^{\overline{gg}}\left(1 - \sum_{j\neq g} \mu^{gj}\cdot dt\right) $$

$$ \Rightarrow \frac{d\,{}_tP_x^{\overline{gg}}}{dt} = -\,{}_tP_x^{\overline{gg}} \sum_{j\neq g} \mu^{gj} $$

Markov models - Multi-state

The result ${}_tp_x^{\overline{gg}} = \exp(-t\sum_{r\neq g}\mu^{gr})$ provides insight into how to derive the likelihood for the multi-state model.

- The contribution to the likelihood due to the time spent in state g by individual i is $\exp(-v_i^g \sum_{r\neq g}\mu^{gr})$
- The contribution to the likelihood for each transition from state g into state h by individual i is ${}_tp_x^{\overline{gg}}\mu^{\delta_i^{gh}}$.

Using these facts the likelihood for the multi-state model is easily derived and the following result obtained:

$$\hat{\mu}^{gh} = \frac{\delta^{gh}}{v^g},$$

and

$$\hat{\mu}^{gh} \approx N(\mu^{gh}, \frac{\delta^{gh}}{(v^g)^2}).$$

Here, δ^{gh} is the total number (over all N individuals) of transitions from state g to state h and v^g is the total time spent in state g.

Deriving Kolmogorov Forward equations

$$\frac{\partial}{\partial t}\,{}_tp_x^{gh} = \sum_{r\neq h}({}_tp_x^{gr}\mu_{x+t}^{rh} -_t p_x^{gh}\mu_{x+t}^{hr})$$

Using the markov assumption we have

$${}_{t+dt}p_x^{gh} = \sum_{r\neq h}\,{}_tp_x^{gr}\;{}_{dt}p_{x+t}^{rh} + \,{}_tp_x^{gh}\;{}_{dt}p_{x+t}^{hh}$$

simple substitution yields that:

$${}_{t+dt}p_x^{gh} = \sum_{r\neq h}\,{}_tp_x^{gr}\;{}_{dt}p_{x+t}^{rh} + \,{}_tp_x^{gh}(1 - \sum_{r\neq h}\,{}_{dt}p_{x+t}^{hr}),$$

Now assuming that ${}_{dt}p_{x+t}^{hr} \approx \mu^{hr}dt$ (the second assumption stated earlier)

$${}_{t+dt}p_x^{gh} \approx \sum_{r\neq h}\,{}_tp_x^{gr}(\mu_{x+t}^{rh}dt) +_t p_x^{gh}\left(1 - \sum_{r\neq h}\mu_{x+t}^{hr}dt\right),$$

From here, the result can be shown by dividing through by dt and taking an appropriate limit.

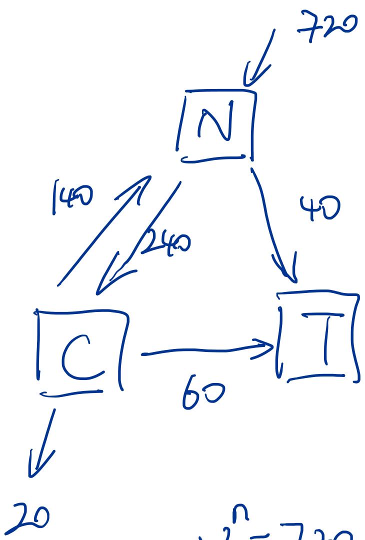
Example

Example 6.1, O'Neill Notes: (Hypothetical) In 1999, 800 first year students entered. The number of commencing counseling sessions was 240 (start, stop, start again is included here). The number of counseling sessions which ended was 140. The total time in counseling for the whole cohort in the year was 20 years, and the total time not in counseling was 720 years. During the year 60 students in counseling and 40 students not in counseling terminated their course. Estimate the transition intensities between states: N(not in counseling), C(in counseling), and T(terminated course). Estimate the probability an entering student will not seek counseling or terminate in the year. Also give 95 percent CI.

Example

The information provided tells us, $v^N = 720$, $v^C = 20$, $\delta^{NC} = 240$, $\delta^{CN} = 140$, $\delta^{NT} = 40$, $\delta^{CT} = 60$. The estimates of transition intensities are,

$$
\begin{aligned}
\hat{\mu}^{NC} &= \delta^{NC}/v^N = 240/720 = 1/3 \\
\hat{\mu}^{CN} &= \delta^{CN}/v^C = 140/20 = 7 \\
\hat{\mu}^{NT} &= \delta^{NT}/v^N = 40/720 = 1/18 \\
\hat{\mu}^{CT} &= \delta^{CT}/v^C = 60/20 = 3
\end{aligned}
$$



$v^n = 720$

$v^c = 20$

$\delta^{NC} = 240$

$\delta^{CN} = 140$

$\delta^{NT} = 40$

$\delta^{CT} = 60$

$$
\begin{aligned}
\hat{p}^{\overline{NN}} &= \exp\{-\hat{\mu}^{NC} - \hat{\mu}^{NT}\} \\
&= \exp\{-(1/3) - (1/18)\} \\
&= e^{-7/18} \\
&= .678
\end{aligned}
$$

The intensity estimates are independent (asymptotically) and thus the variance of a sum of independent random variables equals the sum of the estimated variances,

$$\mathrm{Var}(\hat{p}^{\overline{NN}}) = \mathrm{Var}(\hat{\mu}^{NC} + \hat{\mu}^{NT}) \cdot (\hat{p}^{\overline{NN}})^2 \quad \frac{\delta^{NC}}{v^{N^2}} + \frac{\delta^{NT}}{v^{N^2}},$$

and by the delta method, the estimated standard error for $\hat{p}^{\overline{NN}}$ is,

$$\hat{p}^{\overline{NN}} \sqrt{\frac{\delta_{NC}}{v^{N^2}} + \frac{\delta_{NT}}{v^{N^2}}} = (.678) * \sqrt{240/720^2 + 40/720^2} = .0158$$

This means the 95 percent confidence interval is $.678 \pm .031$.

Delta Method For Multivariate

(O'Neil's note, page 12)

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} \quad p\times 1 \qquad \mu_x = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}$$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{21} & \cdots & \sigma_{p1} \\ \sigma_{12} & \sigma_2^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \cdots & \cdots & \sigma_p^2 \end{bmatrix}$$

$$Var[g(X)] = \underset{1\times p}{\left(\frac{\partial g(X)}{\partial \mu_x}\right)^T} \underset{p\times p}{\Sigma} \underset{p\times 1}{\left(\frac{\partial g(X)}{\partial \mu_x}\right)}$$

In this example:

$$\underset{\substack{\downarrow \\ g(X)}}{\hat{p}^{\overline{NN}}} = e^{-\underset{\substack{\downarrow \\ x_1}}{\hat{\mu}^{NC}} - \underset{\substack{\downarrow \\ x_2}}{\hat{\mu}^{NT}}}$$

$$Var(\hat{p}^{\overline{NN}}) = \begin{pmatrix} -e^{-\hat{\mu}^{NC}-\hat{\mu}^{NT}} \\ -e^{-\hat{\mu}^{NC}-\hat{\mu}^{NT}} \end{pmatrix}^T \begin{bmatrix} Var(\hat{\mu}^{NC}) & 0 \\ 0 & Var(\hat{\mu}^{NT}) \end{bmatrix} \begin{pmatrix} \vdots \end{pmatrix}$$

$$= \begin{bmatrix} -e^{-\hat{\mu}^{NC}-\hat{\mu}^{NT}} \cdot Var(\hat{\mu}^{NC}) \\ -e^{-\hat{\mu}^{NC}-\hat{\mu}^{NT}} \cdot Var(\hat{\mu}^{NT}) \end{bmatrix} \begin{bmatrix} -e^{-\hat{\mu}^{NC}-\hat{\mu}^{NT}} \\ -e^{-\hat{\mu}^{NC}-\hat{\mu}^{NT}} \end{bmatrix}$$

$$= (\hat{p}^{\overline{NN}})^2 \left[Var(\hat{\mu}^{NC}) + Var(\hat{\mu}^{NT}) \right]$$