Issues in the Derivations of Key Expectations for the BG/NBD and Other Latent-Attrition Models

Norbert Schumacher
IRI

Peter S. Fader
www.petefader.com

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

June 2022

1 Introduction

Two key quantities for any model of buyer behavior in noncontractual settings are $E[X(t)]$, the expected number of transactions made by a randomly chosen individual in the interval $(0, t]$, and $E[Y(t) \,|\, x, t_x, T]$, the expected number of transactions in the interval $(T, T + t]$ for an individual who made x transactions in the interval $(0, T]$ with their last transaction occurring at t_x.

Fader et al. (2005) present expressions for these two quantities for the BG/NBD model. While the equations are correct, the associated derivations are incorrect for a portion of the parameter space. The purpose of this note is to present an alternative derivation of these two quantities that holds for the entire parameter space. In Section 2 we identify where the error in the derivation occurs. Correct derivations are presented in Sections 3. Similar derivation errors occur for these quantities under the Pareto/NBD and BG/BB models. Correct derivations are presented in Sections 4 and 5, respectively.

2 Where the Error Occurred

Let us revisit the derivation of $E[X(t) \,|, r, \alpha, a, b]$ for the BG/NBD model.[1] Conditional on λ and p, the expected number of transactions in $(0, t]$ is

$$E[X(t) \,|\, \lambda, p] = \frac{1}{p} - \frac{1}{p} e^{-\lambda p t}. \tag{1}$$

By definition, λ and p are unobserved. The BG/NBD model assumes these individual characteristics are distributed across the population according to a gamma and beta distribution, respectively. We therefore take the expectation of (1) over the distributions of Λ and P.

[1] See Fader et al. (2019) for a detailed derivation of all the various results presented in Fader et al. (2005).

1) We first take the expectation with respect to Λ:

$$E[X(t)\,|\,r,\alpha,p] = \int_0^\infty \left\{\frac{1}{p} - \frac{1}{p}e^{-\lambda pt}\right\} \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)}\, d\lambda$$
$$= \frac{1}{p} - \frac{\alpha^r}{p(\alpha+pt)^r}\,. \tag{2}$$

2) The next step is to take the expectation of this over the distribution of P. To do so, we break apart (2).

   a) We first evaluate

$$\int_0^1 \frac{1}{p} \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}\, dp = \frac{1}{B(a,b)} \int_0^1 p^{a-2}(1-p)^{b-1} dp$$

   which, recalling that the definite integral is the beta function,

$$= \frac{B(a-1,b)}{B(a,b)}$$
$$= \frac{\Gamma(a-1)\Gamma(b)}{\Gamma(a+b-1)} \Big/ \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$
$$= \frac{a+b-1}{a-1}\,.$$

   Note that the final step, which is repeated throughout this note, is required for the following reason. While $\Gamma(z)$ is defined for $-1 < z < 0$, it is negative. As such, it cannot be evaluated directly in computing environments that only offer a "log gamma" function (e.g., `gammaln()` in Excel). Therefore, we cannot compute $\Gamma(a-1)$ in such environments when $a < 1$. The final step removes this computational barrier.

   b) Next, we evaluate

$$\int_0^1 \frac{\alpha^r}{p(\alpha+pt)^r} \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)} dp$$
$$= \frac{\alpha^r}{B(a,b)} \int_0^1 p^{a-2}(1-p)^{b-1}(\alpha+pt)^{-r} dp$$

   letting $q = 1-p$ (which implies $dp = -dq$)

$$= \frac{\alpha^r}{B(a,b)} \int_0^1 q^{b-1}(1-q)^{a-2}(\alpha+t-qt)^{-r} dq$$
$$= \left(\frac{\alpha}{\alpha+t}\right)^r \frac{1}{B(a,b)} \int_0^1 q^{b-1}(1-q)^{a-2}\big(1-\tfrac{t}{\alpha+t}q\big)^{-r} dq$$

   which, recalling Euler's integral for the Gaussian hypergeometric function,

$$= \left(\frac{\alpha}{\alpha+t}\right)^r \frac{B(a-1,b)}{B(a,b)}\, {}_2F_1\big(r,b;a+b-1;\tfrac{t}{\alpha+t}\big)\,.$$

Combining the two interim results gives us

$$E[X(t)\,|\,r,\alpha,a,b] = \frac{a+b-1}{a-1}\left[1-\left(\frac{\alpha}{\alpha+t}\right)^{r}{}_2F_1\big(r,b;a+b-1;\tfrac{t}{\alpha+t}\big)\right]. \tag{3}$$

The first mistake occurs in Step 2a where we "recognize" that the definite integral is the beta function. This result is not true for $a \leq 1$. Furthermore, the expression can't be true because the integrand is strictly positive yet we get a negative number for $a < 1$.

A similar problem occurs in Step 2b when we recognize Euler's integral for the Gaussian hypergeometric function. For ${}_2F_1(a,b;c;z)$, the integral is defined for $c > b$. In this case we have $a+b-1 > b$, which means the integral is not defined for $a \leq 1$. Therefore the expectation is gained fallaciously.[2]

3 A More Careful Derivation

Notice that the error stemmed from the fact that both $1/p$ and $e^{-\lambda pt}/p$ are unbounded on the set $p \in (0,1)$ leading to undefined expectations with respect to some beta distributions $(a \leq 1)$. By contrast, the function $g(p) = \frac{1-e^{-\lambda pt}}{p}$ is bounded in the interval and therefore the expectation of $g(P)$ always exists. We derive this expectation in the following mannner.

We find the Taylor series expansion of (1) around the point $t = 0$. Noting that the kth derivative of (1) wrt t evaluated at $t = 0$ is $(-p)^{k-1}\lambda^k$, we have

$$E[X(t)\,|\,\lambda,p] = \sum_{k=1}^{\infty}\frac{(-p)^{k-1}(\lambda t)^k}{k!}. \tag{4}$$

We first take the expectation of this over the distribution of P. Note that the infinite series is absolutely convergent and therefore (by Fubini's theorem) the integral and sum can be interchanged.

$$\begin{aligned} E[X(t)\,|\,\lambda,a,b] &= \int_0^1 \sum_{k=1}^{\infty}\left\{\frac{(-p)^{k-1}(\lambda t)^k}{k!}\right\}\frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}\,dp \\ &= -\sum_{k=1}^{\infty}\frac{(-\lambda t)^k}{k!}\int_0^1 \frac{p^{a+k-2}(1-p)^{b-1}}{B(a,b)}\,dp \end{aligned}$$

For $k \geq 1$,

$$\begin{aligned} \int_0^1 \frac{p^{a+k-2}(1-p)^{b-1}}{B(a,b)}\,dp &= \frac{B(a+k-1,b)}{B(a,b)} \\ &= \frac{\Gamma(a+k-1)}{\Gamma(a)}\frac{\Gamma(a+b)}{\Gamma(a+b+k-1)} \\ &= \frac{a^{(k-1)}}{(a+b)^{(k-1)}}, \end{aligned}$$

where $a^{(n)}$ is the Pochhammer symbol for the rising factorial $a(a+1)\cdots(a+n-1)$. (Note that $(a-1)a^{(n-1)} = (a-1)^{(n)}$.)

[2]The same problem occurs in the equivalent derivation for the MBG/NBD model (Batislam et al. 2007, Hoppe and Wagner 2007).

Therefore,

$$\begin{aligned}
E[X(t) \mid \lambda, a, b] &= -\sum_{k=1}^{\infty} \frac{a^{(k-1)}}{(a+b)^{(k-1)}} \frac{(-\lambda t)^k}{k!} \\
&= -\frac{a+b-1}{a-1} \sum_{k=1}^{\infty} \frac{(a-1)^{(k)}}{(a+b-1)^{(k)}} \frac{(-\lambda t)^k}{k!} \\
&= \frac{a+b-1}{a-1} \left[1 - \sum_{k=0}^{\infty} \frac{(a-1)^{(k)}}{(a+b-1)^{(k)}} \frac{(-\lambda t)^k}{k!} \right] .
\end{aligned}$$

The last summation is recognized as the confluent hypergeoemetric function ${}_1F_1(a-1, a+b-1, -\lambda t)$ which is an *entire function* whose convergence is guaranteed everywhere in the complex plane $|\lambda t| < \infty$. This gives is

$$E[X(t)|\lambda, a, b] = \frac{a+b-1}{a-1} \big[1 - {}_1F_1(a-1, a+b-1, -\lambda t) \big] .$$

Next, we take the expectation of this over the distribution of Λ. Using Gradshteyn and Ryzhik (2007, equation 7.522.8)[3]:

$$\frac{\alpha^r}{\Gamma(r)} \int_0^{\infty} \lambda^{r-1} e^{-\alpha\lambda} {}_1F_1(a-1; a+b-1; -\lambda t) d\lambda = {}_2F_1(a-1, r; a+b-1; -t/\alpha) ,$$

which gives us

$$E[X(t) \mid r, \alpha, a, b] = \frac{a+b-1}{a-1} \big[1 - {}_2F_1(r, a-1; a+b-1; -t/\alpha) \big] . \tag{5}$$

Many numerical computing environments evaluate ${}_2F_1(a, b; c, z)$ by computing the hypergeometric series directly, which is absolutely convergent for $|z| < 1$ and divergent for $|z| > 1$. We rarely expect this condition for convergence to be met (i.e., $t/\alpha < 1$), so we apply the transformation

$${}_2F_1(a, b; c, z) = (1-z)^{-a} {}_2F_1\big(a, c-b; c, \tfrac{z}{z-1}\big) ,$$

giving us

$$E[X(t) \mid r, \alpha, a, b] = \frac{a+b-1}{a-1} \left[1 - \left(\frac{\alpha}{\alpha+t} \right)^r {}_2F_1\big(r, b; a+b-1; \tfrac{t}{\alpha+t}\big) \right] ,$$

which is exactly the same as (3).

Turning to the conditional expectation, the derivation developed in the appendix of Fader et al. (2005) sees us evaluating

$$\begin{aligned}
E[Y(t) \mid x, t_x, T, r, \alpha, a, b] &= \frac{1}{L(r, \alpha, a, b \mid x, t_x, T)} \\
&\quad \times \int_0^1 \int_0^{\infty} E[Y(t) \mid x, t_x, T, \lambda, p] f(\lambda \mid r, \alpha) f(p \mid a, b) \, d\lambda \, dp ,
\end{aligned} \tag{6}$$

[3]Also see Johnson et al. (1992, equation 1.111).

where

$$E[Y(t) \mid x, t_x, T, \lambda, p] = \frac{1}{p}(1-p)^x \lambda^x e^{-\lambda T} - \frac{1}{p}(1-p)^x \lambda^x e^{-\lambda(T+pt)}\,. \tag{7}$$

The first step taken in solving (6) is to break apart (7) and evaluate

$$\begin{aligned}
&\int_0^1 \int_0^\infty \frac{1}{p}(1-p)^x \lambda^x e^{-\lambda T} f(\lambda \mid r, \alpha) f(p \mid a, b)\, d\lambda\, dp \\
&\quad = \left\{ \int_0^1 \frac{1}{p}(1-p)^x \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}\, dp \right\} \left\{ \int_0^\infty \lambda^x e^{-\lambda T} \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)}\, d\lambda \right\} \\
&\quad = \left\{ \frac{1}{B(a,b)} \int_0^1 p^{a-2}(1-p)^{b+x-1}\, dp \right\} \left\{ \int_0^\infty \frac{\alpha^r \lambda^{r+x-1} e^{-\lambda(\alpha+T)}}{\Gamma(r)}\, d\lambda \right\} \\
&\quad = \frac{B(a-1, b+x)}{B(a,b)} \frac{\Gamma(r+x)}{\Gamma(r)} \left(\frac{\alpha}{\alpha+T}\right)^r \left(\frac{1}{\alpha+T}\right)^x.
\end{aligned}$$

In other words, exactly the same mistake is made as in the derivation of $E[X(t)]$.

As in the derivation of $E[X(t)]$ presented above, the solution is to find the Taylor series expansion of (7) around the point $t = 0$ and take the expectation of the result over the distributions of Λ and P. Rewriting (7) as

$$(1-p)^x \lambda^x e^{-\lambda T} \left\{ \frac{1}{p} - \frac{1}{p} e^{-\lambda p t} \right\}$$

and noting from (4) that

$$\frac{1}{p} - \frac{1}{p} e^{-\lambda p t} = \sum_{k=1}^{\infty} \frac{(-p)^{k-1}(\lambda t)^k}{k!}\,, \tag{8}$$

we get

$$E[Y(t) \mid x, t_x, T, \lambda, p] = -(1-p)^x \lambda^x e^{-\lambda T} \sum_{k=1}^{\infty} p^{k-1} \frac{(-\lambda t)^k}{k!}\,.$$

We first take the expectation of this over the distribution of P:

$$\begin{aligned}
&\int_0^1 \left\{ -\,(1-p)^x \lambda^x e^{-\lambda T} \sum_{k=1}^{\infty} p^{k-1} \frac{(-\lambda t)^k}{k!} \right\} \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}\, dp \\
&\qquad\qquad = -\lambda^x e^{-\lambda T} \sum_{k=1}^{\infty} \frac{(-\lambda t)^k}{k!} \int_0^1 \frac{p^{a+k-2}(1-p)^{b+x-1}}{B(a,b)}\, dp\,.
\end{aligned} \tag{9}$$

For $k \geq 1$,

$$\begin{aligned}
\int_0^1 \frac{p^{a+k-2}(1-p)^{b+x-1}}{B(a,b)}\, dp &= \frac{B(a+k-1, b+x)}{B(a,b)} \\
&= \frac{B(a+k-1, b+x)}{B(a,b)} \frac{B(a, b+x)}{B(a, b+x)} \\
&= \frac{B(a, b+x)}{B(a,b)} \frac{\Gamma(a+k-1)}{\Gamma(a)} \frac{\Gamma(a+b+x)}{\Gamma(a+b+x+k-1)} \\
&= \frac{B(a, b+x)}{B(a,b)} \frac{a^{(k-1)}}{(a+b+x)^{(k-1)}}\,,
\end{aligned}$$

Therefore (9) becomes,

$$
\begin{aligned}
-\frac{B(a,b+x)}{B(a,b)}&\lambda^x e^{-\lambda T}\sum_{k=1}^{\infty}\frac{a^{(k-1)}}{(a+b+x)^{(k-1)}}\frac{(-\lambda t)^k}{k!} \\
&= \frac{a+b+x-1}{a-1}\frac{B(a,b+x)}{B(a,b)}\lambda^x e^{-\lambda T}\left\{-\sum_{k=1}^{\infty}\frac{(a-1)^{(k)}}{(a+b+x-1)^{(k)}}\frac{(-\lambda t)^k}{k!}\right\} \\
&= \frac{a+b+x-1}{a-1}\frac{B(a,b+x)}{B(a,b)}\lambda^x e^{-\lambda T}\left\{1-\sum_{k=0}^{\infty}\frac{(a-1)^{(k)}}{(a+b+x-1)^{(k)}}\frac{(-\lambda t)^k}{k!}\right\} \\
&= \frac{a+b+x-1}{a-1}\frac{B(a,b+x)}{B(a,b)}\lambda^x e^{-\lambda T}\big[1-{}_1F_1(a-1;a+b+x-1;-\lambda t)\big]\,. \qquad (10)
\end{aligned}
$$

Next, we take the expectation of this over the distribution of Λ. Noting that

$$\int_0^{\infty}\lambda^x e^{-\lambda T}\frac{\alpha^r\lambda^{r-1}e^{-\alpha\lambda}}{\Gamma(r)}\,d\lambda = \frac{\Gamma(r+x)}{\Gamma(r)}\frac{\alpha^r}{(\alpha+T)^{r+x}}\,,$$

and

$$
\begin{aligned}
\int_0^{\infty}&\lambda^x e^{-\lambda T}{}_1F_1(a-1;a+b+x-1;-\lambda t)\frac{\alpha^r\lambda^{r-1}e^{-\alpha\lambda}}{\Gamma(r)}d\lambda \\
&= \frac{\alpha^r}{\Gamma(r)}\int_0^{\infty}\lambda^{r+x-1}e^{-\lambda(\alpha+T)}{}_1F_1(a-1;a+b+x-1;-\lambda t)d\lambda \\
&= \frac{\alpha^r}{\Gamma(r)}\Gamma(r+x)(\alpha+T)^{-(r+x)}{}_2F_1(a-1,r+x;a+b+x-1;-\tfrac{t}{\alpha+T}) \\
&= \frac{\Gamma(r+x)}{\Gamma(r)}\frac{\alpha^r}{(\alpha+T)^{r+x}}\left(\frac{\alpha+T}{\alpha+T+t}\right)^{r+x}{}_2F_1(r+x,b+x;a+b+x-1;\tfrac{t}{\alpha+T+t})\,,
\end{aligned}
$$

the expectation of (10) over the distribution of Λ equals

$$
\begin{aligned}
&\frac{a+b+x-1}{a-1}\frac{B(a,b+x)}{B(a,b)}\frac{\Gamma(r+x)}{\Gamma(r)}\frac{\alpha^r}{(\alpha+T)^{r+x}} \\
&\qquad\times\left[1-\left(\frac{\alpha+T}{\alpha+T+t}\right)^{r+x}{}_2F_1\big(r+x,b+x;a+b+x-1;\tfrac{t}{\alpha+T+t}\big)\right]\,.
\end{aligned}
$$

Substituting this in (6) and simplifying yields the expression for the conditional expectation given in Fader et al. (2005), equation (10).

The expressions for $E[X(t)]$ and $E[Y(t)\,|\,x,t_x,T]$ are not defined for $a=1$. We have not been able to derive separate expressions for this specific case. From a practical perspective, this is not an issue. Hoppe and Wagner (2007) show that the limit of these quantities exists as $a\to 1$. If we were ever to find ourselves with $\hat{a}=1$ (given the numerical precision of our computing environment), we would simply add or subtract machine epsilon and use the standard formulas.

4 Revisiting Derivations for the Pareto/NBD Model

Fader and Hardie (2005) present detailed derivations of key expressions for the Pareto/NBD model. It turns out that while the resulting expressions for $E[X(t)]$ and $E[Y(t)\,|\,x,t_x,T]$ are correct, the logic of the derivations is incorrect for a portion of parameter space. We first identify the problem with the original derivations and present alternative derivations that hold for the entire parameter space.

Conditional on λ and μ, the expected number of transactions in $(0, t]$ is

$$E[X(t) \,|\, \lambda, \mu] = \lambda\biggl(\frac{1}{\mu} - \frac{1}{\mu}e^{-\mu t}\biggr)\,. \tag{11}$$

We remove the conditioning on λ and μ by taking the expectation of this over the distributions of Λ and M. The expectation wrt to Λ is simply the mean of the gamma distribution, r/α. The expectation wrt M is performed in the following manner:

1) We first evaluate

$$\begin{aligned}\int_0^\infty \frac{1}{\mu}\frac{\beta^s\mu^{s-1}e^{-\beta\mu}}{\Gamma(s)}\,d\mu &= \frac{\beta}{s-1}\int_0^\infty \frac{\beta^{s-1}\mu^{s-2}e^{-\beta\mu}}{\Gamma(s-1)}\,d\mu \\ &= \frac{\beta}{s-1}\,.\end{aligned}$$

2) Next, we evaluate

$$\begin{aligned}\int_0^\infty \frac{1}{\mu}e^{-\mu t}\frac{\beta^s\mu^{s-1}e^{-\beta\mu}}{\Gamma(s)}\,d\mu &= \int_0^\infty \frac{\beta^s\mu^{s-2}e^{-\mu(\beta+t)}}{\Gamma(s)}\,d\mu \\ &= \frac{\beta}{s-1}\left(\frac{\beta}{\beta+t}\right)^{s-1}\int_0^\infty \frac{(\beta+t)^{s-1}\mu^{s-2}e^{-\mu(\beta+t)}}{\Gamma(s-1)}\,d\mu \\ &= \frac{\beta}{s-1}\left(\frac{\beta}{\beta+t}\right)^{s-1}\,.\end{aligned}$$

Combining these two interim results gives us

$$\frac{\beta}{(s-1)}\left[1-\left(\frac{\beta}{\beta+t}\right)^{s-1}\right],$$

which means the Pareto/NBD expression for the expected number of transactions in the interval $(0, t]$ is

$$E[X(t) \,|\, r, \alpha, s, \beta] = \frac{r\beta}{\alpha(s-1)}\left[1-\left(\frac{\beta}{\beta+t}\right)^{s-1}\right]\,. \tag{12}$$

The first mistake occurs in Step 1 where we "recognize" that the integrand is a gamma pdf and therefore the solution to the integral is 1. However, it is only a pdf for $s > 1$. The same mistake is made is Step 2. Therefore, this result is not true for $s \le 1$. (Furthermore, the expression can't be true because the initial integrand is strictly positive yet we get a negative number for $s < 1$.)

An alternative derivation uses the following result (Gradshteyn and Ryzhik 2007, equation 3.434.1):

$$\int_0^\infty \frac{e^{-\nu x} - e^{-\mu x}}{x^{\rho+1}}dx = \frac{\mu^\rho - \nu^\rho}{\rho}\Gamma(1-\rho)\,,\ \rho < 1.$$

Therefore, for $s > 0$,

$$\begin{aligned}
\int_0^\infty \left\{ \frac{1}{\mu} - \frac{1}{\mu} e^{-\mu t} \right\} \frac{\beta^s \mu^{s-1} e^{-\beta\mu}}{\Gamma(s)} \, d\mu &= \frac{\beta^s}{\Gamma(s)} \int_0^\infty \frac{e^{-\beta\mu} - e^{-(\beta+t)\mu}}{\mu^{(1-s)+1}} \, d\mu \\
&= \frac{\beta^s}{\Gamma(s)} \frac{(\beta+t)^{1-s} - \beta^{1-s}}{1-s} \Gamma\big(1 - (1-s)\big) \\
&= \frac{\beta}{s-1} \left[\left(\frac{\beta}{\beta+t} \right)^{s-1} \right] . \qquad (13)
\end{aligned}$$

Multiplying this by the mean of the distribution of Λ gives us (12).

The derivation of the conditional expectation sees us evaluating

$$\begin{aligned}
E[Y(t) \,|\, x, t_x, T, r, \alpha, s, \beta] = &\frac{1}{L(r, \alpha, s, \beta \,|\, x, t_x, T)} \\
&\times \int_0^\infty \int_0^\infty E[Y(t) \,|\, x, t_x, T, \lambda, \mu] f(\lambda \,|\, r, \alpha) f(\mu \,|\, s, \beta) \, d\lambda \, d\mu \, , \qquad (14)
\end{aligned}$$

where

$$\begin{aligned}
E[Y(t) \,|\, x, t_x, T, \lambda, \mu] &= \lambda^x e^{-(\lambda+\mu)T} \left\{ \frac{\lambda}{\mu} - \frac{\lambda}{\mu} e^{-\mu t} \right\} \\
&= \lambda^{x+1} e^{-\lambda T} \left\{ \frac{e^{-\mu T} - e^{-\mu(T+t)}}{\mu} \right\} .
\end{aligned}$$

Now,

$$\int_0^\infty \lambda^{x+1} e^{-\lambda T} \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)} \, d\lambda = \frac{\Gamma(r+x+1)}{\Gamma(r)} \frac{\alpha^r}{(\alpha+T)^{r+x+1}} . \qquad (15)$$

Similarly, following the logic of (13),

$$\int_0^\infty \left\{ \frac{e^{-\mu T} - e^{-\mu(T+t)}}{\mu} \right\} \frac{\beta^s \mu^{s-1} e^{-\beta\mu}}{\Gamma(s)} \, d\mu = \frac{1}{s-1} \left(\frac{\beta}{\beta+T} \right)^s \left[1 - \left(\frac{\beta+T}{\beta+T+t} \right)^{s-1} \right] . \qquad (16)$$

Substituting (15) and (16) in (14) and simplifying yields the expression for the conditional expectation reported in Fader and Hardie (2005).

4.1 The Case of $s = 1$

The expressions for $E[X(t)]$ and $E[Y(t) \,|\, x, t_x, T]$ are not defined for $s = 1$. We can derive an expression for $E[X(t)]$ for the case of $s = 1$ using the following result (Gradshteyn and Ryzhik 2007, equation 3.434.2):

$$\int_0^\infty \frac{e^{-\nu x} - e^{-\mu x}}{x} dx = \ln \left(\frac{\nu}{\mu} \right) .$$

Therefore, for $s = 1$,

$$\begin{aligned}
\int_0^\infty \left\{ \frac{1}{\mu} - \frac{1}{\mu} e^{-\mu t} \right\} \frac{\beta^s \mu^{s-1} e^{-\beta\mu}}{\Gamma(s)} \, d\mu &= \frac{\beta}{\Gamma(1)} \int_0^\infty \frac{e^{-\beta\mu} - e^{-(\beta+t)\mu}}{\mu} \, d\mu \\
&= \beta \ln \left(\frac{\beta+t}{\beta} \right) .
\end{aligned}$$

The resulting expression for the expected number of transactions in $(0, t]$ is

$$E[X(t) \,|\, r, \alpha, s = 1, \beta] = \frac{r\beta}{\alpha} \ln\left(\frac{\beta + t}{\beta}\right) . \tag{17}$$

This is a new result.

Following the same logic, we can derive an expression for $E[Y(t) \,|\, x, t_x, T]$ for the special case of $s = 1$.

5 Revisiting Derivations for the BG/BB Model

Finally, we revisit the derivations of the mean and conditional expectation for the BG/BB model (Fader et al. 2010). Once again, while the resulting expressions are correct, the logic of the derivations is incorrect for a portion of parameter space. We now present an alternative derivation that holds for the entire parameter space.

Conditional on p and θ, the expected number of transactions over n transaction opportunities is

$$\begin{aligned} E[X(n) \,|\, p, \theta] &= p \sum_{t=1}^{n} (1-\theta)^t \qquad (18) \\ &= \frac{p(1-\theta)}{\theta} - \frac{p(1-\theta)^{n+1}}{\theta} . \qquad (19) \end{aligned}$$

Taking the expectation of (19) over the distributions of P and Θ — beta distributions with parameters (α, β) and (γ, δ), respectively — gives us, after a few initial steps,

$$\begin{aligned} E[X(n) \,|\, \alpha, \beta, \gamma, \delta] &= \frac{\alpha}{\alpha+\beta} \left\{ \int_0^1 \frac{\theta^{\gamma-2}(1-\theta)^{\delta}}{B(\gamma, \delta)} \, d\theta - \int_0^1 \frac{\theta^{\gamma-2}(1-\theta)^{\delta+n}}{B(\gamma, \delta)} \, d\theta \right\} \\ &= \left(\frac{\alpha}{\alpha+\beta}\right) \left\{ \frac{B(\gamma-1, \delta+1) - B(\gamma-1, \delta+n+1)}{B(\gamma, \delta)} \right\} \\ &= \left(\frac{\alpha}{\alpha+\beta}\right) \left(\frac{\delta}{\gamma-1}\right) \left\{ 1 - \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma+\delta+n)} \frac{\Gamma(1+\delta+n)}{\Gamma(1+\delta)} \right\} . \qquad (20) \end{aligned}$$

Exactly the same mistake is made as in the derivation of the mean of the BG/NBD model: we "recognize" that the definite integrals are beta functions, but the result is not true for $\gamma \leq 1$.

An alternative derivation sees us taking the expectation of (18) over the distributions of P and Θ, giving us

$$E[X(n) \,|\, \alpha, \beta, \gamma, \delta] = \left(\frac{\alpha}{\alpha+\beta}\right) \sum_{t=1}^{n} \frac{B(\gamma, \delta+t)}{B(\gamma, \delta)} \tag{21}$$

If we can show that

$$\sum_{t=1}^{n} B(\gamma, \delta+t) = B(\gamma-1, \delta+1) - B(\gamma-1, \delta+n+1) \tag{22}$$

then (20) is a solution to (21). We do so by induction.

Let us first consider the case of $n = 1$:

$$
\begin{aligned}
B(\gamma, \delta+1) &= \frac{\Gamma(\gamma)\Gamma(\delta+1)}{\Gamma(\gamma+\delta+1)} \\
&= \frac{\Gamma(\gamma-1)\Gamma(\delta+1)}{\Gamma(\gamma+\delta)}\left(\frac{\gamma-1}{\gamma+\delta}\right) \\
&= \frac{\Gamma(\gamma-1)\Gamma(\delta+1)}{\Gamma(\gamma+\delta)}\left(1-\frac{\delta+1}{\gamma+\delta}\right) \\
&= B(\gamma-1, \delta+1) - B(\gamma-1, \delta+2)\,.
\end{aligned}
$$

Therefore (22) is true when $n = 1$.

We now assume that (22) is true for n:

$$\sum_{t=1}^{n} B(\gamma, \delta+t) = B(\gamma-1, \delta+1) - B(\gamma-1, \delta+n+1)\,.$$

It follows that

$$\sum_{t=1}^{n+1} B(\gamma, \delta+t) = B(\gamma-1, \delta+1) - B(\gamma-1, \delta+n+1) + B(\gamma, \delta+n+1)\,. \tag{23}$$

Now

$$
\begin{aligned}
B(\gamma, \delta+n+1) &= \frac{\Gamma(\gamma)\Gamma(\delta+n+1)}{\Gamma(\gamma+\delta+n+1)} \\
&= \frac{\Gamma(\gamma-1)\Gamma(\delta+n+1)}{\Gamma(\gamma+\delta+n)}\left(\frac{\gamma-1}{\gamma+\delta+n}\right) \\
&= \frac{\Gamma(\gamma-1)\Gamma(\delta+n+1)}{\Gamma(\gamma+\delta+n)}\left(1-\frac{\delta+n+1}{\gamma+\delta+n}\right) \\
&= B(\gamma-1, \delta+n+1) - B(\gamma-1, \delta+n+2)\,.
\end{aligned}
$$

Substituting this in (23) gives us

$$\sum_{t=1}^{n+1} B(\gamma, \delta+t) = B(\gamma-1, \delta+1) - B(\gamma-1, \delta+n+2)\,.$$

In summary, we have shown in the basis step that the relationship is true for $n = 1$. We have also shown in the induction step that if the relationship is true for integer n, then it is true for $n + 1$. Thus it must be true for all positive integers n.

QED.

Therefore, while the standard derivation is fallacious when $\gamma \leq 1$, the resulting equation is still correct.

Turning our attention to the conditional expectation, Fader et al. (2010) use a slightly different notation for the expected number of transactions across the next n^* transaction opportunities for an individual with observed behavior (x, t_x, n), $E[X(n, n+n^*) \,|\, x, t_x, n]$.

Three important intermediate results are:

- Conditional on (p, θ) and being alive at n,

$$E[X(n, n+n^*) \,|\, p, \theta, \text{alive at } n] = \sum_{s=1}^{n^*} p(1-\theta)^s \tag{24}$$

$$= \frac{p(1-\theta)}{\theta} - \frac{p(1-\theta)^{n^*+1}}{\theta}\,. \tag{25}$$

- Conditional on (p, θ) , the probability that a customer with purchase history (x, t_x, n) is "alive" at n is given by

$$P(\text{alive at } n \,|\, p, \theta; x, t_x, n) = \frac{p^x(1-p)^{n-x}(1-\theta)^n}{L(p, \theta \,|\, x, t_x, n)}\,. \tag{26}$$

- By Bayes' theorem, the joint posterior distribution of P and Θ for a customer with purchase history (x, t_x, n) is

$$g(p, \theta \,|\, \alpha, \beta, \gamma, \delta; x, t_x, n) = \frac{L(p, \theta \,|\, x, t_x, n) g(p \,|\, \alpha, \beta) g(\theta \,|\, \gamma, \delta)}{L(\alpha, \beta, \gamma, \delta \,|\, x, t_x, n)}\,. \tag{27}$$

An expression for the conditional expectation is obtained by solving

$$\begin{aligned} &E[X(n, n+n^*) \,|\, \alpha, \beta, \gamma, \delta; x, t_x, n] \\ &\quad = \int_0^1 \int_0^1 \Big\{ E[X(n, n+n^*) \,|\, p, \theta, \text{alive at } n] \, P(\text{alive at } n \,|\, p, \theta; x, t_x, n) \\ &\qquad\qquad \times g(p, \theta \,|\, \alpha, \beta, \gamma, \delta; x, t_x, n) \Big\} dp \, d\theta\,. \end{aligned} \tag{28}$$

Substituting (25)–(27) in (28) and solving gives us

$$\begin{aligned} &E[X(n, n+n^*) \,|\, \alpha, \beta, \gamma, \delta; x, t_x, n] \\ &\quad = \frac{1}{L(\alpha, \beta, \gamma, \delta \,|\, x, t_x, n)} \frac{B(\alpha+x+1, \beta+n-x)}{B(\alpha, \beta)} \\ &\qquad \times \left(\frac{\delta}{\gamma-1}\right) \frac{\Gamma(\gamma+\delta)}{\Gamma(1+\delta)} \left\{ \frac{\Gamma(1+\delta+n)}{\Gamma(\gamma+\delta+n)} - \frac{\Gamma(1+\delta+n+n^*)}{\Gamma(\gamma+\delta+n+n^*)} \right\}. \end{aligned}$$

The final line of this expression results from simplifying the result

$$\begin{aligned} &\int_0^1 \left\{ \frac{(1-\theta)}{\theta} - \frac{(1-\theta)^{n^*+1}}{\theta} \right\} (1-\theta)^n \frac{\theta^{\gamma-1}(1-\theta)^{\delta-1}}{B(\gamma, \delta)} \, d\theta \\ &\qquad = \frac{B(\gamma-1, \delta+n+1) - B(\gamma-1, \delta+n+n^*+1)}{B(\gamma, \delta)}\,. \end{aligned}$$

The now-familiar mistake is made when solving the integral. Using (24) in place of (25) and changing the order of integration and summation results in

$$\sum_{s=1}^{n^*} \int_0^1 (1-\theta)^s (1-\theta)^n \frac{\theta^{\gamma-1}(1-\theta)^{\delta-1}}{B(\gamma, \delta)} \, d\theta = \sum_{s=1}^{n^*} \frac{B(\gamma, \delta+n+s)}{B(\gamma, \delta)}\,.$$

It follows from our derivation of $E[X(n)]$ that

$$\sum_{s=1}^{n^*} \frac{B(\gamma, \delta+n+s)}{B(\gamma, \delta)} = \frac{B(\gamma-1, \delta+n+1) - B(\gamma-1, \delta+n+n^*+1)}{B(\gamma, \delta)}\,.$$

Therefore, while the standard derivation is fallacious when $\gamma \leq 1$, the resulting equation is still correct.

5.1 The Case of $\gamma = 1$

We note that (20) is not defined for $\gamma = 1$. If we were ever to face such a situation, we could compute the mean using (21). However, we can derive an expression for $E[X(n)]$ when $\gamma = 1$ in the following manner.

We first note that

$$\psi(z+n) = \psi(z) + \sum_{i=0}^{n-1} \frac{1}{z+i}\,,$$

which follows from the digamma function recurrence formula

$$\psi(z+1) = \psi(z) + \frac{1}{z}\,.$$

Now, for $\gamma = 1$,

$$\begin{aligned}
\sum_{t=1}^{n} \frac{B(\gamma, \delta + t)}{B(\gamma, \delta)} &= \sum_{t=1}^{n} \frac{B(1, \delta + t)}{B(1, \delta)} \\
&= \sum_{t=1}^{n} \frac{\delta}{\delta + t} \\
&= \delta \sum_{t=0}^{n} \frac{1}{\delta + t} - 1 \\
&= \delta\psi(\delta + n + 1) - \delta\psi(\delta) - 1
\end{aligned}$$

which by the recurrence relationship of the digamma function

$$= \delta\psi(\delta + n + 1) - \delta\psi(\delta + 1)\,.$$

It follows that

$$E[X(n) \,|\, \alpha, \beta, \gamma = 1, \delta] = \left(\frac{\alpha\delta}{\alpha + \beta}\right) \big[\psi(\delta + n + 1) - \psi(\delta + 1)\big]\,. \tag{29}$$

This is a new result.

Following the same logic, we can derive an expression for $E[X(n, n + n^*) \,|\, x, t_x, n]$. for the special case of $\gamma = 1$.

6 Conclusion

We have shown how the derivations of expressions for the mean and conditional expectation for the BG/NBD, Pareto/NBD, and BG/BB models are incorrect for a portion of the parameter space, and have presented alternatives derivations that holds for the entire parameter space.

We suspect that the same error has been made in the derivation of expressions for these quantities in other latent-attrition models of buyer behavior in noncontractual settings where the lifetimes are distributed geometric with beta heterogeneity, or exponential/Erlang/gamma with gamma heterogeneity. Our results would suggest that the derived expressions are not incorrect; it is simply the case that the derivations themselves are not true for the entire parameter space.

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.

Fader, Peter S. and Bruce G. S. Hardie (2005), “A Note on Deriving the Pareto/NBD Model and Related Expressions.” <http://brucehardie.com/notes/009/>

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.

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2019), “A Step-by-Step Derivation of the BG/NBD Model.” <http://brucehardie.com/notes/039/>

Fader, Peter S., Bruce G. S. Hardie, and Jen Shang (2010), “Customer-Base Analysis in a Discrete-Time Noncontractual Setting,” *Marketing Science*, **29** (November–December), 1086–1108.

Gradshteyn, I. S. and I. M. Ryzhik (2007), *Table of Integrals, Series, and Products*, 7th edition, Burlington, MA: Academic Press.

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.

Johnson, Norman L., Samuel Kotz, and Adrienne W. Kemp (1992), *Univariate Discrete Distributions*, 2nd edition, New York, NY: John Wiley & Sons, Inc.