John's Session I starts here.

Single State Duration Models without Unobserved Hetrogeneity.

Applications of Hazard Models that are of interest to economists and sociologists:

1. Time until first birth
2. Time until marriage for single individuals
3. Time until divorce for married couples
4. Time until relapse for graduates of an alcohol/drug treatment program

5. Time until return to jail for those released from jail
6. Time until failure for a new firm
7. Time until retirement
8. Time until entering employment for un employed
9. Time until drilling for oil on a site that a firm has an option on.

What questions do we ask in hazard model analysis

- What is the probability in a given week that one leaves a state (e.g. unemployment) given you have been in a state for t weeks
- Or (equivalently) what is the probability an individual will stay in a state for T weeks
- What is the expected length of a spell and how does it change with a change in the explanatory variable – an important policy effect.

- How does the probability of leaving change with the length of the spell – duration dependence - important for interventions.
- In a two state model – e.g. employment and nonemployment - What fraction of time does an individual spend in employment and how does it change with a change in the explanatory variable – a second important policy effect.
- Duration models have a reputation for being difficult, and it is certainly true that one's intuition from a linear model does not carry over well.
- However, if one builds the models up one step at a time, they become much more accessible. On my website I list papers by former students and their papers often involve duration models, and are published in good journals.

- We will want to look at transition rates at given durations and given calendar times – they won't be equivalent unless everyone enters the state at the same calendar time.
- We especially want to worry about time changing explanatory variables – unemployment rate in a study of employment or unemployment duration. This is not done in much European empirical work right now, but there is really no reason for this omission.

- Why the emphasis on time changing explanatory variables - if we don't have them, we could simply run regressions/Tobit Models where the Tobit Index function is

$$Dur_i^* = X_i \beta + \varepsilon_i$$

and the relation to the true data is given by

$$Dur_i^* = Dur_i \quad \text{for complete spells}$$
$$Dur_i^* > Dur_i \quad \text{for incomplete spells}$$

(Incomplete spells occur if someone runs out of benefits and drops out of the data, or in survey data if they are still in the spell at the end of the period.)
I am sure you all know how to estimate a Tobit model in Stata. However, one cannot use Tobit Model do this and let the explanatory variables change over the spell

- A second, but in my view less strong, motivation for Hazard models is that they may be more closely linked to theory – Example of a simple search model in unemployment. Here the probability someone leaves unemployment after a spell of t days is

$\lambda(t) = \mu(t)(1 - F(w^r(t))$

where $F(w(t))$ is the distribution function for wages

and $w^r(t)$ is the reservation wage.

How $\lambda(t)$ changes with t may let us learn about $\mu(t)$ and $w^r(t))$

simple e.g. set $\mu(t) = 1$ and learn about how $\mathrm{w}^r(t)$

changes with time

- We will work in discrete time. Why? First it is much more intuitive – our goal is to take away impression that duration models are inherently very hard or complicated.

- Second, As we discuss below transitions (outside financial markets) are observed in discrete time (e.g. German socio-economic panel where duration is monthly). Further, time changing explanatory variables change discretely (e.g. by month or quarter). The advantages of continuous time disappear once one has to account for these problems.

(Simplest) Case 1: λ as constant across time and individuals.

- The probability of a transition is given by

  $\lambda = 1/(1+e^{-\delta})$.

We use logit structure to insure that hazard stays between 0 and 1. (It's easy to change the functional form to, e.g., probit.) Note that this does not depend on length of time in a spell.

- **We will want to ask**

1. Probability that a spell lasts T weeks (density function)

$$\Pr(Dur = T) = \lambda(1-\lambda)^{T-1}$$

2. Prob a spell is still ongoing after K weeks (Survivor function)

$$\text{Prob (Dur>K)}=(1-\lambda)^{K}$$

(Note: in discrete time $\text{Prob (Dur>K)} \neq \text{Prob (Dur} \geq K).$)

3. How long will a spell last on average (expected duration)

$$E(Dur) = \lambda^{-1} = (1+e^{-\delta}).$$

Comparison to a continuous time Hazard

$\lambda = e^{\theta}$

(logit converges to this as time period approaches 0 in length.)

1. Density for a spell lasting exactly t periods

$$h(t) = \lambda \exp\{-\int_0^t \lambda dr\}$$

2. Survivor function

$$S(t) = \exp\{-\int_0^t \lambda dr\}$$

3. Expected duration

$$E(dur) = \int_0^\infty \tau f(\tau) d\tau.$$

But in data like the German Socio-Economic Panel we observe that spell ends between month T and T+1. The probability of this event is

$$\Pr(T \leq dur \leq T+1) = \Pr(dur \geq T) - \Pr(dur \geq T+1) = S(T) - S(T+1),$$

This will be the contribution of the spell to the likelihood, and the continuous time model starts to look like the discrete time model; sometimes people use the approximation that the transition takes place at time T+.5. This grouping can be incorporated in Stata programs for estimating continuous time hazard models (which have a number of limitations noted below).

Case 2: differences among individuals

Let $\lambda_i = \dfrac{1}{1+\exp\left(-\alpha Z_i\right)}$;

note *if* $\alpha_k > 0,\ \text{then}\ \dfrac{\partial \lambda_i}{\partial X_k} > 0$

Why would we expect the hazard to differ across individuals

- arrival rate may be higher for highly educated men, workers in their 20's and 30's, white vs. black men, men vs. women.

Continuous time analogue $\lambda_i = e^{\delta X_i}$.

Again

1. Prob a spell lasts T weeks (density function)

$$\Pr(Dur_i = T) = h(t) = \lambda_i (1-\lambda_i)^{T-1}$$

2. Prob a spell is still ongoing after K weeks (Survivor function)

$$\text{Prob (Dur}_i \text{>K)=}(1-\lambda_i)^K$$

3. How long will a spell last on average (expected duration)

$$E(Dur) = \lambda_i^{-1} = 1 + \exp\left(\alpha Z_i\right).$$

Note: Expected Duration for the sample

$$ED^* = \frac{1}{N}\sum_{i=1}^{N} Dur_i = \frac{1}{N}\sum_{i=1}^{N} 1 + \exp\left(\alpha Z_i\right) \neq 1 + \exp\left(\alpha \frac{1}{N}\sum_{i=1}^{N} Z_i\right).$$

Policy Effect of increasing one of the X's by one unit

$\frac{\Delta \mathrm{E}(\mathrm{dur}_i)}{\Delta X_i}$. We often will take the derivative numerically.

Expected Duration for the sample

$$\frac{\Delta \mathrm{ED}^*}{\Delta X} = \frac{1}{N}\sum_{i=1}^{N}\frac{\Delta \mathrm{E}(\mathrm{dur}_i)}{\Delta X_i}.$$

Much more informative that simply looking at the hazard coefficient.

Note since there are no unobservables, there is no possibility of an endogenous variable. Also it doesn't make sense to cluster the by individual when calculating the standard errors. The problem with clustering is that ignoring unobserved heterogeneity in estimation will lead to inconsistent parameter estimates, so it doesn't make sense to account for it in calculating the standard errors.

Case 3: The hazard depends on calendar time

$\lambda_i(t) = \dfrac{1}{1+\exp(-Z_{i\tau}\eta)}$, $\tau = \tau_0 + t$ and

τ_0 is the calendar time start date of the spell.

Continuous time analogue $\lambda_i\left(t\right)=e^{Z_{i\tau}}\quad \tau=\tau_0+t$.

Why might the hazard function depend on calendar time?

Variables such as the unemployment rate in employment or unemployment duration. Monte carlo work indicates that ignoring the time changing nature of variables such as the unemployment rate biases the effect of the business cycle substantially. Note we are not considering time changing X's that are controlled (to some extent) by the individual – e.g. remaining time eligible for benefits. We need time changing X's that are external to the model, and these will help in identification.

We have

1. Prob a spell lasts T weeks (density function)

$$\Pr(Dur_i = T) = h_i(t) = \lambda_i(t+\tau_0)\prod_{r=1}^{t-1}(1-\lambda_i(r+\tau_0))$$

2. Prob a spell is still ongoing after K weeks (Survivor function)

$$\text{Prob (Dur}_i > \text{K)} = \prod_{r=1}^{K}(1-\lambda_i(r+\tau_0))$$

3. How long will a spell last on average (expected duration)

$$\text{E(dur}_i\text{)} = \sum_{r=1}^{\infty} r f_i(r).$$

One issue – what to set the time changing X's to once one gets out of the sample period.

What happens in continuous time with time changing explanatory variables.

Suppose $x_{it+\tau_0}$ changes only once at duration time t_1, calendar time $t_1+\tau_0$. Then even if outcome is continuously measured, the contribution to the likelihood for a completed spell

$$f(t)=\lambda_i(t+\tau_0)\exp\{-\int_0^{t_1}\lambda(r+\tau_0)dr+\int_{t_1}^{t}\lambda(r+\tau_0)dr\}.$$

Again the continuous time approach loses it's nice form. Stata's program will let you break up the interval a limited number of times, but it's very easy to make a mistake here. With general duration dependence it gets even trickier. *With discrete time you take care of this in the data step.*

Case 4: In discrete time let λ_i *depend on duration and calendar time*

$$\lambda_i(t) = \frac{1}{1+\exp\left(-Z_{i\tau}\eta - f(t)\right)}$$

Why might λ_i depend on duration?

Define

negative duration dependence - $\partial\lambda_i(t)/\partial t < 0$

positive duration dependence $\partial\lambda_i(t)/\partial t > 0$.

Why might we expect negative duration dependence in unemployment duration?

- Discouraged workers, negative signaling to employers.

Why might we expect positive duration dependence?

Assets dwindling, learning about the wage distribution and shifting down reservation wages, running out of benefits, being subject to 'sticks' (penalties) if they stay unemployed too long.

The relevant formulae are no more complicated than with time changing X's. Big question is how to specify duration dependence function $f(t)$. Early studies took simple functional forms for $f(t)$ – linear in t, quadratic in t, or t^{α}. However, you will want to be more flexible. Can either use a polynomial in t or a step function in t. Latter seems to be better if you are only considering one type of spell, the former may be better if you are estimating the parameters of several types of spells simultaneously since it tends to lead to more parsimonious specifications.

Important point: suppose you have weekly data, and the highest duration you can see is 52 weeks. Temptation put in 52 dummies for duration. That will lead to very noisy estimates and make it impossible to allow for unobserved heterogeneity. Instead make sure each dummy corresponds to 3-5% of the transitions. Suppose we choose 3% i.e, make a dummy for 30-35 weeks if 3% of the transitions occur in these weeks.

Relevant formulae

1. Prob a spell lasts T weeks (density function)

$$\Pr(Dur_i = T) = h_i(t) = \lambda_i(t+\tau_0)\prod_{r=1}^{t-1}(1-\lambda_i(r+\tau_0)).$$

2. Prob a spell is still ongoing after K weeks (Survivor function)

$$\text{Prob (Dur}_i\text{>K)}=\prod_{r=1}^{K}(1-\lambda_i(r+\tau_0)).$$

3. How long will a spell last on average (expected duration)

$$\text{E(dur}_i\text{)}=\sum_{r=1}^{\infty} r f_i(r).$$

One possible problem – a defective distribution occurs if the duration dependence drives the hazard to 0 for t>t*, the expected duration won't exist. A couple of alternatives – use median duration, or use a truncuated duration:

$$\text{TrunE}(\text{dur}_i) = \sum_{r=1}^{T*} r f_i(r) + S_i(T*)T*.$$

In discrete time one still uses a logit program, the only difference across the models is in the data step.

It is tricky to have general duration dependence and time changing X's in continuous time, but not to have general duration dependence and only time constant X's in continuous time.

Up to this point estimation is straight-forward as long as you set up the data as Xianghong will suggest, and there is no identification problem.

Turnover to Xianghong

Xianghong's first session starts here.

I. Data management

A recommended STATA book:

Microeconometrics Using Stata, by Colin Cameron and Pravin Trivedi (revised edition)

Download data sets in the book:

Open Stata when internet in connected.

. net install mus

. net get mus

Search where the files installed in your computer and put into a folder of your choice.

I stored the data sets under:

C:\DurationWorkshop\MUS

Data Format Needed to Run a Discrete Time Duration Model in Logistic Functional Form

Logistic discrete time hazard function

$$\lambda_i(t) = \frac{1}{1 + \exp\left(-Z_{it}\eta - f(t)\right)}$$

Where Z_{it} is a vector of (possibly) time-changing explanatory variables, η is a vector of parameters, and $f(t)$ represents duration dependence.

In a single spell setting, individuals with a completed spell of has K periods:

$$L_i(K) = \lambda_i(K)\prod_{t=1}^{K-1}\left(1-\lambda_i(t)\right)$$

Individuals with a right-censored (as John discussed) spell of length K:

$$L_i(K) = \prod_{t=1}^{K}\left(1-\lambda_i(t)\right)$$

- This structure of likelihood function implies we need a panel data (or long form data, the panel is not balanced) to estimate a discrete time duration model.

- Usually survey data come in wide-form, each observation occupies one line.

- For discrete time model, data preparation is the same with or without time varying covariate.

Stata commands used to organize the data for estimating a continuous time model (**stset** when no time-varying covariates and **stsplit** to handle presence of time varying covariates for episode-splitting) not applicable to discrete time model.

How to create a panel (long form) data in STATA

Example #1 (page 280 of the book)

An example from the book.

C:\DurationWorkshop\STATA programs\Example1.do

Example #2: a single spell example

- A dataset I created. For each individual, we observe a single unemployment spell.
- A randomly assigned treatment, D_i , before sample starts.
- A time constant variable, X_i
- A time changing variable, W_{it} – local unemployment rates.
- Duration dependence $f\left(t\right)$

$$\lambda_i\left(t\right)=\frac{1}{1+\exp\left(-\theta-D_i\gamma-X_i\beta-W_{it}\delta-f\left(t\right)\right)}$$

- γ is the coefficient for the treatment dummy variable. This is an "intention-to-treat" (ITT) design and γ captures the effect of being assigned to a treatment not the actual treatment effect on employment status.

- W_{it} is the Canadian provincial prime age male monthly unemployment rate from January 2000 to December 2007.

- Each individual enters the spell at different calendar time (from January 2000 to June 2001) that is determined by a random draw.

- Right censored at 40 periods.

C:\DurationWorkshop\STATA programs\Example2.do

II. Estimate a single-spell duration model

2.1 A single-spell duration model without unobserved heterogeneity

This can be done easily in STATA using “logit” command.

Duration dependence $f(t)$: the step function we just created, can be incorporated as additional regressors.

C:\DurationWorkshop\STATA programs\ EstimationNoUH.do

Hazard coefficient interpretation

Effects of covariate on hazard rate

Directional (not magnitude) effect of covariates on expected duration.

Usually researchers and policy makers care about:

Expected duration

Magnitude effect of changing covariates (most interestingly policy change) on expected duration

Additional computation is needed, and we will cove them in the next section.

(Switch to John)

John's Section II starts here

Case 5: what about unobserved differences across individuals – this will lead to a considerable increase in difficulty conceptually and in estimation. Xianghong's R programs should be extremely helpful for you here

$$\lambda_i\left(t\middle|\theta_i\right) = \frac{1}{1+\exp\left(-\theta_i - Z_{i\tau}\eta - f\left(t\right)\right)}$$

What does θ_i capture

-optimism

-work ethic

-ability

Note: need $\text{cov}\left(\theta_i, Z_{i\tau}\right)=0$ at the start of the spell, i.e. $\tau=\tau_0$

Will talk later about what, if anything, on can do if $\text{cov}\left(\theta_i, Z_{i\tau}\right) \neq 0$ at the start of the spell, i.e. $\tau=\tau_0$.

We will assume that the heterogeneity is drawn from a discrete distribution with points of support $\theta_1, \theta_2, ..., \theta_{J-1}, \theta_J$ and associated probabilities $P_1, P_2, ..., P_{J-1}, P_J$, where $P_J = 1 - \sum_{j=1}^{J-1} P_j$ (Heckman and Singer Ecmta 1984).

Need to control for unobserved heterogeneity if we want to measure duration dependence $f(t)$ and the coefficients.

Why need to control for unobs heterogeneity duration dependence:

Draw graph on board

Why need to control for unobs heterogeneity for coefficients:

Even if $\text{cov}(\theta_i, Z_{i\tau}) = 0$ at the start of the spell, i.e. $\tau = \tau_0$ there will be dynamic selection it won't be true that $\text{cov}(\theta_i, Z_{i\tau})$ for $\tau > \tau_0$. If a high education person is still in unemployment after several periods, they probably have a bad draw on the unobserved heterogeneity, assuming education leads to a faster exit from unemployment holding θ_i constant. This correlation will lead to biased parameter estimates.

Prob a spell lasts T weeks (density function) conditional on the person being of type θ_j

$$h\left(T \mid \theta_j\right) = \lambda_i(T \mid \theta_j)\prod_{r=1}^{T-1}(1-\lambda_i(r \mid \theta_j))$$

Unconditional probability of the a spell lasting T weeks

$$h(T) = \sum_{j=1}^{J} P_j \lambda_i(T \mid \theta_j)\prod_{r=1}^{T-1}(1-\lambda_i(r \mid \theta_j)),$$

i.e. average or integrate out θ_j.

Prob a spell is still ongoing after K weeks (Survivor function)

conditional on the person being of type θ_j

$$S\left(T \mid \theta_j\right) = \prod_{r=1}^{T}(1-\lambda_i(r \mid \theta_j))$$

The unconditional Survivor function is given by $S(T)=\sum_{j=1}^{J}P_j\prod_{r=1}^{T-1}(1-\lambda_i(r\,|\,\theta_j))$,

i.e. average or integrate out θ_j.

Where how long will a spell last on average (expected duration)

The conditional expression is

$$E\left(dur\,|\,\theta_j\right)=\sum_{r=1}^{\infty}rh\left(r\,|\,\theta_j\right)$$

and the unconditional expression

$$E\left(dur\right)=\sum_{j=1}^{J}P_jE\left(dur\,|\,\theta_j\right).$$

Identification

Most of the results are for the mixed proportional hazard model

$$\lambda_i\left(t\middle|\theta_i\right)=\exp(\theta_i)\exp(Z_i\eta)\exp(f\left(t\right)=\exp\left(\theta_i+Z_i\eta+f\left(t\right)\right).$$

Basic result – the distribution θ_i and $f\left(t\right)$ are nonparametrically identified. However, note that this is conditional on the MPH assumption. Time changing X's will help in the sense of getting smaller standard errors, as will having two spells for the same individual, as we describe later. But since the model is identified formally without either of these factors, we are getting basic identification off functional form assumption of MPH. Fortunately Monte Carlo evidence suggests that discrete approach works relatively well – I haven't seen Monte Carlo evidence for continuous time model.

Some Estimation Issues

-consider first flow sample – look at those entering unemployment within a certain calendar window, i.e. Jan 2011-May 2012.

Empirical Hazard and Survivor Functions

Implicitly assume only duration matters – Kaplan-Meier estimates

$\hat{\mu}(t) =$number leaving at week t/number still in unemployment entering t.

Get standard errors for $\{V(\hat{\mu}(t)) = \hat{\mu}(t)[1 - \hat{\mu}(t)] /$ N(t)$\}$. N(t) is the number still in unemployment entering t.

Note cov($(\hat{\mu}(t), \hat{\mu}(t')) \neq 0$), but could use the bootstrap if you need this covariance.

e.g. of this Empirical Hazard Function from ham and rea 1987 e.g. of Empirical Survivor Functions from HLSS

Estimation with Unobserved heterogeneity

Problems

1. As Xianghong will show you this is a nontrivial nonlinear estimation problem. It's easy to overfit the problem – use objective criterion like Schwartz, AIC that penalizes parameterization to choose the model.
2. There is an incidental parameter problem since number of support points J goes up with the sample size.
3. Hypothesis testing – Suppose we have two points of support

And want to test $H_0 : \theta_1 = \theta_2$ then P=$\Pr(\theta = \theta_1)$ is not identified, so you end up with a non-standard testing problem see Davies (1985).

Policy effects

$$E(dur_i \mid \theta_j) = \sum_{t=1}^{\infty} t h_i(t \mid \theta_j)$$

$$E(dur_i) = \sum_{j=1}^{J} P_j E(dur_i \mid \theta_j).$$

$$ED^* = \frac{1}{N} \sum_{i=1}^{N} E(dur_i)$$

$$\frac{\Delta ED^*}{\Delta X_k} = \frac{1}{N} \sum_{i=1}^{N} \frac{\Delta E(dur_i)}{\Delta X_k}.$$

It is not appropriate to put the mean of theta in the hazard and look at it.

Very nice simplification for policy analysis – estimate the model with and without unobserved heterogeneity. Will get different parameters but estimates of $\frac{\Delta ED^{*}}{\Delta X_k}$ will be very similar.

Standard errors for policy effects.

In all models use the second derivative matrix to get standard errors – without unobserved heterogeneity, Stata will do this for you with the logit program. For standard errors of the policy effect, the expected durations are differentiable functions of the estimated parameters, with bounded and non-zero derivatives, so we can use the delta method

If $y = g(Z)$ and Var(Z)=Ω, then Var(y)=$\left[\frac{\partial g(Z)}{\partial Z}\right]' \Omega \left[\frac{\partial g(Z)}{\partial Z}\right]$.

Xianghong's second session starts here.

2.2 Estimating a single-spell duration model with unobserved heterogeneity

For the model with unobserved heterogeneity, we will need to use MLE estimation. Some do this with Fortran or C, but R and Matlab are becoming much more popular and also they are much easier to use. I will first give you a brief introduction to R and then we consider a simple example of maximization problem.

R – A statistical software for computing and graphics

- **What is R?**
  - **R** is an open source programming language and software environment for statistical computing and graphics.

- It is originally used in academics as the predominant language for statisticians, but also widely used by engineers and scientists.
- R is rapidly gaining ground the business world. Companies as diverse as Google, Pfizer, Merck, Bank of America, the InterContinental Hotels Group and Shell use it.

- **Main features – a flexible and versatile language**

  - R uses a command line interface similar to Stata. You may type your command at the command prompt and see side effects immediately. Thus with a bit of preparation it is as easy to use as Stata, but

    - Programs are much better vetted for mistakes
    - Always backward compatible unlike Stata
    - Much better than Stata for non-parametric estimation.

  - R has quickly found a following because statisticians, engineers and scientists who find it much easier to use than C, C++ and Fortran 90.

- For computationally intensive tasks, C, C++, and Fortran code can be linked and called at run time.
- Another strength of R is static graphics, which can produce publication-quality graphs, including mathematical symbols.

- R's matrix features and performance are comparable to Matlab and Gauss. (there is an online manual on how to translate between R and Matlab).

R – Why it could be very useful for economists

I assume you already have a language such as Stata to clean data and run some basic models.

R will be extremely helpful in the following aspects if:

- You want to implement a more complicated model requiring numerical optimization (such as duration models we are covering now).

- R has extremely reliable and relatively user-friendly optimization algorithms and a very active online forum where you can seek help if a problem occurs.
- Here you get fast responses from technically strong statisticians and mathematicians who routinely handle numerical maximization. Along this direction the resources are more abundant on the R forum than on the Stata forum.

- You want to adopt a cutting-edge nonparametric approach, such as matching, LATE, regression discontinuity. R's reliability in nonparametric implementation has probably the highest standard among all statistical softwares (I would avoid Stata for such tasks as simple as kernel density estimation).

R Official website: http://www.r-project.org/

One of R's optimization routine: optim

- R is well-suited for programming your own maximum likelihood routines.
- There are several procedures for optimizing likelihood functions. I will focus on the "optim" command, which implements the BFGS and simulated and annealing (which we will discuss below), among others.
- Optimization through optim is relatively straightforward, since it only requires user provided likelihood functions.

BFGS

The BFGS method approximates Newton's method, a class of hill-climbing optimization techniques that seeks a stationary point of a (preferably twice continuously differentiable) function. It relies on gradients and hessian to determine the next move in the process of searching for a optimum.

Define your log-likelihood function

Syntax:

```
Name = function(pars,object){
declarations
logl = loglikelihood function
return(-logl)
}
```

A simple example

Suppose we have a sample that was drawn from a normal distribution with unknown mean μ and variance σ^2. The objective is to estimate these parameters relying on numerical maximization. The normal log-likelihood function (of a sample with n observations) is given by:

$$l = -0.5n\ln\left(2\pi\right) - 0.5n\ln\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_i \left(y_i - \mu\right)^2$$

Let's consider this numerical maximization problem in R.

C:\DurationWorkshop\R programs\SimpleMLE.r

Now our single-spell duration model with unobserved heterogeneity

As John discussed, we often assume a nonparametric distribution with finite support point for unobserved heterogeneity:

θ follows a discrete distribution with points of support $\theta_1, \theta_2, ..., \theta_J$ and associated probabilities $p_1, p_2, ..., p_J$ respectively, where $p_J = 1 - \sum_{k=1}^{J-1} p_k$.

It is distributed independently across individuals and is fixed over time for a given individual.

Individual likelihood contribution with a complete spell:

$$l_i\left(K\right) = \log\left\{\sum_{j=1}^{J} p_j \cdot \left[\lambda_i(K \mid \theta_j)\right]\prod_{t=1}^{K-1}\left[1-\lambda_i(t \mid \theta_j)\right]\right\}$$

This functional form creates a very difficult numerical maximization problem (multiple local optima).

Our simple single spell data actually has unobserved heterogeneity and individuals belong to two (unobserved) types with the following parameter values:

$$\theta_1 = -0.8, \theta_2 = -2.8, p_1 = p_2 = 0.5, D \sim Bernoulli\left(p = 0.5\right), \gamma = 1,$$

$$\beta = 1, \delta = -0.05, X \sim N\left(0, \sigma^2 = 0.25\right), f\left(t\right) = \exp\left(\frac{1-t}{5}\right) - 1$$

Now let's consider estimating this model in R.

C:\DurationWorkshop\R programs\EstimationSingleSpell.r

(Switch to John)