Computing P(alive) Using the BG/NBD Model

Peter S. Fader
www.petefader.com

Bruce G. S. Hardie
www.brucehardie.com

Ka Lok Lee†
kaloklee@wharton.upenn.edu

December 2008

One of the key results we associate with the Pareto/NBD model (Schmittlein et al. 1987) is an expression for $P(\text{alive}\,|\,x, t_x, T)$, the probability that a customer with purchase history (x, t_x, T) is "alive" at time T.

In Fader, Hardie and Lee (2005) we present the BG/NBD model as an alternative to the Pareto/NBD. The positioning of this work is that the model yields very similar results to the Pareto/NBD while being vastly easier to implement. Although we present analogous expressions for most of the quantities reported in Schmittlein et al. (1987), we do not report an expression for $P(\text{alive})$. This has led some to claim that this quantity cannot be computed using the BG/NBD model; as we shall see, this is not the case. (Our primary reason for not presenting this result is simply that we see the quantity as a means to an end, not an end in and of itself. However, continued interest in it leads us to write this note.)

Deriving the expression for $P(\text{alive})$ under the BG/NBD model is a trivial exercise. Given the assumptions of the BG/NBD model, the probability that a customer with purchase history (x, t_x, T) is alive at time T is simply the probability that he did not "die" at t_x.

Recall the BG/NBD likelihood function:

$$L(r, \alpha, a, b\,|\,x, t_x, T) = \frac{B(a, b+x)}{B(a,b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+T)^{r+x}} + \delta_{x>0}\, \frac{B(a+1, b+x-1)}{B(a,b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+t_x)^{r+x}}\,. \quad (1)$$

Noting that the first half of (1) represents the likelihood under the assumption that the customer did not die at t_x (and is therefore alive at T), while the second half represents the likelihood under the assumption that the customer died at t_x, the application of Bayes' theorem gives us

$$
\begin{aligned}
&P(\text{alive} \mid x, t_x, T, r, \alpha, a, b) \\
&\quad = \frac{B(a, b+x)}{B(a,b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+T)^{r+x}} \Big/ L(r, \alpha, a, b \mid x, t_x, T) \\
&\quad = 1 \Big/ \left\{ 1 + \delta_{x>0} \frac{a}{b+x-1} \left(\frac{\alpha+T}{\alpha+t_x} \right)^{r+x} \right\}.
\end{aligned}
$$

We note that $P(\text{alive}) = 1$ for a customer who made no purchases in the interval $(0, T]$; this follows from the model's assumptions that death occurs after a purchase and that customers are alive at the beginning of the observation period. Those interested in the $P(\text{alive})$ metric may view this as a shortcoming of the model.

There are two ways to overcome this potential problem.

1. One variant of the basic BG/NBD model relaxes the assumption that all customers are alive at the beginning of the observation period. In particular, we assume that $\pi \times 100\%$ of the customer base is "dead" at the beginning of the observation period, with the behaviour of the remaining $(1-\pi) \times 100\%$ of the customer base is governed by the BG/NBD model. (There is an obvious parallel with the zero-inflated NBD model — see, for example, Morrison (1969).)

   The likelihood function for this variant of the basic model is

$$
\begin{aligned}
L(r, \alpha, a, b, \pi \mid x, t_x, T) = \pi \delta_{x=0} + (1-\pi) \bigg\{ & \frac{B(a, b+x)}{B(a,b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+T)^{r+x}} \\
& + \delta_{x>0} \frac{B(a+1, b+x-1)}{B(a,b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+t_x)^{r+x}} \bigg\}.
\end{aligned}
$$

   It follows that

$$
\begin{aligned}
&P(\text{alive} \mid x, t_x, T, r, \alpha, a, b, \pi) \\
&\quad = \begin{cases}
(1-\pi) \Big/ \left\{ \pi \left(\dfrac{\alpha+T}{\alpha} \right)^r + (1-\pi) \right\} & x = 0 \\[2ex]
1 \Big/ \left\{ 1 + \dfrac{a}{b+x-1} \left(\dfrac{\alpha+T}{\alpha+t_x} \right)^{r+x} \right\} & x > 0
\end{cases}
\end{aligned}
$$

   (We have not found support for $\pi > 0$ in our limited empirical testing of this model variant.)

2. Another variant of the basic BG/NBD model sees a flip of the "death coin" occurring at time 0. The likelihood function for this variant of the basic model is

$$
\begin{aligned}
L(r,\alpha,a,b \,|\, x,t_x,T) = {} & \frac{B(a,b+x+1)}{B(a,b)}\frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+T)^{r+x}} \\
& + \frac{B(a+1,b+x)}{B(a,b)}\frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+t_x)^{r+x}}\,.
\end{aligned}
$$

As this variant did not yield a better fit to the CDNOW dataset, we chose not to explore it in our paper. However, it has has been examined by Batislam et al. (2007, 2008), Hoppe and Wagner (2007), and Wagner and Hoppe (2008).

It follows that

$$
P(\text{alive} \,|\, x,t_x,T,r,\alpha,a,b) = 1 \Bigg/ \left\{1+\frac{a}{b+x}\left(\frac{\alpha+T}{\alpha+t_x}\right)^{r+x}\right\},
$$

which is less than 1 for a customer who made no purchases in the interval $(0,T]$.

References

Batislam, Emine Persentili, Meltem Denizel, and Alpay Filiztekin (2007), "Empirical Validation and Comparison of Models for Customer Base Analysis," *International Journal of Research in Marketing*, **24** (September), 201–209.

Batislam, E. P. M. Denizela and A. Filiztekin (2008), "Formal response to "Erratum on the MBG/NBD Model"," *International Journal of Research in Marketing* **25** (September), 227.

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005), ""Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model," *Marketing Science*, **24** (Spring), 275–284.

Hoppe, Daniel and Udo Wagner (2007), "Customer Base Analysis: The Case for a Central Variant of the Betageometric/NBD Model," *Marketing — Journal of Research and Management*, **3** (2), 75–90.

Morrison, Donald G. (1969), "Conditional Trend Analysis: A Model that Allows for Nonusers," *Journal of Marketing Research*, **6**, 342–346.

Schmittlein, David C., Donald G. Morrison, and Richard Colombo (1987), "Counting Your Customers: Who Are They and What Will They Do Next?" *Management Science*, **33** (January), 1–24.

Wagner, Udo and Daniel Hoppe (2008), "Erratum on the MBG/NBD Model," *International Journal of Research in Marketing* **25** (September), 225–226.