An introduction to discrete duration analysis: part 2

Bryan S. Graham, UC - Berkeley & NBER

February 1, 2023

Set-up: All notation is as defined in part 1, (i) Y_i is the actual duration of unit i and (ii) C_i is unit i's random censoring time. For each of $i = 1, \ldots, N$ randomly sampled units we observe

$$Z_i = \min \{Y_i, C_i\},$$

a vector of covariates, X_i, and the non-censored indicator

$$D_i = \begin{cases} 1 & \text{if } Y_i \leq C_i \\ 0 & \text{if } Y_i > C_i \end{cases}.$$

When $D_i = 1$ we observe unit $i's$ actual spell length, whereas when $D_i = 0$ we only know that $Y_i > Z_i$. Estimation and inference is based upon the random sample $(Z_1, D_1, X_1'), (Z_2, D_2, X_2'), \ldots, (Z_N, D_N, X_N')$. Time is discrete with $Y_i \in \mathbb{Y} = \{y_1, y_2, \ldots, y_J\}$.

Our goal is to incorporate the unit-specific covariate vector X_i into our analysis. For example X_i might include measures of unemployment benefits which we would like to relate to the length of unemployment spells. Alternatively X_i might include indicators for participation in post-release prisoner re-entry support programs whose relationship with recidivism may be of interest.

Let $\lambda(y|X;\theta)$ be some parametric family of functions indexed by the unknown parameter $\theta \in \Theta$. We will assume that the conditional hazard function is a member of this family with parameter $\theta = \theta_0$:

$$\lambda(y|X;\theta_0) = \Pr(Y = y|Y \geq y, X). \tag{1}$$

Note that, due to the discrete treatment of time,

$$\Pr(Y = y|Y \geq y, X) = \Pr(Y = y|Y > y - 1, X). \tag{2}$$

The survival function is also indexed by θ

$$S\left(y|\,X;\theta_0\right)=\Pr\left(Y>y|\,X\right), \tag{3}$$

since, recalling our analysis in part 1,

$$\Pr\left(Y>y|\,X\right)=\prod_{t=1}^{y}\left[1-\lambda\left(t|\,X;\theta_0\right)\right]. \tag{4}$$

Equalities (2) and (4), along with some basic probability manipulations, yield

$$\begin{aligned}
\Pr\left(Y=y|\,X\right) &= \Pr\left(Y=y,Y>y-1|\,X\right)\\
&= \Pr\left(Y=y|\,Y>y-1,X\right)\Pr\left(Y>y-1|\,X\right)\\
&= \lambda\left(y|\,X;\theta_0\right)S\left(y-1|\,X;\theta_0\right)\\
&= \lambda\left(y|\,X;\theta_0\right)\prod_{t=1}^{y-1}\left[1-\lambda\left(t|\,X;\theta_0\right)\right]\\
&= \frac{\lambda\left(y|\,X;\theta_0\right)}{1-\lambda\left(y|\,X;\theta_0\right)}\prod_{t=1}^{y}\left[1-\lambda\left(t|\,X;\theta_0\right)\right].
\end{aligned}$$

This expression will be used to construct the log-likelihood function below.

Maximum Likelihood Estimation (MLE)

Likelihood

Recall that (X,Z,D) corresponds to a generic random draw from the sampled population. The form of the conditional probability of the event $Z=y,D=d$ given X depends on whether $d=1$ (uncensored case) or $d=0$ (censored case). In the uncensored case, we have

$$\begin{aligned}
\Pr\left(Z=y,D=1|\,X\right) &= \Pr\left(Y=y,D=1|\,X\right)\\
&= \Pr\left(Y=y,C\geq y|\,X\right)\\
&= \Pr\left(Y=y|\,X\right)\left[1-G\left(y|\,X\right)\right]
\end{aligned}$$

where $G\left(y|\,X\right)=\Pr\left(C\leq y|\,X\right)$ is the conditional cumulative distribution function (cdf) for the censoring times.

When Y is censored we instead have

$$\begin{aligned}\Pr\left(\left.Z=y,D=0\right|X\right) &= \Pr\left(\left.Y>y,D=0\right|X\right)\\ &= \Pr\left(\left.Y>y,C=y\right|X\right)\\ &= \Pr\left(\left.Y>y\right|X\right)g\left(\left.y\right|X\right),\end{aligned}$$

where $g\left(\left.y\right|X\right)=\Pr\left(\left.C=y\right|X\right)$ is the conditional probability mass function (pmf) for the censoring times. Note that role played by the assumption that censoring is uninformative (i.e., $C\perp\left.Y\right|X$) in the derivations above.

Putting things together, and using (1) and (3) above, we get a likelihood of the event $Z=y,D=d$ given X equal to

$$\begin{aligned}\Pr\left(\left.Z=y,D=d\right|X\right) &= \Pr\left(\left.Y=y\right|X\right)^{d}\Pr\left(\left.Y>y\right|X\right)^{1-d}\left[1-G\left(\left.y\right|X\right)\right]^{d}g\left(\left.y\right|X\right)^{1-d}\\ &= \left\{\frac{\lambda\left(\left.y\right|X;\theta_0\right)}{1-\lambda\left(\left.y\right|X;\theta_0\right)}\prod_{t=1}^{y}\left[1-\lambda\left(\left.t\right|X;\theta_0\right)\right]\right\}^{d}\left\{\prod_{t=1}^{y}\left[1-\lambda\left(\left.t\right|X;\theta_0\right)\right]\right\}^{1-d}\\ &\quad\times\left[1-G\left(\left.y\right|X\right)\right]^{d}g\left(\left.y\right|X\right)^{1-d}.\end{aligned}$$

If we take a random draw of (Z,D) from a subpopulation of units with covariates X, the probability that $Z=y$ and $D=d$ is given by the expression above (Just like the probability of observing "heads" when we flip a fair coin is 1/2).

The likelihood of the entire sample $\{Z_i,D_i\}_{i=1}^{N}$ conditional on the covariates $\mathbf{X}=(X_1,\ldots,X_N)'$ is, when $\theta_0=\theta$,

$$\begin{aligned}L_N\left(\theta\right) &= \prod_{i=1}^{N}\left\{\frac{\lambda\left(\left.Z_i\right|X_i;\theta\right)}{1-\lambda\left(\left.Z_i\right|X_i;\theta\right)}\prod_{y=1}^{Z_i}\left[1-\lambda\left(\left.y\right|X_i;\theta\right)\right]\right\}^{D_i}\left\{\prod_{y=1}^{Z_i}\left[1-\lambda\left(\left.y\right|X_i;\theta\right)\right]\right\}^{1-D_i}\\ &\quad\times\prod_{i=1}^{N}\left[1-G\left(\left.Z_i\right|X_i\right)\right]^{D_i}g\left(\left.Z_i\right|X_i\right)^{1-D_i}\\ &= \prod_{i=1}^{N}\left\{\frac{\lambda\left(\left.Z_i\right|X_i;\theta\right)}{1-\lambda\left(\left.Z_i\right|X_i;\theta\right)}\right\}^{D_i}\left\{\prod_{y=1}^{Z_i}\left[1-\lambda\left(\left.y\right|X_i;\theta\right)\right]\right\}\\ &\quad\times\prod_{i=1}^{N}\left[1-G\left(\left.Z_i\right|X_i\right)\right]^{D_i}g\left(\left.Z_i\right|X_i\right)^{1-D_i}. \qquad (5)\end{aligned}$$

Let $\left\{(X_i',Z_i,D_i)'\right\}_{i=1}^{N}$ be the simple random sample in hand. If we evaluate (5) using the value of X_i, Z_i and D_i in the collected sample we get the *ex ante* probability that a random sample of size N would take the form of the one in hand.

Our approach to estimation will be to choose $\hat{\theta}$ to maximize (5) with respect to θ. This estimate is the maximum likelihood estimate (MLE). Imagine we believe that $\theta_0 = \theta^*$. If we evaluate (5) at $\theta = \theta^*$ and get a very small number this suggest that the chance of observing a sample like the one in hand is very low in populations with $\theta_0 = \theta^*$. This suggests that perhaps our beliefs are incorrect. If we flip a coin 100 times and it comes up heads each and every time, perhaps the coin is not fair?

The random event observed is the configuration of durations and censoring indicators in the sample in hand. The *ex ante* probability of this event is $L_N(\theta_0)$. By setting $\hat{\theta} = \arg\max_{\theta \in \Theta} L_N(\theta)$ we select a value for $\hat{\theta}$ that maximizes the *ex ante* chance ("likelihood") of observing a sample just like the one in hand. This, it turns out, is often a good way to construct parameter estimates. Values of θ far from $\hat{\theta}$ have the implication that samples like the one in hand would be very unlikely in populations with such a value for θ.

Note that the second part of (5) doesn't depend on θ; so this part of (5) can be dropped since doing so will not change the arg max. Taking the logarithm of (5) and dropping the parts which don't vary with θ we get a sample log-likelihood proportional to

$$\begin{aligned} l_N(\theta) = \ln L_N(\theta) &\propto \sum_{i=1}^{N} D_i \ln\left\{\frac{\lambda\left(Z_i|\, X_i;\theta_0\right)}{1-\lambda\left(Z_i|\, X_i;\theta_0\right)}\right\} + \sum_{i=1}^{N}\sum_{y=1}^{Z_i} \ln\left\{1-\lambda\left(y|\, X;\theta_0\right)\right\} \\ &\propto \sum_{i=1}^{N}\sum_{y=1}^{Z_i} W_{iy} \ln\left\{\frac{\lambda\left(y|\, X_i;\theta_0\right)}{1-\lambda\left(y|\, X_i;\theta_0\right)}\right\} + \sum_{i=1}^{N}\sum_{y=1}^{Z_i} \ln\left\{1-\lambda\left(y|\, X;\theta_0\right)\right\}, \end{aligned} \quad (6)$$

where in the second line we set $W_{iy} = 1$ if $y = Z_i$ *and* $D_i = 1$ and $W_{iy} = 0$ otherwise.

Data organization

Say our first observation is $Z_1 = 4, D_1 = 1$ (a unit with a completed spell of four months); our second observation is $Z_2 = 3$, $D_2 = 0$ (a unit censored after three months) and so on. For estimation purposes it is helpful to use our raw data to construct a "person-period" dataset of the form:

i	y	Z_i	D_i	W_{iy}	X_i	V_{i1}	V_{i2}	$\cdots$	V_{iK}
1	1	4	1	0	X_1	1	0	$\cdots$	0
1	2	$\vdots$	$\vdots$	0	X_1	0	1	$\cdots$	0
1	3	$\vdots$	$\vdots$	0	X_1	$\vdots$	$\vdots$	$\ddots$	$\vdots$
1	4	4	1	1	X_1	0	0	$\cdots$	0
2	1	3	0	0	X_2	1	0	$\cdots$	0
2	2	$\vdots$	$\vdots$	0	X_2	0	1	$\cdots$	0
2	3	3	0	0	X_2	$\vdots$	$\vdots$	$\ddots$	$\vdots$

Here V_{iy} is an indicator for whether the given row of the dataset corresponds to period y or not (as in the last lecture K denotes the length of the maximum uncensored duration).

Conditional hazard model

For our conditional hazard function we can use any number of familiar binary choice models. A common choice is the *logit* specification with

$$\lambda\left(y|\, X;\theta_0\right)=\frac{\exp\left(\sum_{y=1}^{K}V_{iy}\gamma_y+X_i'\beta\right)}{1+\exp\left(\sum_{y=1}^{K}V_{iy}\gamma_y+X_i'\beta\right)}.$$

The $\gamma=(\gamma_1,\ldots,\gamma_K)'$ coefficients allow the general shape of the (baseline) hazard function to vary with y, while β parameterizes how the covariates alter the hazard of exit. This gives $\theta=(\beta',\gamma')'$.

Another specification that is sometimes used in discrete duration analysis is the *complementary log-log* one where

$$\lambda\left(y|\, X;\theta_0\right)=1-\exp\left[-\exp\left(\sum_{y=1}^{K}V_{iy}\gamma_y+X_i'\beta\right)\right].$$

Estimation

With the appropriate parametric choice for the hazard function and the data organized into "person-period" form estimation is straightforward. Simply compute the logit (or complementary log-log) fit of W_{iy} onto $V_{i1},V_{i2},\ldots,V_{iK}$ and X_i (with no constant included). The coefficients and standard errors reported by a standard "logit" program will be consistent under our assumptions.

How might you restrict the form of the baseline hazard?

Further reading

Singer & Willett (1993), Jenkins (1995) and Efron (1988) are discuss parametric discrete duration analysis. Ashenfelter & Card (2002) is an interesting and accessible application of discrete hazard methods.

References

Ashenfelter, O. & Card, D. (2002). Did the elimination of mandatory retirement affect faculty retirement? *American Economic Review*, 92(4), 957 – 980.

Efron, B. (1988). Logistic regression, survival analysis, and the kaplan-meier curve. *Journal of the American Statistical Association*, 83(402), 414 – 425.

Jenkins, S. P. (1995). Easy estimaton methods for discrete-time duration models. *Oxford Bulletin of Economics and Statistics*, 57(1), 129 – 137.

Singer, J. D. & Willett, J. B. (1993). It's about time: using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics*, 18(2), 155 – 195.

