A Step-by-Step Derivation of the BG/NBD Model

Peter S. Fader
www.petefader.com

Bruce G. S. Hardie
www.brucehardie.com

Ka Lok Lee†

November 2019

1 Introduction

This note presents a detailed derivation of the various mathematical results presented in Fader et al. (2005), along with some additional results. As such, it should be of interest to those students of probability models who are interested in expanding their math skills beyond those required for simple models such as the beta-binomial and NBD.

Since the publication of Fader et al. (2005), the notation and terminology used in our work on latent-attrition models has evolved. In this note we use the notation and terminology used in the original paper.

2 Model Assumptions

The BG/NBD model is based on the following assumptions (the first two of which are identical to the corresponding Pareto/NBD assumptions):

i. While active, the number of transactions made by a customer follows a Poisson process with transaction rate λ. This is equivalent to assuming that the time between transactions is distributed exponential with transaction rate λ:

$$f(t_j \,|\, t_{j-1}; \lambda) = \lambda e^{-\lambda(t_j - t_{j-1})}\,, \quad t_j > t_{j-1} \geq 0\,.$$

ii. Heterogeneity in λ follows a gamma distribution with pdf

$$f(\lambda \,|\, r, \alpha) = \frac{\alpha^r \lambda^{r-1} e^{-\lambda\alpha}}{\Gamma(r)}\,, \quad \lambda > 0\,. \tag{1}$$

iii. After any transaction, a customer becomes inactive with probability p. Therefore the point at which the customer "drops out" is distributed across transactions according to a (shifted) geometric distribution with pmf

$$P(\text{inactive immediately after } j\text{th transaction}) \\ = p(1-p)^{j-1}\,, \quad j = 1, 2, 3, \ldots\,.$$

We assume that a customer is alive at the beginning of the observation period. Therefore, a customer cannot "die" before he makes his first transaction.

iv. Heterogeneity in p follows a beta distribution with pdf

$$f(p\,|\,a,b) = \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}\,, \quad 0 \le p \le 1\,, \tag{2}$$

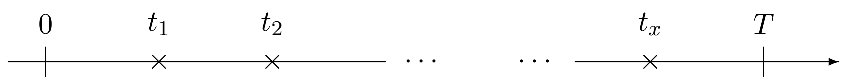
where $B(a,b)$ is the beta function, which can be expressed in terms of gamma functions: $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$.

v. The transaction rate λ and the dropout probability p vary independently across customers.

3 Model Development at the Individual Level

3.1 Derivation of the Likelihood Function

Consider a customer making x transactions in the the time interval $(0, T]$ with the transactions occurring at $t_1, t_2, \ldots, t_x$:



We derive the individual-level likelihood function in the following manner:

- the likelihood of the first transaction occurring at t_1 is the standard exponential likelihood component, which equals $\lambda e^{-\lambda t_1}$.
- the likelihood of the second transaction occurring at t_2 is the probability of not dying at t_1 times the standard exponential likelihood component, which equals $(1-p)\lambda e^{-\lambda(t_2-t_1)}$.

  $\cdots$
- the likelihood of the xth transaction occurring at t_x is the probability of not dying at t_{x-1} times the standard exponential likelihood component, which equals $(1-p)\lambda e^{-\lambda(t_x-t_{x-1})}$.
- the likelihood of observing zero purchases in $(t_x, T]$ is the probability the customer died at t_x, plus the probability he survived *and* made no purchases in this interval, which equals $p + (1-p)e^{-\lambda(T-t_x)}$.

Therefore,

$$\begin{aligned}L(\lambda, p \,|\, t_1, t_2, \ldots, t_x, T) &= \lambda e^{-\lambda t_1}(1-p)\lambda e^{-\lambda(t_2-t_1)} \cdots (1-p)\lambda e^{-\lambda(t_x-t_{x-1})} \\ &\quad \times \left\{p + (1-p)e^{-\lambda(T-t_x)}\right\} \\ &= p(1-p)^{x-1}\lambda^x e^{-\lambda t_x} + (1-p)^x \lambda^x e^{-\lambda T}\,.\end{aligned}$$

We note that information on when each of the x transactions occurred is not required; a sufficient summary of the customer's purchase history is $(X = x, t_x, T)$.

What about the case of a customer making no purchases in $(0, T]$? Given our assumption that the customer was alive at the beginning of the observation period, the associated likelihood function is the exponential survival function:

$$L(\lambda \,|\, X = 0, T) = e^{-\lambda T}\,.$$

We can therefore write the individual-level likelihood function as

$$L(\lambda, p \,|\, X = x, T) = (1-p)^x \lambda^x e^{-\lambda T} + \delta_{x>0}\, p(1-p)^{x-1}\lambda^x e^{-\lambda t_x}\,, \tag{3}$$

where $\delta_{x>0} = 1$ if $x > 0$, 0 otherwise.

3.2 Derivation of $P(X(t) = x)$

Let the random variable $X(t)$ denote the number of transactions occurring in a time period of length t (with a time origin of 0). To derive an expression for $P(X(t) = x)$, we recall the fundamental relationship between inter-transaction times and the number of transactions,

$$X(t) \geq x \Leftrightarrow T_x \leq t\,,$$

where T_x is the random variable denoting the time of the xth transaction. This implies

$$\begin{aligned}P(X(t) = x) &= P(X(t) \geq x) - P(X(t) \geq x+1) \\ &= P(T_x \leq t) - P(T_{x+1} \leq t)\,.\end{aligned}$$

Given our assumption regarding the nature of the death process,

$$\begin{aligned}P(X(t) = x) &= P(\text{alive after } x\text{th purchase}) \times P(T_x \leq t \text{ and } T_{x+1} > t) \\ &\quad + \delta_{x>0} \times P(\text{dies after } x\text{th purchase}) \times P(T_x \leq t)\,.\end{aligned}$$

Given the assumption of exponentially distributed inter-transaction times, the second term above is simply the Poisson probability that $X(t) = x$ and the final term is the Erlang-x cdf. Therefore

$$\begin{aligned}&P(X(t) = x \,|\, \lambda, p) \\ &\qquad = (1-p)^x \frac{(\lambda t)^x e^{-\lambda t}}{x!} + \delta_{x>0}\, p(1-p)^{x-1}\left[1 - e^{-\lambda t}\sum_{j=0}^{x-1}\frac{(\lambda t)^j}{j!}\right]. \end{aligned}\tag{4}$$

As a useful exercise in mathematics, let us prove to ourselves that this expression is indeed a pmf. To do so, we need to show that

$$\sum_{x=0}^{\infty} P(X(t) = x \mid \lambda, p) = 1 .$$

Recalling (4), we can write

$$\sum_{x=0}^{\infty} P(X(t) = x \mid \lambda, p) = \mathsf{A} + \mathsf{B}$$

where

$$\mathsf{A} = \sum_{x=0}^{\infty} (1-p)^x \frac{(\lambda t)^x e^{-\lambda t}}{x!} , \text{ and}$$

$$\mathsf{B} = \sum_{x=1}^{\infty} p(1-p)^{x-1} \left[1 - e^{-\lambda t} \sum_{j=0}^{x-1} \frac{(\lambda t)^j}{j!} \right] .$$

- Rearranging the terms in A, we have

$$\mathsf{A} = e^{-\lambda p t} \sum_{x=0}^{\infty} \frac{[\lambda(1-p)t]^x e^{-\lambda(1-p)t}}{x!}$$

  which, since the summand is a Poisson pmf with mean $\lambda(1-p)t$ and the sum therefore equals 1,

$$= e^{-\lambda p t} .$$

- The trick to solving B is to recognize that the bracketed term in the summand is the Erlang-x cdf and to replace it with the associated integral representation:

$$\begin{aligned}
\mathsf{B} &= \sum_{x=1}^{\infty} p(1-p)^{x-1} \int_0^t \frac{\lambda^x u^{x-1} e^{-\lambda u}}{(x-1)!} \, du \\
&= \int_0^t \lambda p \left\{ \sum_{x=1}^{\infty} \frac{[\lambda(1-p)u]^{x-1} e^{-\lambda u}}{(x-1)!} \right\} du \\
&= \int_0^t \lambda p e^{-\lambda p u} \left\{ \sum_{y=0}^{\infty} \frac{[\lambda(1-p)u]^y e^{-\lambda(1-p)u}}{y!} \right\} du
\end{aligned}$$

  which, since the summand is a Poisson pmf,

$$\begin{aligned}
&= \int_0^t \lambda p e^{-\lambda p u} \, du \\
&= 1 - e^{-\lambda p t} .
\end{aligned}$$

- Since $\mathsf{A} + \mathsf{B} = 1$, we can say that (4) is a pmf.

3.3 Derivation of P(alive at t)

Given the model assumptions, how can a customer be alive at time t? They made no purchases in the time interval $(0, t]$. Or they made only one purchase and survived the flip of the "death" coin. Or they made only two purchases and survived the two flips of the "death" coin. And so on. Therefore, conditional on λ and p, the probability that a customer is alive at time t is simply

$$\begin{aligned} P(\text{alive at } t \,|\, \lambda, p) &= \sum_{j=0}^{\infty} (1-p)^j \frac{(\lambda t)^j e^{-\lambda t}}{j!} \\ &= e^{-\lambda p t} \sum_{j=0}^{\infty} \frac{(\lambda(1-p)t)^j e^{-\lambda(1-p)t}}{j!} \end{aligned}$$

which. since the summand is a Poisson pmf,

$$= e^{-\lambda p t} \,. \tag{5}$$

Letting τ denote the time at which the customer dies, we therefore have

$$P(\tau > t \,|\, \lambda, p) = e^{-\lambda p t} \,.$$

This implies that the pdf of the death time is given by

$$g(\tau \,|\, \lambda, p) = \lambda p e^{-\lambda p \tau} \,.$$

(Note that this is dependent on the transaction rate λ. In contrast, the Pareto/NBD has an exponential death process independent of the transaction rate.)

3.4 Derivation of $E[X(t)]$

Given that the number of transactions follows a Poisson process,

- if τ, the time at which the customer dies, is greater than t, $E[X(t)]$ is simply λt.
- if $\tau \leq t$, the expected number of transactions in $(0, \tau]$ is $\lambda \tau$.

Therefore,

$$\begin{aligned} E(X(t) \,|\, \lambda, p) &= (\lambda t) P(\tau > t) + \int_0^t \lambda \tau g(\tau \,|\, \lambda, p) d\tau \\ &= \lambda t e^{-\lambda p t} + \lambda^2 p \int_0^t \tau e^{-\lambda p \tau} d\tau \,. \end{aligned} \tag{6}$$

Integrating by parts,

$$\begin{aligned} \int_0^t \tau e^{-\lambda p \tau} d\tau &= \frac{-\tau}{\lambda p} e^{-\lambda p \tau} \Big|_0^t + \int_0^t \frac{1}{\lambda p} e^{-\lambda p \tau} d\tau \\ &= -\frac{t}{\lambda p} e^{-\lambda p t} + \left\{ \frac{-1}{\lambda^2 p^2} e^{-\lambda p \tau} \Big|_0^t \right\} \\ &= -\frac{t}{\lambda p} e^{-\lambda p t} + \frac{1}{\lambda^2 p^2} - \frac{1}{\lambda^2 p^2} e^{-\lambda p t} \,. \end{aligned}$$

Substituting this in (6), we get

$$E(X(t)\,|\,\lambda,p)=\frac{1}{p}-\frac{1}{p}e^{-\lambda pt}\,. \tag{7}$$

An alternative approach to the derivation of this expression is to use first principles:

$$E[X(t)]=\sum_{x=0}^{\infty} xP(X(t)=x)\,.$$

As an exercise in mathematics, let's do this. Recalling (4), we can write

$$\begin{aligned}E[X(t)\,|\,\lambda,p] &= \sum_{x=0}^{\infty} xP(X(t)=x\,|\,\lambda,p)\\ &= \mathsf{A}+\mathsf{B}\,,\end{aligned}$$

where

$$\begin{aligned}\mathsf{A} &= \sum_{x=0}^{\infty} x(1-p)^x\frac{(\lambda t)^x e^{-\lambda t}}{x!}\,,\ \text{ and}\\ \mathsf{B} &= \sum_{x=1}^{\infty} xp(1-p)^{x-1}\left[1-e^{-\lambda t}\sum_{j=0}^{x-1}\frac{(\lambda t)^j}{j!}\right].\end{aligned}$$

- Rearranging the terms in A, we have

  $$\mathsf{A}=e^{-\lambda pt}\sum_{x=0}^{\infty} x\underbrace{\frac{[\lambda(1-p)t]^x e^{-\lambda(1-p)t}}{x!}}_{*}\,.$$

  Since $*$ is a Poisson pmf with mean $\lambda(1-p)t$, the sum gives us the mean of this Poisson distribution. Therefore,

  $$\mathsf{A}=\lambda(1-p)te^{-\lambda pt}\,.$$

- Recognizing that the bracketed term in the summand of B is the Erlang-x cdf, and replacing it with the associated integral representation,

  $$\begin{aligned}\mathsf{B} &= \sum_{x=1}^{\infty} xp(1-p)^{x-1}\int_0^t \frac{\lambda^x u^{x-1}e^{-\lambda u}}{(x-1)!}\,du\\ &= \int_0^t \lambda p\left\{\sum_{x=1}^{\infty} x\frac{[\lambda(1-p)u]^{x-1}e^{-\lambda u}}{(x-1)!}\right\}du\\ &= \int_0^t \lambda pe^{-\lambda pu}\left\{\sum_{y=0}^{\infty}(y+1)\frac{[\lambda(1-p)u]^y e^{-\lambda(1-p)u}}{y!}\right\}du\end{aligned}$$

which, recognizing that the bracketed term is $E(Y)+1$ where Y is distributed Poisson with mean $\lambda(1-p)u$,

$$
\begin{aligned}
&= \int_0^t \lambda p e^{-\lambda p u}[\lambda(1-p)u + 1]\, du \\
&= \mathsf{B}_1 + \mathsf{B}_2 \,.
\end{aligned}
$$

where

$$
\mathsf{B}_1 = \int_0^t \lambda^2 p(1-p) u e^{-\lambda p u}\, du \quad \text{and } \mathsf{B}_2 = \int_0^t \lambda p e^{-\lambda p u}\, du \,.
$$

- Now

$$
\mathsf{B}_1 = \frac{1-p}{p} \int_0^t (\lambda p)^2 u e^{-\lambda p u}\, du
$$

which, recognizing that the integrand is an Erlang-2 pdf,

$$
= \left(\frac{1}{p} - 1\right)\left[1 - e^{-\lambda p t}(1 + \lambda p t)\right] .
$$

- Turning to B_2,

$$
\begin{aligned}
\mathsf{B}_2 &= \int_0^t \lambda p e^{-\lambda p u}\, du \\
&= 1 - e^{-\lambda p t} \,.
\end{aligned}
$$

- It follows that

$$
\mathsf{A} + \mathsf{B}_1 + \mathsf{B}_2 = \frac{1}{p} - \frac{1}{p} e^{-\lambda p t}
$$

which is the expression for $E(X(t) \mid \lambda, p)$ given in (7).

4 Moving to a Randomly Chosen Individual

All the expressions developed above are conditional on the transaction rate λ and the death probability p, both of which are unobserved. To arrive at the equivalent expressions for a randomly chosen customer, we take the expectation of the individual-level results over the distributions of λ and p.

Before doing this, let us note the following two results:

$$
\int_0^\infty \lambda^j e^{-\lambda t} \frac{\alpha^r \lambda^{r-1} e^{-\lambda \alpha}}{\Gamma(r)}\, d\lambda = \frac{\Gamma(r+j)\alpha^r}{\Gamma(r)(\alpha+t)^{r+j}} \,, \tag{8}
$$

and

$$
\int_0^1 p^j (1-p)^k \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}\, dp = \frac{B(a+j, b+k)}{B(a,b)} \,. \tag{9}
$$

4.1 Derivation of the Likelihood Function

The likelihood function for a randomly chosen customer with purchase history $(X = x, t_x, T)$ is given by

$$L(r, \alpha, a, b \mid X = x, t_x, T) = \int_0^1 \int_0^\infty L(\lambda, p \mid X = x, t_x, T) f(\lambda \mid r, \alpha) f(p \mid a, b) d\lambda dp \,. \quad (10)$$

Substituting (1), (2), and (3) in (10), and recalling (8) and (9), we have

$$\begin{aligned} L(r, \alpha, a, b \mid X = x, t_x, T) &= \frac{B(a, b+x)}{B(a,b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+T)^{r+x}} \\ &\quad + \delta_{x>0} \frac{B(a+1, b+x-1)}{B(a,b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+t_x)^{r+x}} \,. \end{aligned} \quad (11)$$

4.2 Derivation of $P(X(t) = x)$

For a randomly chosen customer,

$$P(X(t) = x \mid r, \alpha, a, b) = \int_0^1 \int_0^\infty P(X(t) = x \mid \lambda, p) f(\lambda \mid r, \alpha) f(p \mid a, b) d\lambda \, dp \,. \quad (12)$$

Substituting (1), (2), and (4) in (12), and recalling (8) and (9), we have

$$\begin{aligned} &P(X(t) = x | r, \alpha, a, b) \\ &= \frac{B(a, b+x)}{B(a,b)} \frac{\Gamma(r+x)}{\Gamma(r)x!} \left(\frac{\alpha}{\alpha+t}\right)^r \left(\frac{t}{\alpha+t}\right)^x \\ &\quad + \delta_{x>0} \frac{B(a+1, b+x-1)}{B(a,b)} \\ &\quad \times \left[1 - \left(\frac{\alpha}{\alpha+t}\right)^r \left\{\sum_{j=0}^{x-1} \frac{(\Gamma(r+j)}{\Gamma(r)j!} \left(\frac{t}{\alpha+t}\right)^j\right\}\right] . \end{aligned} \quad (13)$$

4.3 Derivation of P(alive at t)

For a randomly chosen customer

$$P(\text{alive at } t \mid r, \alpha, a, b) = \int_0^1 \int_0^\infty P(\text{alive at } t \mid \lambda, p) f(\lambda \mid r, \alpha) f(p \mid a, b) d\lambda \, dp \,. \quad (14)$$

Substituting (1), (2), and (5) in (14), we have

$$\begin{aligned} P(\text{alive at } t \mid r, \alpha, a, b) &= \int_0^1 \int_0^\infty e^{-\lambda pt} \frac{\alpha^r \lambda^{r-1} e^{-\lambda\alpha}}{\Gamma(r)} \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)} d\lambda dp \\ &= \int_0^1 \left\{ \int_0^\infty \frac{\alpha^r \lambda^{r-1} e^{-\lambda(\alpha+pt)}}{\Gamma(r)} d\lambda \right\} \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)} dp \\ &= \frac{\alpha^r}{B(a,b)} \int_0^1 p^{a-1}(1-p)^{b-1}(\alpha+pt)^{-r} dp \end{aligned}$$

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

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

Recalling Euler's integral for the Gaussian hypergeometric function,

$${}_2F_1(a,b;c;z) = \frac{1}{B(b,c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-zt)^{-a} dt\,, \quad c > b\,,$$

we have

$$P(\text{alive at } t \mid r, \alpha, a, b) = \left(\frac{\alpha}{\alpha+t}\right)^r {}_2F_1\bigl(r, b; a+b; \tfrac{t}{\alpha+t}\bigr)\,. \tag{15}$$

4.4 Derivation of $E[X(t)]$

To arrive at an expression for $E[X(t)]$ for a randomly chosen customer, we need to take the expectation of (7) over the distributions of λ and p. First we take the expectation with respect to λ, giving us

$$E(X(t) \mid r, \alpha, p) = \frac{1}{p} - \frac{\alpha^r}{p(\alpha + pt)^r}\,.$$

The next step is to take the expectation of this over the distribution of p. We first evaluate

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

Next, we evaluate

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

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

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

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

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

It follows that

$$E(X(t)\,|\,r,\alpha,a,b)$$
$$= \frac{a+b-1}{a-1}\left[1-\left(\frac{\alpha}{\alpha+t}\right)^{r}{}_2F_1\big(r,b;a+b-1;\tfrac{t}{\alpha+t}\big)\right]. \quad (16)$$

5 Making Conditional Predictions

In order for the BG/NBD model to be of greatest use in a forward-looking customer-base analysis exercise, we need to derive expressions for i) the probability that an individual with observed behavior $(X = x, t_x, T)$ is still active at time T, and ii) the expected number of transactions in a future period of length t for an individual with observed behavior $(X = x, t_x, T)$. We now derive the expressions for these quantities, along with an expression for the probability of making y purchases in a future time period given the observed past behavior.

5.1 Derivation of $P(\text{alive at } T\,|\,X = x, t_x, T)$

Given the (implicit) model assumption that a customer cannot die before he has made any transactions,

$$P(\text{alive at } T\,|\,X = 0, T, r, \alpha, a, b) = 1\,.$$

For the case where purchases were made in $(0, T]$, the probability that a customer with purchase history $(X = x, t_x, T)$ is alive at T, conditional on λ and p, is simply the probability that he did not die at t_x *and* made no purchase in $(t_x, T]$, divided by the probability of making no purchases in this same period. Recalling that this second probability is simply the probability that the customer died at t_x, plus the probability he survived *and* made no purchases in this interval, we have

$$P(\text{alive at } T\,|\,X = x, t_x, T, \lambda, p) = \frac{(1-p)e^{-\lambda(T-t_x)}}{p+(1-p)e^{-\lambda(T-t_x)}}\,.$$

Multiplying this by $[(1-p)^{x-1}\lambda^x e^{-\lambda t_x}]/[(1-p)^{x-1}\lambda^x e^{-\lambda t_x}]$ gives us

$$P(\text{alive at } T\,|\,X = x, t_x, T, \lambda, p) = \frac{(1-p)^x\lambda^x e^{-\lambda T}}{L(\lambda, p\,|\,X = x, t_x, T)}\,, \quad (17)$$

where the expression for $L(\lambda, p\,|\,X = x, t_x, T)$ is given in (3). (Note that when $x = 0$, the expression given in (17) equals 1.)

As the transaction rate λ and death probability p are unobserved, we compute $P(\text{alive}\,|\,X = x, t_x, T)$ for a randomly chosen customer by taking the expectation of (17) over the (joint) distribution of λ and p, updated to take account of the information $(X = x, t_x, T)$:

$$P(\text{alive at } T\,|\,X = x, t_x, T, r, \alpha, a, b)$$
$$= \int_0^1\int_0^\infty \Big\{P(\text{alive at } T\,|\,X = x, t_x, T, \lambda, p)$$
$$\times f(\lambda, p\,|\,r, \alpha, a, b, X = x, t_x, T)\Big\}d\lambda\, dp\,. \quad (18)$$

By Bayes theorem, the joint posterior distribution of λ and p is given by

$$f(\lambda, p \mid r, \alpha, a, b, X = x, t_x, T) = \frac{L(\lambda, p \mid X = x, t_x, T) f(\lambda \mid r, \alpha) f(p \mid a, b)}{L(r, \alpha, a, b \mid X = x, t_x, T)} \,. \tag{19}$$

Substituting (17) and (19) in (18), we get

$$\begin{aligned} P(\text{alive at } T \mid X = x, t_x, T, r, \alpha, a, b) &= \frac{1}{L(r, \alpha, a, b \mid X = x, t_x, T)} \\ &\quad \times \int_0^1 \int_0^\infty (1-p)^x \lambda^x e^{-\lambda T} f(\lambda \mid r, \alpha) f(p \mid a, b) d\lambda \, dp \,. \end{aligned} \tag{20}$$

Now,

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

which, recalling (8) and (9),

$$= \frac{B(a, b+x)}{B(a, b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+T)^{r+x}} \tag{21}$$

Substituting (11) and (21) in (20) and simplifying, we get

$$\begin{aligned} &P(\text{alive at } T \mid X = x, t_x, T, r, \alpha, a, b) \\ &\qquad = 1 \Big/ \left[1 + \delta_{x>0} \frac{a}{b+x-1} \left(\frac{\alpha+T}{\alpha+t_x} \right)^{r+x} \right] . \end{aligned} \tag{22}$$

5.2 Derivation of $E(Y(t) \mid X = x, t_x, T)$

Let the random variable $Y(t)$ denote the number of purchases made in $(T, T+t]$. We are interested in computing the conditional expectation $E(Y(t) \mid X = x, t_x, T)$, the expected number of purchases in $(T, T+t]$ for a customer with purchase history $(X = x, t_x, T)$.

If the customer is alive at T, it follows from (7) that

$$E(Y(t) \mid \lambda, p) = \frac{1}{p} - \frac{1}{p} e^{-\lambda p t} \,. \tag{23}$$

Multiplying this expression by the probability that a customer with purchase history $(X = x, t_x, T)$ is alive at T, (17), gives us

$$\begin{aligned} &E(Y(t) \mid X = x, t_x, T, \lambda, p) \\ &\qquad = \frac{(1-p)^x \lambda^x e^{-\lambda T} \left(\frac{1}{p} - \frac{1}{p} e^{-\lambda p t} \right)}{L(\lambda, p \mid X = x, t_x, T)} \\ &\qquad = \frac{p^{-1}(1-p)^x \lambda^x e^{-\lambda T} - p^{-1}(1-p)^x \lambda^x e^{-\lambda(T+pt)}}{L(\lambda, p \mid X = x, t_x, T)} \,. \end{aligned} \tag{24}$$

(Note that this reduces to (23) when $x = 0$, which follows from the assumption that a customer who made zero purchases in $(0, T]$ is alive at time T.)

As the transaction rate λ and death probability p are unobserved, we compute $E(Y(t) \mid X = x, t_x, T)$ for a randomly chosen customer by taking the expectation of (24) over the posterior distribution of λ and p:

$$
\begin{aligned}
&E(Y(t) \mid X = x, t_x, T, r, \alpha, a, b) \\
&\quad = \int_0^1 \int_0^\infty E(Y(t) \mid X = x, t_x, T, \lambda, p) f(\lambda, p \mid r, \alpha, a, b, X = x, t_x, T) d\lambda\, dp\,. \qquad (25)
\end{aligned}
$$

Substituting (19) and (24) in (25), we get

$$
E(Y(t) \mid X = x, t_x, T, r, \alpha, a, b) = \frac{\mathsf{A} - \mathsf{B}}{L(r, \alpha, a, b \mid X = x, t_x, T)}\,, \qquad (26)
$$

where

$$
\begin{aligned}
\mathsf{A} &= \int_0^1 \int_0^\infty p^{-1}(1-p)^x \lambda^x e^{-\lambda T} f(\lambda \mid r, \alpha) f(p \mid a, b) d\lambda dp\,, \text{and} \\
\mathsf{B} &= \int_0^1 \int_0^\infty p^{-1}(1-p)^x \lambda^x e^{-\lambda(T+pt)} f(\lambda \mid r, \alpha) f(p \mid a, b) d\lambda dp\,.
\end{aligned}
$$

- Recalling (8) and (9),

$$
\mathsf{A} = \frac{B(a-1, b+x)}{B(a,b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+T)^{r+x}}\,. \qquad (27)
$$

- Now,

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

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

$$
\begin{aligned}
&= \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)B(a,b)} \int_0^1 q^{b+x-1}(1-q)^{a-2}(\alpha+T+t-qt)^{-(r+x)} dq \\
&= \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)B(a,b)(\alpha+T+t)^{r+x}} \\
&\qquad \times \int_0^1 q^{b+x-1}(1-q)^{a-2}\left(1 - \tfrac{t}{\alpha+T+t} q\right)^{-(r+x)} dq
\end{aligned}
$$

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

$$
\begin{aligned}
&= \frac{B(a-1, b+x)}{B(a,b)} \frac{\Gamma(r+x)\alpha^r}{\Gamma(r)(\alpha+T+t)^{r+x}} \\
&\qquad \times {}_2F_1\left(r+x, b+x; a+b+x-1; \tfrac{t}{\alpha+T+t}\right). \qquad (28)
\end{aligned}
$$

Substituting (11), (27) and (28) in (26) and simplifying, we get

$$E(Y(t)\,|\,X=x,t_x,T,r,\alpha,a,b)=$$
$$\frac{\dfrac{a+b+x-1}{a-1}\left[1-\left(\dfrac{\alpha+T}{\alpha+T+t}\right)^{r+x}{}_2F_1\big(r+x,b+x;a+b+x-1;\tfrac{t}{\alpha+T+t}\big)\right]}{1+\delta_{x>0}\,\dfrac{a}{b+x-1}\left(\dfrac{\alpha+T}{\alpha+t_x}\right)^{r+x}}\,. \tag{29}$$

5.3 Derivation of $P(Y(t)=y\,|\,X=x,t_x,T)$

Finally, we are interested in deriving an expression for the probability of observing y purchases in the interval $(T,T+t]$ by a customer with purchase history $(X=x,t_x,T)$.

If the customer is alive at T, it follows from (4) that

$$P(Y(t)=y\,|\,\lambda,p)=(1-p)^y\frac{(\lambda t)^y e^{-\lambda t}}{y!} + \delta_{y>0}\,p(1-p)^{y-1}\left[1-e^{-\lambda t}\sum_{j=0}^{y-1}\frac{(\lambda t)^j}{j!}\right]. \tag{30}$$

On the other hand, if the customer is dead at T,

$$P(Y(t)=y\,|\,\lambda,p)=\begin{cases}1 & \text{if } y=0\\ 0 & \text{otherwise}\end{cases} \tag{31}$$

Multiplying (30) by the probability that a customer with purchase history $(X=x,t_x,T)$ is alive at T, (17), and (31) by the probability that this same customer is dead at T (which equals 0 if $x=0$), gives us

$$P(Y(t)=y\,|\,X=x,t_x,T,\lambda,p)=\Bigg\{\delta_{x>0,y=0}\,p(1-p)^{x-1}\lambda^x e^{-\lambda t_x} + (1-p)^{x+y}\frac{\lambda^{x+y}t^y e^{-\lambda(T+t)}}{y!}+\delta_{y>0}\,p(1-p)^{x+y-1} \times\left[\lambda^x e^{-\lambda T}-e^{-\lambda(T+t)}\sum_{j=0}^{y-1}\frac{\lambda^{x+j}t^j}{j!}\right]\Bigg\}\Big/L(\lambda,p\,|\,X=x,t_x,T)\,. \tag{32}$$

(Note that this reduces to (30) when $x=0$, which follows from the assumption that a customer who made zero purchases in $(0,T]$ is alive at time T.)

As the transaction rate λ and death probability p are unobserved, we compute $P(Y(t)=y\,|\,X=x,t_x,T)$ for a randomly chosen customer by taking the expectation of (32) over the posterior distribution of λ and p:

$$P(Y(t)=y\,|\,X=x,t_x,T,r,\alpha,a,b) =\int_0^1\int_0^\infty\Big\{P(Y(t)=y\,|\,X=x,t_x,T,\lambda,p) \times f(\lambda,p\,|\,r,\alpha,a,b,X=x,t_x,T)\Big\}\,d\lambda dp\,. \tag{33}$$

Substituting (19) and (32) in (33), we get

$$P(Y(t)=y\,|\,X=x,t_x,T,r,\alpha,a,b)=\frac{\delta_{x>0,y=0}\,\mathsf{A}+\mathsf{B}+\delta_{y>0}\,\mathsf{C}}{L(r,\alpha,a,b\,|\,X=x,t_x,T)} \tag{34}$$

where, recalling (1), (2), (8), and (9),

$$\begin{aligned}
\mathsf{A} &= \int_0^1\int_0^\infty p(1-p)^{x-1}\lambda^x e^{-\lambda t_x} f(\lambda\,|\,r,\alpha)f(p\,|\,a,b)d\lambda\,dp\\
&= \biggl\{\int_0^1 \frac{p^a(1-p)^{b+x-2}}{B(a,b)}\,dp\biggr\}\biggl\{\int_0^\infty \frac{\alpha^r\lambda^{r+x-1}e^{-\lambda(\alpha+t_x)}}{\Gamma(r)}\,d\lambda\biggr\}\\
&= \frac{B(a+1,b+x-1)}{B(a,b)}\frac{\Gamma(r+x)}{\Gamma(r)}\frac{\alpha^r}{(\alpha+t_x)^{r+x}}\,,
\end{aligned} \tag{35}$$

and

$$\begin{aligned}
\mathsf{B} &= \int_0^1\int_0^\infty (1-p)^{x+y}\frac{\lambda^{x+y}t^y e^{-\lambda(T+t)}}{y!} f(\lambda\,|\,r,\alpha)f(p\,|\,a,b)d\lambda\,dp\\
&= \biggl\{\int_0^1 \frac{p^{a-1}(1-p)^{b+x+y-1}}{B(a,b)}\,dp\biggr\}\biggl\{\int_0^\infty \frac{\alpha^r\lambda^{r+x+y-1}t^y e^{-\lambda(\alpha+T+t)}}{y!\Gamma(r)}\,d\lambda\biggr\}\\
&= \frac{B(a,b+x+y)}{B(a,b)}\frac{\Gamma(r+x+y)}{\Gamma(r)y!}\frac{\alpha^r t^y}{(\alpha+T+t)^{r+x+y}}\,,
\end{aligned} \tag{36}$$

and

$$\begin{aligned}
\mathsf{C} &= \int_0^1\int_0^\infty \biggl\{p(1-p)^{x+y-1}\biggl[\lambda^x e^{-\lambda T}-e^{-\lambda(T+t)}\sum_{j=0}^{y-1}\frac{\lambda^{x+j}t^j}{j!}\biggr]\\
&\qquad\qquad \times f(\lambda\,|\,r,\alpha)f(p\,|\,a,b)\biggr\}d\lambda dp\\
&= \biggl\{\int_0^1 \frac{p^a(1-p)^{b+x+y-2}}{B(a,b)}\,dp\biggr\}\biggl\{\int_0^\infty \frac{\alpha^r\lambda^{r+x-1}e^{-\lambda(\alpha+T)}}{\Gamma(r)}\,d\lambda\\
&\qquad\qquad -\int_0^\infty\sum_{j=0}^{y-1}\frac{\alpha^r\lambda^{r+x+j-1}t^j e^{-\lambda(\alpha+T+t)}}{j!\Gamma(r)}\,d\lambda\biggr\}\\
&= \frac{B(a+1,b+x+y-1)}{B(a,b)}\\
&\quad\times\biggl\{\frac{\Gamma(r+x)}{\Gamma(r)}\frac{\alpha^r}{(\alpha+T)^{r+x}}-\sum_{j=0}^{y-1}\frac{\Gamma(r+x+j)}{\Gamma(r)j!}\frac{\alpha^r t^j}{(\alpha+T+t)^{r+x+j}}\biggr\}\,,
\end{aligned} \tag{37}$$

and the expression for $L(r,\alpha,a,b\,|\,X=x,t_x,T)$ is given in (11).

References

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.