Incorporating Time-Varying Covariates in a Simple Mixture Model for Discrete-Time Duration Data

Peter S. Fader
www.petefader.com

Bruce G. S. Hardie
www.brucehardie.com[†]

July 2020

1 Introduction

The beta-geometric (BG) distribution is a robust model for characterising discrete duration-time data. Important marketing applications include modeling the duration of a customer's contractual relationship with a firm (Fader and Hardie 2007) and the calculation of customer lifetime value in contractual settings (Fader and Hardie 2010, 2017).

A number of analysts have expressed a desire to incorporate the effects of time-varying covariates in the BG model. Unfortunately, it is not practical to do so. This note shows how such effects can be accommodated using a lesser-known model for discrete duration-time data which, for all intents and purposes, is identical to the BG.

2 Three Continuous Mixtures of Geometrics

The beta-geometric distribution is derived in the following manner:

- Let the random variable T denote the (discrete) time period in which the event of interest occurs.
- At the level of the individual, we assume that T is distributed geometric with pmf

$$P(T = t \,|\, \theta) = \theta(1-\theta)^{t-1}\,,\ 0 < \theta < 1,\ t = 1, 2, 3, \ldots \tag{1}$$

and survivor function

$$S(t \,|\, \theta) = (1-\theta)^t\,,\ 0 < \theta < 1,\ t = 0, 1, 2, \dots \tag{2}$$

- From the analyst's perspective, the unobserved (and unobservable) θ is a realization of the random variable Θ. We assume that Θ is characterized by the beta distribution:

$$g(\theta \,|\, \gamma, \delta) = \frac{\theta^{\gamma-1}(1-\theta)^{\delta-1}}{B(\gamma, \delta)}\,,\ \gamma, \delta > 0. \tag{3}$$

It follows that for a randomly chosen individual,

$$\begin{aligned} P(T = t \,|\, \gamma, \delta) &= \int_0^1 P(T = t \,|\, \theta) g(\theta \,|\, \gamma, \delta)\, d\theta \\ &= \frac{B(\gamma+1, \delta+t-1)}{B(\gamma, \delta)}\,,\ t = 1, 2, 3, \dots \end{aligned} \tag{4}$$

and

$$\begin{aligned} S(t \,|\, \gamma, \delta) &= \int_0^1 S(t \,|\, \theta) g(\theta \,|\, \gamma, \delta)\, d\theta \\ &= \frac{B(\gamma, \delta+t)}{B(\gamma, \delta)}\,,\ t = 0, 1, 2, \dots \end{aligned} \tag{5}$$

The choice of the beta distribution to characterize Θ is driven by mathematical convenience; it is a flexible distribution that results in closed-form marginal distributions. A less-obvious alternative is to use the transformation $\theta = 1 - \exp(-\lambda)$ where differences in λ across individuals are characterized by a gamma distribution with parameters r and α. This is equivalent to assuming that heterogeneity in θ is captured by a Grassia(II) distribution (Grassia 1977).

The resulting mixture model has the following pmf and survivor function:

$$\begin{aligned} P(T = t \,|\, r, \alpha) &= \int_0^\infty (1 - e^{-\lambda})\big(e^{-\lambda}\big)^{t-1} \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)}\, d\lambda \\ &= \left(\frac{\alpha}{\alpha+t-1}\right)^r - \left(\frac{\alpha}{\alpha+t}\right)^r, \end{aligned} \tag{6}$$

$$\begin{aligned} S(t \,|\, r, \alpha) &= \int_0^\infty \big(e^{-\lambda}\big)^t \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)}\, d\lambda \\ &= \left(\frac{\alpha}{\alpha+t}\right)^r. \end{aligned} \tag{7}$$

Fader et al. (2019) call this the Grassia(II)-geometric (G2G) distribution and show that the fit and predictive performance of this model is equivalent to that of the BG. (Some readers may recognize that this distribution could also be called a discretised Pareto distribution of the second kind.)

Another alternative distribution for characterising Θ is the logit-normal. This is equivalent to using the transformation

$$\theta = \frac{e^\eta}{e^\eta + 1}$$

and assuming that differences in η across individuals are characterized by a normal distribution with mean μ and variance σ^2. For a randomly chosen individual,

$$P(T = t \mid \mu, \sigma^2) = \int_0^1 \left(\frac{e^\eta}{e^\eta + 1}\right)\left(1 - \frac{e^\eta}{e^\eta + 1}\right)^{t-1} g(\eta \mid \mu, \sigma^2)\, d\eta\,, \quad (8)$$

and

$$S(t \mid \mu, \sigma^2) = \int_0^1 \left(1 - \frac{e^\eta}{e^\eta + 1}\right)^{t} g(\eta \mid \mu, \sigma^2)\, d\eta \quad (9)$$

We call this the logit-normal-geometric (LNG) distribution. Note that there are no closed-form solutions to these integrals; they must be evaluated numerically.

3 The Modeling Problem and a Simple Solution

The geometric distribution is the distribution of the number of iid Bernoulli trials needed to get one "success". If the probability of success is allowed to vary across trials, we have

$$P(T = t) = \theta_t \prod_{j=1}^{t-1} (1 - \theta_j)$$

and

$$S(t) = \begin{cases} 1 & t = 0 \\ \prod_{j=1}^{t} (1 - \theta_j) & t = 1, 2, 3, \ldots \end{cases}$$

Let $\mathbf{z}(j)$ denote the vector of covariates at time j and $\boldsymbol{\beta}$ the effects of these covariates. (It is **very** important to note that this vector of covariates does not include an intercept.) Let $\mathbf{Z}(t) = \{\mathbf{z}(1), \mathbf{z}(2), \ldots, \mathbf{z}(t)\}$ represent the covariate path up to time t.

As we reflect on how to incorporate these covariate effects in our model of duration times, it is natural to think that the probability of the event happening at time j, given that it has not occurred so far, is a function of the covariates at time j; i.e., $\theta_j = f(\mathbf{z}(j))$. We wish to incorporate the effects of these covariates at the individual-level and then account for unobserved cross-sectional heterogeneity. It is not practical to do so using the beta distribution to characterize unobserved heterogeneity. (That is, we do not end up with a simple closed-form solution.)

The most obvious way of making θ_j a function of $\mathbf{z}(j)$ is to use the logit link function:

$$\theta_j = \frac{e^{\eta+\mathbf{z}(j)\boldsymbol{\beta}'}}{e^{\eta+\mathbf{z}(j)\boldsymbol{\beta}'}+1}\,.$$

This means that, at the level of the individual, we have

$$P(T = t \,|\, \eta, \boldsymbol{\beta}; \mathbf{Z}(t)) = \left(\frac{e^{\eta+\mathbf{z}(t)\boldsymbol{\beta}'}}{e^{\eta+\mathbf{z}(t)\boldsymbol{\beta}'}+1}\right)\prod_{j=1}^{t-1}\left(1 - \frac{e^{\eta+\mathbf{z}(j)\boldsymbol{\beta}'}}{e^{\eta+\mathbf{z}(j)\boldsymbol{\beta}'}+1}\right)$$

and

$$S(t \,|\, \eta, \boldsymbol{\beta}; \mathbf{Z}(t)) = \prod_{j=1}^{t}\left(1 - \frac{e^{\eta+\mathbf{z}(j)\boldsymbol{\beta}'}}{e^{\eta+\mathbf{z}(j)\boldsymbol{\beta}'}+1}\right).$$

Assuming differences in η across individuals are characterized by a normal distribution with mean μ and variance σ^2, it follows that for a randomly chosen individual,

$$P(T = t \,|\, \mu, \sigma^2, \boldsymbol{\beta}; \mathbf{Z}(t)) = \int_0^1 P(T = t \,|\, \eta, \boldsymbol{\beta}; \mathbf{Z}(t))\, g(\eta \,|\, \mu, \sigma^2)\, d\eta \qquad (10)$$

and

$$S(t \,|\, \mu, \sigma^2, \boldsymbol{\beta}; \mathbf{Z}(t)) = \int_0^1 S(t \,|\, \eta, \boldsymbol{\beta}; \mathbf{Z}(t))\, g(\eta \,|\, \mu, \sigma^2)\, d\eta\,. \qquad (11)$$

When $\boldsymbol{\beta} = \mathbf{0}$, (10) and (11) reduce to (8) and (9), respectively. As with the LNG, there are no closed-form solutions to these integrals. Historically, this was a barrier to the adoption of such models. However, this is less of an issue these days as the integrals can easily be evaluated using simulation methods. Nevertheless, a closed-form solution is still desirable. (For example, a "simple" closed-form solution is less computationally burdensome.)

A less familiar alternative to the logit link function is the complementary log-log link function

$$\theta_j = 1 - e^{-e^{\eta+\mathbf{z}(j)\boldsymbol{\beta}'}}\,,$$

which we can rewrite as

$$\theta_j = 1 - e^{-\lambda e^{\mathbf{z}(j)\boldsymbol{\beta}'}}\,.$$

This means that, at the level of the individual, we have

$$\begin{aligned}
P(T = t \,|\, \lambda, \boldsymbol{\beta}; \mathbf{Z}(t)) &= \left(1 - e^{-\lambda e^{\mathbf{z}(t)\boldsymbol{\beta}'}}\right) \prod_{j=1}^{t-1} \left(e^{-\lambda e^{\mathbf{z}(j)\boldsymbol{\beta}'}}\right) \\
&= e^{-\lambda C(t-1)} - e^{-\lambda C(t)}
\end{aligned}$$

and

$$\begin{aligned}
S(t \,|\, \lambda, \boldsymbol{\beta}; \mathbf{Z}(t)) &= \prod_{j=1}^{t} \left(e^{-\lambda e^{\mathbf{z}(j)\boldsymbol{\beta}'}}\right) \\
&= e^{-\lambda C(t)} ,
\end{aligned}$$

where

$$C(t) = \sum_{j=1}^{t} e^{\mathbf{z}(j)\boldsymbol{\beta}'} .$$

Assuming differences in λ across individuals are characterized by a gamma distribution with parameters r and α, it follows that for a randomly chosen individual,

$$\begin{aligned}
P(T = t \,|\, r, \alpha, \boldsymbol{\beta}; \mathbf{Z}(t)) &= \int_0^\infty P(T = t \,|\, \lambda, \boldsymbol{\beta}; \mathbf{Z}(t)) g(\lambda \,|\, r, \alpha) \\
&= \left(\frac{\alpha}{\alpha + C(t-1)}\right)^r - \left(\frac{\alpha}{\alpha + C(t)}\right)^r \qquad (12)
\end{aligned}$$

and

$$\begin{aligned}
S(t \,|\, r, \alpha, \boldsymbol{\beta}; \mathbf{Z}(t)) &= \int_0^\infty S(t \,|\, \lambda, \boldsymbol{\beta}; \mathbf{Z}(t)) g(\lambda \,|\, r, \alpha) \\
&= \left(\frac{\alpha}{\alpha + C(t)}\right)^r . \qquad (13)
\end{aligned}$$

We call this the G2G+covariates model. When $\boldsymbol{\beta} = \mathbf{0}$, $C(t) = t$, and (12) and (13) reduce to (6) and (7), respectively.

As with the G2G, some readers will see that the G2G+covariates model could be viewed as a discretised version of the continuous-time model in which time-varying covariates are "added" to an exponential distribution using the proportional hazards framework and unobserved heterogeneity is characterized by the gamma distribution (e.g., Fader et al. 2003).

This G2G+covariates model has very simple closed-form expressions for its pmf and survivor function, and is our recommended go-to model for anyone modeling (single-event) discrete-time duration data with time-varying covariates. More generally, it is a natural replacement for the BG whenever we want the (timing-related) phenomenon being characterized by the BG to be a function of time-varying covariates. For example, Braun et al. (2015) extend the BG/NBD model using this model in place of the BG to examine the impact of time-varying covariates on (latent) attrition.

References

Braun, Michael, David A. Schweidel, and Eli Stein (2015), "Transaction Attributes and Customer Valuation," *Journal of Marketing Research*, **52** (December), 848–864.

Fader, Peter S. and Bruce G. S. Hardie (2007), "How To Project Customer Retention," *Journal of Interactive Marketing*, **21** (Winter), 76–90.

Fader, Peter S. and Bruce G. S. Hardie (2010), "Customer-Base Valuation in a Contractual Setting: The Perils of Ignoring Heterogeneity," *Marketing Science*, **29** (January–February), 85–93.

Fader, Peter S. and Bruce G.S. Hardie (2017), "Exploring the Distribution of Customer Lifetime Value (in Contractual Settings)." [`http://brucehardie.com/notes/035/`]

Fader, Peter S., Bruce G. S. Hardie, Daniel McCarthy, and Ramnath Vaidyanathan (2019), "Exploring the Equivalence of Two Common Mixture Models for Duration Data," *The American Statistician*, **73** (August), 288–295.

Fader, Peter S., Bruce G. S. Hardie, and Robert Zeithammer (2003), "Forecasting New Product Trial in a Controlled Test Market Environment," *Journal of Forecasting*, **22** (August), 391–410.

Grassia, A. (1977), "On a Family of Distributions with Argument Between 0 and 1 Obtained by Transformation of the Gamma and Derived Compound Distributions," *Australian Journal of Statistics*, **19** (2), 108–114.