Transition and Duration Models

Introduction to Counting Processes: The Poisson Process

Christian Schluter

Counting processes

We will study a simple counting process, that will serve as a basic building block for duration analysis.

The counting process counts the number of randomly arriving events. The random new arrival will increment the counter by one, leading to a transition. The time between two subsequent randomly occurring transitions is a duration.

For a Poisson process, the increments are independent, and the durations have a very simple form.

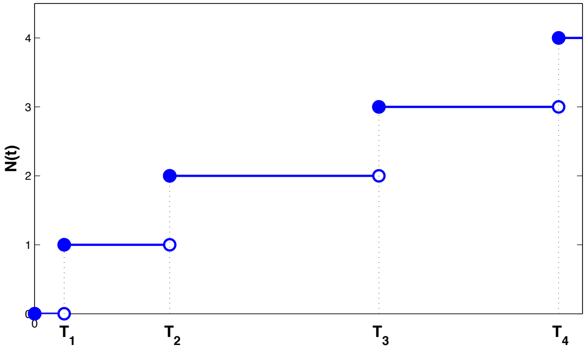
Consider a process $N(t)$ that counts the number of randomly arriving events as time t evolves continuously.

This **counting process** $\{N(t) : t \geq 0\}$ takes values in $S = \{0, 1, 2, ..\}$.

The number counted at time t, $N(t)$ is therefore a random variable, as is the time of the nth arrival T_n, as is the time between arrivals (the duration until a new arrival) $X_n = T_n - T_{n-1}$.

Once we observe a new arrival, the counter N increments by one.

The random variables $N(t)$, T_n, and X_n:



The Poisson process

The Poisson process is a particular counting process that has three defining features:

1. Since it is a counting process, it starts at 0, and can only increase; it cannot decrease.
2. The probability of observing more than one increment in a short space of time is negligible.
3. Increments are independent.

Formally: The **Poisson process** with intensity $\lambda > 0$ is the **counting process** $\{N(t) : t \geq 0\}$ taking values in $S = \{0, 1, 2, ..\}$ such that:

1. $N(0) = 0$; if $s < t$ then $N(s) \leq N(t)$
2. $$\Pr\{N(t+h) = n + m | N(t) = n\} = \begin{cases} \lambda h + o(h) & \text{if } m = 1 \\ o(h) & \text{if } m > 1 \\ 1 - \lambda h + o(h) & \text{if } m = 0 \end{cases}$$
3. if $s < t$ then the number $N(t) - N(s)$ of counted event during the interval $(s, t]$ is independent of the number of counted events during $[0, s)$.

The little-o notation captures the idea that a quantity is negligible and being dominated by anther negligible quantity. We can define this formally.

Definition: A function $g(h)$ is said to be $o(h)$, $g(h) = o(h)$, if $g(h)/h \to 0$ as $h \to 0$.

He have defined that the probability of observing more than one occurence in the interval $(t, t+h]$ is $o(h)$. It follows that for the Poisson process the term $o(h)$ must decline to zero faster than h.

The three features of the Poisson process are sufficient to fully characterise the distribution of $N(t)$.

Theorem: If $\{N(t) : t \geq 0\}$ follows the Poisson process, then $N(t)$ has the Poisson distribution with parameter λt:

$$\Pr\{N(t) = j\} = \frac{(\lambda t)^j}{j!} e^{-\lambda t}, \qquad j = 0, 1, 2, \ldots$$

Proof

Why is this process called Poisson ?

Recall the following definition of the **Poisson** distribution.

Definition: The discrete random variable X follows the Poisson law with parameter $\lambda > 0$ if its density is given by

$$f(k;\lambda) = \Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!},$$

for $k = 0, 1, 2, \ldots$.

It can be shown that $\mathrm{E}(X) = \mathrm{Var}(X) = \lambda$. This is an important limitation. In many applications, the variance of count data is larger than the mean; the data are said to be over-dispersed (below, we will briefly consider the Negative Binomial model).

Let us briefly revisit the second property of the Poisson process, i.e. the probability of observing more than one occurence in the interval $(t, t+h]$ being $o(h)$: the term $o(h)$ must decline to zero faster than h.

We can verify this using our theorem:

$$\begin{aligned}\Pr\{\text{more than 1 occurence in}(t, t+h]\} &= \sum_{j=2}^{\infty} \Pr\{N(h) = j\} \\ &= \sum_{j=2}^{\infty} \frac{(\lambda h)^j}{j!} e^{-\lambda t} \\ &= o(h)\end{aligned}$$

Since the Poisson process is a counting process with independent increments, it should be intuitively clear that the sum of two independent Poisson processes is also a Poisson process:

Result': If $\{N_i(t), t \geq 0\}$ are independent Poisson processes with rate λ_i, $i = 1, 2$, then $\{N(t) = N_1(t) + N_2(t), t \geq 0\}$ is a Poisson process with rate $\lambda_1 + \lambda_2$.

Proof: Exercise set 1.

Durations

Consider the duration between two subsequent arrivals.

Formally, the **time of the nth arrival** T_n is given by

$$T_n = \inf \{t : N(t) = n\}.$$

with $T_0 = 0$. In turn, we have $N(t) = \max\{n : T_n \leq t\}$.

The **inter-arrival times** are defined by $X_n = T_n - T_{n-1}$.

The following equivalence is often useful

$$N(t) \geq j \text{ if and only if } T_j \leq t.$$

For instance, we have

$$\begin{aligned}
\Pr\{N(t)=j\} &= \frac{(\lambda t)^j}{j!}e^{-\lambda t} \\
&= \Pr\{T_j \le t < T_{j+1}\} \\
&= \Pr\{T_j \le t\} - \Pr\{T_{j+1} < t\}
\end{aligned}$$

This follows since

$$\Pr\{T_j \le t\} = \Pr\{T_j \le t, T_{j+1} > t\} + \Pr\{T_j \le t, T_{j+1} < t\},$$

and $\Pr\{T_j \le t, T_{j+1} < t\} = \Pr\{T_{j+1} < t\}$.

The inter-arrival times of the Poisson process follow an **exponential** law.

Theorem: $X_1, X_2, ...$ are independent exponential random variables with parameter λ.

Recall the definition of the **exponential** distribution.

Definition: The continuous and positively valued random variable X follows the exponential law with parameter $\lambda > 0$ if its density is given by

$$f(x) = \lambda e^{-\lambda x} I_{(0,\infty)}(x)$$

The cumulative distribution function (cdf)

$$F(x) = 1 - e^{-\lambda x} \qquad x > 0,$$

It can be shown that $E(X) = \lambda^{-1}$ and $var(X) = \lambda^{-2}$.

Proof. We proceed inductively, and start with X_1:

$$\Pr\{X_1 > t\} = \Pr\{N(t) = 0\} = e^{-\lambda t}.$$

For the analysis of X_2 condition on $X_1 = t_1$.

$$\begin{aligned}\Pr\{X_2 > t | X_1 = t_1\} &= \Pr\{\text{no arrival in } (t_1, t_1 + t] | X_1 = t_1\} \\ &= \Pr\{\text{no arrival in } (t_1, t_1 + t]\}\end{aligned}$$

since, using Property 3, the event $\{X_1 = t_1\}$ which relates to time interval $[0, t_1]$ is independent of the event $\{\text{no arrival in } (t_1, t_1 + t]\}$ as this relates to $(t_1, t_1 + t]$.

Therefore

$$\begin{aligned}\Pr\{X_2 > t | X_1 = t_1\} &= \Pr\{\text{no arrival in } (t_1, t_1 + t]\} \\ &= e^{-\lambda t},\end{aligned}$$

and X_2 is independent of X_1.

Similarly

$$\begin{aligned}\Pr\{X_{n+1} > t | X_1 = t_1, \ldots, X_n = t_n\} &= \Pr\{\text{no arrival in } (T, T + t]\} \\ &= e^{-\lambda t}.\end{aligned}$$

where $T = t_1 + \cdots + t_n$. $\{X_i\}$ are iid.

Let $\{X(t), t \geq 0\}$ be a Poisson process with rate parameter $\lambda = 0.7$, and let T_k be the time of the kth event. Compute:

1. $\Pr(X(4) = 5 | X(3.5) = 4)$,
2. $\Pr(X(4) = 4 | X(3.5) = 5)$,
3. $\Pr(T_1 < 5)$,
4. $\Pr(T_3 < 5 | T_2 = 3.5)$,
5. $\Pr(X(7) = 8 | T_3 = 6)$,
6. $\Pr(X(7) = 8 | T_9 = 6)$.

The Markov property, and Markov chains

The **Markov property** refers to the loss of memory of a process.

Since the Poisson process has independent increments, it follows that the process will satisfy the Markov property.

We will then return to the jumps of the Poisson process, which we can think of as a study of transition probabilities.

Before we do this, we look more systematically at Markovian processes.

Let $X = \{X(t) : t \geq 0\}$ be a family of random variables taking values in some countable state space S.

Definition. The process X satisfies the **Markov property** if

$$\begin{aligned} & \Pr(X(t_n) = j | X(t_1) = i_1, X(t_2) = i_2, \ldots, X(t_{n-1}) = i_{n-1}) \\ = \quad & \Pr(X(t_n) = j | X(t_{n-1}) = i_{n-1}). \end{aligned}$$

Conditional on the most recent position, the preceding history is irrelevant.

If X satisfies the Markov property, it is called a (continuous time) **Markov chain**.

The **transition probability** $p_{ij}(s,t)$ is defined as

$$p_{ij}(s,t) = \Pr(X(t) = j | X(s) = i) \qquad \text{for } s \leq t.$$

The chain is called **homogeneous** if

$$p_{ij}(s,t) = p_{ij}(0, t-s), \qquad \text{for all } i, j, s, t$$

so the transition probability only depends on the gap $t-s$ and not on calendar time.

To simplify notation write $p_{ij}(t-s)$ for $p_{ij}(0, t-s)$.

We have the following result:

Result: The Poisson process $\{N(t) : t \geq 0\}$ is a homogenous Markov Chain with state space $S = \{0, 1, \dots\}$ and transition probability $p_{i(i+k)}(t)$ given by

$$p_{i(i+k)}(t) = \Pr(X(t) = i + k | X(0) = i) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

The Poisson process has stationary independent increments, since (i) the distribution of $N(t) - N(s)$ depends only on $t - s$, and (ii) the increments $\{N(t_i) - N(s_i)\}$ are independent if $s_1 \leq t_1 < s_2 \leq t_2$.

Poisson processes are the only **renewal** processes that are Markov chains.

Definition: A **renewal process** $N = \{N(t) : t \geq 0\}$ is a process for which

$$N(t) = \max\{n : T_n \leq t\}$$

where $T_0 = 0$, $T_n = \sum_{i=1}^{n} X_i$, and X_i are iid non-negative random variables.

For the Poisson process, X_i has a particular interpretation, namely a duration (or inter-arrival time).

Suppose we interrupt the renewal process N at some specified time s. By this time $N(s)$ occurrences have already taken place and we are waiting for the $(N(s)+1)^{th}$; i.e. s belongs to the random interval

$$I_s = [T_{N(s)}, T_{N(s)+1}).$$

- The excess or **residual lifetime** of I_s is:

$$Ex(s) = T_{N(s)+1} - s.$$

- the current lifetime, or age, or **elapsed time** of I_s:

$$C(s) = s - T_{N(s)}$$

- The total lifetime of I_s: $D(s) = Ex(s) + C(s)$.

Consider the excess or residual lifetime of a Poisson process $Ex(s)$.

Theorem: $E\{Ex(s)\} = \lambda^{-1}$.

Proof. N is a Markov chain, so the distribution of $Ex(s)$ does not depend on arrivals prior to s.

Thus $Ex(s)$ has the same mean as $Ex(0) = T_1 = X_1$.

We have already shown that X_i are iid and follow the exponential law with parameter λ. The mean of X_1 is λ^{-1}.

Poisson Regressions

Poisson models are estimated using Maximum Likelihood (ML).

Revision

Let $Y_i \sim^{i.i.d} Poi(\lambda_i)$.

Recall the Poisson density (with parameter $\lambda > 0$) is

$$f(y_i; \lambda) = \Pr(Y_i = y_i) = \frac{\lambda^{y_i} e^{-\lambda}}{y_i!},$$

for $y_i = 0, 1, 2, \ldots$.

Since we have a random sample, the likelihood is

$$L = \prod_i f(y_i; \lambda_i)$$

The log likelihood function is then simply

$$\log L = \sum_i y_i \log(\lambda_i) - \lambda_i$$

Assume that the individual-specific mean $\lambda_i > 0$ varies with observables x_i.

To ensure that the mean is positive, it is customary to specify a log-linear model

$$\log \lambda_i = x_i' \beta$$

Recall that the mean of a Poisson RV Poi(λ) is simply λ.

Write $\lambda = E(Y|X)$.

In our model, what is the marginal effect of a change in the control variable / regressor $x_{i,j}$?

Since $\log \lambda_i = x_i'\beta$, the marginal effect on the conditional mean is

$$\frac{dE(Y|x)}{dx_j} = \exp(x'\beta)\beta_j = E(Y|x)\beta_j,$$

so

$$\beta_j = \frac{E(Y|x)/dx_j}{E(Y|x)}$$

is a proportional marginal change.

We will now show that $(\exp(\beta_j) - 1)$ is the proportionate change in the conditional mean by a *unit* change in x_j. We have

$$\frac{E(Y|x=1) - E(Y|x=0)}{E(Y|x=0)} = \frac{E(Y|x=1)}{E(Y|x=0)} - 1.$$

Using the preceding result $\beta_j dx_j = \frac{E(Y|x)}{E(Y|x)}$, and integrating this from 0 to 1, we obtain

$$\log E(Y|x=1) - \log E(Y|x=0) = \log \frac{E(Y|x=1)}{E(Y|x=1)} = \beta_j$$

and hence the claimed result.

We will use this result below where x_j is a dummy variable indicating a policy change.

All this is implemented in R in the function `glm` (short for generalised linear model). We need to specify the empirical model for the log of the mean of the process $\log \lambda_i$.

Example: The number of doctor visits.

The dataset `mdvisitsx` is a cross-sectional subsample from the GSOEP, which collected data on doctor visits before and after a major health care reform that took place in 1997.

The reform increased the copayments for prescription drugs by up to 200% and imposed upper limits on the reimbursement of physicians by the state insurance.

The outcome is the number of doctor visits in a three month period. The predictors of interest are `reform`, a dummy variable that takes the value 1 after the reform and 0 before, age in years, `education` in years, a dummy variable for `bad health`, and the log of `income`.

We fit a Poisson regression model to estimate the effect of the reform.

```
library(foreign)

md = read.dta("mdvisitsx.dta")
mp = glm(numvisit ~ reform+age+educ+badh+loginc, family=poisson, data=md)
summary(mp)

               Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.1742048  0.3167720  -0.550 0.582362
reform      -0.2273629  0.0315264  -7.212 5.52e-13 ***
age          0.0049815  0.0014733   3.381 0.000722 ***
educ        -0.0006806  0.0068735  -0.099 0.921127
badh         1.1726352  0.0352552  33.261  < 2e-16 ***
loginc       0.1119396  0.0427068   2.621 0.008764 **


## The coefficient of reform shows a 20.3\% (= 100 * (exp(-0.227) -1))
## reduction in the number of visits when we compare subjects with the
## same age, education, health status and income.
```

The Poisson model imposes the restriction that the mean and the variance of the Poisson variable are the same (λ). In many applications, the variance of count data is larger than the mean; the data are said to be over-dispersed.

An alternative count model assumes that the count variable follows a Negative Binomial law. The conditional mean and variances are

$$\begin{aligned} E(Y|X) &= \lambda \\ V(Y|X) &= \lambda + \alpha\lambda^2 \end{aligned}$$

with $\lambda = \exp(x'\beta)$. α is the over-dispersion parameter.

The negative binomial regression is implemented in `R` in the function `glm.nb` in the `MASS` library.

```
require(MASS)
mnb = glm.nb(numvisit ~reform+age+educ+badh+loginc, data=md)
summary(mnb)

              Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.081726   0.625016  -0.131 0.895967
reform      -0.215337   0.061841  -3.482 0.000497 ***
age          0.006624   0.002911   2.275 0.022894 *
educ         0.009255   0.013353   0.693 0.488232
badh         1.166646   0.088645  13.161  < 2e-16 ***
loginc       0.076613   0.084642   0.905 0.365389

1/mnb$theta ## the overdispersion parameter.
##  1.00752
```

Example. The recidivism data. The individuals are convicts released from prison. We consider the duration (in months) from release to re-arrest.

The data are individual person-months observations. Partition duration into J intervals with cutpoints $(0 = \tau_0 < \tau_1 < \ldots < \tau_J = \infty)$. This defines the j-th interval $[\tau_{j-1}, \tau_j)$.

So far, we have considered a Poisson process with parameter λ. A more flexible models allows the parameter of the Poisson process to change over these intervals, so λ_j is the parameter for the individual in interval j. We will return later to such *duration dependence*.

For later reference we will call this the **Piece-Wise Exponential (PWE)** model.

Define measures of exposure as follows. Let t_{ij} denote the time lived by the i-th individual in the j-th interval, that is, between τ_{j-1} and τ_j.

(i) If the individual lived beyond the end of the interval, so that $t_i > \tau_j$, then $t_{ij} = \tau_j - \tau_{j-1}$.

(ii) If the individual died or was censored in the interval, i.e. if $\tau_{j-1} < t_i < \tau_j$, then the timed lived in the interval is $t_{ij} = t_i - \tau_{j-1}$.

Define the death indicators. Let d_{ij} take the value one if individual i dies in interval j and zero otherwise.

We treat the death indicators d_{ij} as if they were independent Poisson observations with means

$$\mu_{ij} = t_{ij}\lambda_{ij},$$

where t_{ij} is the exposure time, and λ_{ij} equals

$$\lambda_{ij} = \lambda_j \exp\{\boldsymbol{x}_i'\boldsymbol{\beta}\},$$

In later sections we will refer to this as the hazard for individual i in interval j.

The likelihood contribution is:

$$\log L_{ij} = d_{ij} \log \mu_{ij} - \mu_{ij} = d_{ij} \log(t_{ij}\lambda_{ij}) - t_{ij}\lambda_{ij}.$$

```
load("recidx.RData")

## examine the data structure:
## duration=25 months for individual with id=4
recidx[(recidx$id == 4),c("id","durat", "exposure","interval","fail")]

     id durat exposure interval fail
4     4    12       12   (0,12]    0
1448  4    24       12  (12,24]    0
2710  4    25        1  (24,36]    1

mf2 <- fail ~ interval + workprg + priors + tserved + felon + alcohol +
               drugs + black + married + educ + age
```

```
summary(glm(mf2, offset = log(exposure), data = recidx, family = poisson))
                 Estimate Std. Error z value Pr(>|z|)

(Intercept)     -3.8301275  0.2802673 -13.666  < 2e-16 ***
interval(12,24]  0.0365320  0.1093618   0.334  0.73834
interval(24,36] -0.3738156  0.1296119  -2.884  0.00393 **
interval(36,48] -0.8115436  0.1564015  -5.189 2.12e-07 ***
interval(48,60] -0.9382311  0.1683212  -5.574 2.49e-08 ***
interval(60,81] -1.5471779  0.2033489  -7.608 2.77e-14 ***
workprg          0.0838291  0.0907942   0.923  0.35586
priors           0.0872458  0.0134735   6.475 9.46e-11 ***
tserved          0.0130089  0.0016859   7.716 1.20e-14 ***
felon           -0.2839252  0.1061488  -2.675  0.00748 **
alcohol          0.4324425  0.1057211   4.090 4.31e-05 ***
drugs            0.2747141  0.0978635   2.807  0.00500 **
black            0.4335560  0.0883623   4.907 9.27e-07 ***
married         -0.1540477  0.1092119  -1.411  0.15838
educ            -0.0214162  0.0194440  -1.101  0.27071
age             -0.0035800  0.0005222  -6.855 7.13e-12 ***
## the risk of recidivism is about the same in the first two years,
##   but then decreases substantially with duration since release.
## Subjects imprisoned for alcohol or drug related offenses have much higher
##   risk of recidivism, everything else being equal.
```

Theorem: If $\{N(t) : t \geq 0\}$ follows the Poisson process, then $N(t)$ has the Poisson distribution with parameter λt:

$$\Pr\{N(t) = j\} = \frac{(\lambda t)^j}{j!} e^{-\lambda t}, \qquad j = 0, 1, 2, \ldots$$

Proof. To prove the theorem condition $N(t+h)$ on $N(t)$:

$$\begin{aligned}
Pr(N(t+h) = j) &= \sum_i P(N(t) = i) P(N(t+h) = j | N(t) = i) \\
&= \sum_i P(N(t) = i) P((j-i) \text{ arrivals in } (t, t+h]) \\
&= P(N(t) = j-1) P(\text{ one arrival }) \\
&\quad + P(N(t) = j) P(\text{ no arrival }) + o(h) \\
&= P(N(t) = j-1) \lambda h \\
&\quad + P(N(t) = j)(1 - \lambda h) + o(h).
\end{aligned}$$

This equation can be re-written more compactly using the notation $p_j(t) = P(N(t) = j)$.

We have

$$p_j(t+h) = \lambda h p_{j-1}(t) + (1 - \lambda h) p_j(t) + o(h)$$

if $j \neq 0$. Subtract $p_j(t)$, divide by h, and let $h \to 0$ to get the equation

$$p_j'(t) = \lambda p_{j-1}(t) - \lambda p_j(t).$$

This is a differential difference equation.

Similarly, for $j = 0$ we have $p_0(t+h) = (1 - \lambda h) p_0(t) + o(h)$, leading to

$$p_0'(t) = -\lambda p_0(t).$$

To solve this system, start with solving

$$p_0'(t) = -\lambda p_0(t)$$

subject to $p_0(0) = 1$, to get

$$p_0(t) = e^{-\lambda t}.$$

Then substitute this into $p_j'(t) = \lambda p_{j-1}(t) - \lambda p_j(t)$ for $j = 1$, and solve to get

$$p_1(t) = \lambda t e^{-\lambda t}.$$

and continue inductively for $j = 2, \ldots$.

Return

Consider a random sample $X_1, .., X_n$ from a known density f with unknown parameters θ. We seek to estimate this unknown parameter(s) θ.

The likelihood is the joint density of the sample evaluated at the observed values

$$L(\theta) = f_{X_1 \dots X_n}(x_1, \dots, x_n; \theta)$$

By independence, the joint density factorises, and using the fact that all X_i are identically distributed we have

$$\begin{aligned} L(\theta) &= f_{X_1 \dots X_n}(x_1, \dots, x_n; \theta) = f_{X_1}(x_1; \theta) \cdot \dots \cdot f_{X_n}(x_n; \theta) \\ &= \prod_{i=1}^{n} f_X(x_i; \theta) \end{aligned}$$

Since f is assumed known, this is simply a function of the unknown parameters.

It can be shown that the resulting estimator obeys the Normal law:

$$\sqrt{n}\left(\widehat{\theta}-\theta_0\right) \to^{d} \mathcal{N}\left(0, A_0^{-1}\right)$$

where $A_0 = -E\left(H_i\left(\theta_0\right)\right)$,i.e.

$$\mathcal{I}(\theta) = -\operatorname{E}\left[\left.\frac{\partial^2}{\partial\theta^2}\log f(X;\theta)\right|\theta\right] = -\operatorname{E}[H_i(\theta)].$$

The variance of the score $\left(s_i\left(\theta\right) = \nabla_\theta l_i\left(\theta\right)^{\mathsf{T}}\right)$ is often called the "Fisher information" $\mathcal{I}(\theta)$, which can be linked to the expected Hessian, using the information matrix equality (provided l_i is twice differentiable).

Return