A Note on Deriving the Pareto/NBD Model and Related Expressions

Peter S. Fader
www.petefader.com

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

November 2005

1 Introduction

The Pareto/NBD model was developed by Schmittlein et al. (1987), hereafter SMC, to describe repeat-buying behavior in a noncontractual setting. They derive expressions for, amongst other things, (i) the probability that a customer with a given transaction history is still "alive", and (ii) the expected number of future transactions for a randomly-chosen customer, conditional on his transaction history.

Many readers of SMC find the derivations presented in the paper to be rather daunting. The objective of this note is to guide the reader through the derivations of the key results and to present some new related results. In many cases, our approach to deriving a specific expression differs from that used by SMC; our reason for taking an alternative derivation route is that we feel it is simpler to follow.

In Section 2 we outline the assumptions of the Pareto/NBD model and derive a key intermediate result. In Sections 3–6, we derive expressions for the model likelihood function (something not presented in SMC), the mean and variance, the probability that a customer is "alive", and the conditional expectation.

But before we start, let us "introduce" the Gaussian hypergeometric function,[1] which is the power series of the form

$$
{}_2F_1(a,b;c;z) = \sum_{j=0}^{\infty} \frac{(a)_j (b)_j}{(c)_j} \frac{z^j}{j!}\,, \quad c \neq 0, -1, -2, \ldots,
$$

where $(a)_j$ is Pochhammer's symbol, which denotes the ascending factorial $a(a+1)\cdots(a+j-1)$. The series converges for $|z| < 1$ and is divergent for $|z| > 1$; if $|z| = 1$, the series converges for $c - a - b > 0$.

- Since an ascending factorial can be represented as a ratio of two gamma functions,

$$
(a)_j = \frac{\Gamma(a+j)}{\Gamma(a)}\,,
$$

[1] The standard reference is the *Handbook of Mathematical Functions*, edited by Abramowitz and Stegun (1972); the interested reader is directed to this book for further information on this function. Additional information can be found in Gradshteyn and Ryzhik (1994) and Andrews, Askey, and Roy (1999). An excellent online reference is the Wolfram functions site (http://functions.wolfram.com/).

we can write the Gaussian hypergeometric function as

$$ {}_2F_1(a,b;c;z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)} \sum_{j=0}^{\infty} \frac{\Gamma(a+j)\Gamma(b+j)}{\Gamma(c+j)} \frac{z^j}{j!} \,. $$

Looking at this, it should be clear that the function is symmetric in the upper parameters a and b, i.e., ${}_2F_1(a,b;c;z) = {}_2F_1(b,a;c;z)$. (The reader should keep this in mind when working through the derivations in this note.)

- Euler's integral representation of the Gaussian hypergeometric function is

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

where $B(\cdot,\cdot)$ is the beta function. While the symmetry of the Gaussian hypergeometric function in the parameters a and b is not obvious in this integral, be assured that it does hold.

2 Model Assumptions

The Pareto/NBD model is based on the following assumptions:

i. Customers go through two stages in their "lifetime" with a specific firm: they are "alive" for some period of time, then become permanently inactive.

ii. While alive, the number of transactions made by a customer follows a Poisson process with transaction rate λ; therefore the probability of observing x transactions in the time interval $(0,t]$ is given by

$$ P(X(t) = x \mid \lambda) = \frac{(\lambda t)^x e^{-\lambda t}}{x!} \,, \quad x = 0,1,2,\ldots . $$

This is equivalent to assuming that the time between transactions is distributed exponential with transaction rate λ,

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

where t_j is the time of the jth purchase.

iii. A customer's unobserved "lifetime" of length τ (after which he is viewed as being inactive) is exponentially distributed with dropout rate μ:

$$ f(\tau \mid \mu) = \mu e^{-\mu\tau} \,. $$

iv. Heterogeneity in transaction rates across customers follows a gamma distribution with shape parameter r and scale parameter α:

$$ g(\lambda \mid r, \alpha) = \frac{\alpha^r \lambda^{r-1} e^{-\lambda\alpha}}{\Gamma(r)} \,. \tag{2} $$

v. Heterogeneity in dropout rates across customers follows a gamma distribution with shape parameter s and scale parameter β.

$$ g(\mu \mid s, \beta) = \frac{\beta^s \mu^{s-1} e^{-\mu\beta}}{\Gamma(s)} \,. \tag{3} $$

vi. The transaction rate λ and the dropout rate μ vary independently across customers.

Assumptions (ii) and (iv) give us the NBD model for the distribution of the number of transactions while the customer is alive,

$$\begin{aligned} P(X(t)=x \mid r,\alpha) &= \int_0^\infty P(X(t)=x \mid \lambda)\, g(\lambda \mid r,\alpha)\, d\lambda \\ &= \frac{\Gamma(r+x)}{\Gamma(r)\, x!}\left(\frac{\alpha}{\alpha+t}\right)^r \left(\frac{t}{\alpha+t}\right)^x , \end{aligned} \tag{4}$$

while assumptions (iii) and (v) give us "Pareto distribution of the second kind",

$$\begin{aligned} f(\tau \mid s,\beta) &= \int_0^\infty f(\tau \mid \mu)\, g(\mu \mid s,\beta)\, d\mu \\ &= \frac{s}{\beta}\left(\frac{\beta}{\beta+\tau}\right)^{s+1} , \text{ and} \end{aligned} \tag{5}$$

$$\begin{aligned} F(\tau \mid s,\beta) &= \int_0^\infty F(\tau \mid \mu)\, g(\mu \mid s,\beta)\, d\mu \\ &= 1-\left(\frac{\beta}{\beta+\tau}\right)^{s} . \end{aligned} \tag{6}$$

The NBD and Pareto labels for each of the sub-models naturally leads to the name of the integrated model.

2.1 A Key Intermediate Result

As we proceed with the derivations, we will need evaluate a double integral of the following form a number of times:

$$\mathsf{A} = \int_0^\infty \int_0^\infty \frac{\lambda^\gamma \mu^\delta}{\lambda+\mu} e^{-(\lambda+\mu)t} g(\lambda \mid r,\alpha) g(\mu \mid s,\beta)\, d\lambda\, d\mu\,. \tag{7}$$

Let us consider the transformation $p = \frac{\mu}{\lambda+\mu}$ and $z = \lambda+\mu$, with corresponding inverse transformation $\lambda = (1-p)z$ and $\mu = pz$. The Jacobian of this transformation is

$$J = \begin{vmatrix} \frac{\partial\lambda}{\partial p} & \frac{\partial\lambda}{\partial z} \\ \frac{\partial\mu}{\partial p} & \frac{\partial\mu}{\partial z} \end{vmatrix} = z\,.$$

It follows that

$$\mathsf{A} = \int_0^1 \int_0^\infty p^\delta (1-p)^\gamma z^{\gamma+\delta-1} e^{-zt} g(p,z \mid \alpha,\beta,r,s)\, dz\, dp\,, \tag{8}$$

where the joint distribution of p and z

$$g(p,z \mid \alpha,\beta,r,s) = \frac{\alpha^r \beta^s}{\Gamma(r)\Gamma(s)} p^{s-1}(1-p)^{r-1} z^{r+s-1} e^{-z(\alpha-(\alpha-\beta)p)} \tag{9}$$

is derived using the transformation technique (Casella and Berger 2002, Section 4.3, pp. 156–162; Mood et al. 1974, Section 6.2, p. 204ff). Substituting (9) in (8) gives us

$$\mathsf{A} = \frac{\alpha^r \beta^s}{\Gamma(r)\Gamma(s)} \times \mathsf{B}$$

where

$$
\begin{aligned}
\mathsf{B} &= \int_0^1 \int_0^\infty p^{s+\delta-1}(1-p)^{r+\gamma-1} z^{r+s+\gamma+\delta-2} e^{-z(\alpha+t-(\alpha-\beta)p)}\,dz\,dp \\
&= \int_0^1 p^{s+\delta-1}(1-p)^{r+\gamma-1}\left\{\int_0^\infty z^{r+s+\gamma+\delta-2} e^{-z(\alpha+t-(\alpha-\beta)p)}\,dz\right\} dp \\
&= \Gamma(r+s+\gamma+\delta-1)\int_0^1 p^{s+\delta-1}(1-p)^{r+\gamma-1}(\alpha+t-(\alpha-\beta)p)^{-(r+s+\gamma+\delta-1)}\,dp \\
&= \frac{\Gamma(r+s+\gamma+\delta-1)}{(\alpha+t)^{r+s+\gamma+\delta-1}}\int_0^1 p^{s+\delta-1}(1-p)^{r+\gamma-1}\big[1-\big(\tfrac{\alpha-\beta}{\alpha+t}\big)p\big]^{-(r+s+\gamma+\delta-1)}\,dp
\end{aligned}
$$

which recalling Euler's integral for the Gaussian hypergeometric function, (1),

$$
= \frac{\Gamma(r+s+\gamma+\delta-1)}{(\alpha+t)^{r+s+\gamma+\delta-1}}\frac{\Gamma(r+\gamma)\Gamma(s+\delta)}{\Gamma(r+s+\gamma+\delta)}\,{}_2F_1\big(r+s+\gamma+\delta-1, s+\delta; r+s+\gamma+\delta; \tfrac{\alpha-\beta}{\alpha+t}\big)\,,
$$

and therefore

$$
\begin{aligned}
\mathsf{A} = \frac{\alpha^r\beta^s}{(\alpha+t)^{r+s+\gamma+\delta-1}}&\frac{\Gamma(r+\gamma)\Gamma(s+\delta)}{\Gamma(r)\Gamma(s)}\left(\frac{1}{r+s+\gamma+\delta-1}\right) \\
&\times {}_2F_1\big(r+s+\gamma+\delta-1, s+\delta; r+s+\gamma+\delta; \tfrac{\alpha-\beta}{\alpha+t}\big)\,. \qquad (10)
\end{aligned}
$$

Looking closely at (10), we see that the z argument of the Gaussian hypergeometric function, $\frac{\alpha-\beta}{\alpha+t}$, is guaranteed to be bounded between 0 and 1 when $\alpha \geq \beta$ (since $\beta > 0$ and $t > 0$), thus ensuring convergence of the series representation of the function. However, when $\alpha < \beta$ we can be faced with the situation where $|z| > 1$, in which case the series is divergent.

Therefore, for the case of $\alpha \leq \beta$, let us consider the transformation $p = \frac{\lambda}{\lambda+\mu}$ and $z = \lambda+\mu$, with corresponding inverse transformation $\lambda = pz$ and $\mu = (1-p)z$. The Jacobian of this transformation is

$$
J = \begin{vmatrix} \frac{\partial\lambda}{\partial p} & \frac{\partial\lambda}{\partial z} \\ \frac{\partial\mu}{\partial p} & \frac{\partial\mu}{\partial z} \end{vmatrix} = z\,.
$$

It follows that

$$
\mathsf{A} = \int_0^1 \int_0^\infty p^\gamma(1-p)^\delta z^{\gamma+\delta-1} e^{-zt} g(p, z \,|\, \alpha, \beta, r, s)\,dz\,dp\,,
$$

where the joint distribution of p and z

$$
g(p, z \,|\, \alpha, \beta, r, s) = \frac{\alpha^r\beta^s}{\Gamma(r)\Gamma(s)} p^{r-1}(1-p)^{s-1} z^{r+s-1} e^{-z(\beta-(\beta-\alpha)p)}
$$

is derived using the transformation technique. This gives us

$$
\mathsf{A} = \frac{\alpha^r\beta^s}{\Gamma(r)\Gamma(s)} \times \mathsf{B}
$$

where

$$
\begin{aligned}
\mathsf{B} &= \int_0^1 \int_0^\infty p^{r+\gamma-1}(1-p)^{s+\delta-1} z^{r+s+\gamma+\delta-2} e^{-z(\beta+t-(\beta-\alpha)p)}\,dz\,dp \\
&= \int_0^1 p^{r+\gamma-1}(1-p)^{s+\delta-1}\left\{\int_0^\infty z^{r+s+\gamma+\delta-2} e^{-z(\beta+t-(\beta-\alpha)p)}\,dz\right\} dp \\
&= \Gamma(r+s+\gamma+\delta-1)\int_0^1 p^{r+\gamma-1}(1-p)^{s+\delta-1}(\beta+t-(\beta-\alpha)p)^{-(r+s+\gamma+\delta-1)}\,dp \\
&= \frac{\Gamma(r+s+\gamma+\delta-1)}{(\beta+t)^{r+s+\gamma+\delta-1}}\int_0^1 p^{r+\gamma-1}(1-p)^{s+\delta-1}\big[1-\big(\tfrac{\beta-\alpha}{\beta+t}\big)p\big]^{-(r+s+\gamma+\delta-1)}\,dp
\end{aligned}
$$

which recalling Euler's integral for the Gaussian hypergeometric function, (1),

$$= \frac{\Gamma(r+s+\gamma+\delta-1)}{(\beta+t)^{r+s+\gamma+\delta-1}} \frac{\Gamma(r+\gamma)\Gamma(s+\delta)}{\Gamma(r+s+\gamma+\delta)} {}_2F_1\big(r+s+\gamma+\delta-1, r+\gamma; r+s+\gamma+\delta; \tfrac{\beta-\alpha}{\beta+t}\big)\,,$$

and therefore

$$\begin{aligned} \mathsf{A} = {} & \frac{\alpha^r \beta^s}{(\beta+t)^{r+s+\gamma+\delta-1}} \frac{\Gamma(r+\gamma)\Gamma(s+\delta)}{\Gamma(r)\Gamma(s)} \left(\frac{1}{r+s+\gamma+\delta-1}\right) \\ & \times {}_2F_1\big(r+s+\gamma+\delta-1, r+\gamma; r+s+\gamma+\delta; \tfrac{\beta-\alpha}{\beta+t}\big)\,. \end{aligned} \tag{11}$$

We see that the z argument of the above Gaussian hypergeometric function is bounded between 0 and 1 when $\alpha \leq \beta$. We therefore present (10) and (11) as solutions to (7): we use (10) when $\alpha \geq \beta$ and (11) when $\alpha \leq \beta$.

3 Deriving the Model Likelihood Function

3.1 Deriving the Likelihood Function Conditional on λ and μ

Let us assume we know when each of a customer's x transactions occurred during the period $(0, T]$; we denote these times by $t_1, t_2, \ldots, t_x$:

There are two possible ways this pattern of transactions could arise:

i. The customer is still alive at the end of the observation period (i.e., $\tau > T$), in which case the individual-level likelihood function is simply the product of the (inter-transaction-time) exponential density functions and the associated survivor function:

$$\begin{aligned} L(\lambda \,|\, t_1, \ldots, t_x, T, \tau > T) &= \lambda e^{-\lambda t_1} \lambda e^{-\lambda(t_2 - t_1)} \cdots \lambda e^{-\lambda(t_x - t_{x-1})} e^{-\lambda(T - t_x)} \\ &= \lambda^x e^{-\lambda T}\,. \end{aligned}$$

ii. The customer became inactive at some time τ in the interval $(t_x, T]$, in which case the individual-level likelihood function is

$$\begin{aligned} L(\lambda \,|\, t_1, \ldots, t_x, T, \text{inactive at } \tau \in (t_x, T]) &= \lambda e^{-\lambda t_1} \lambda e^{-\lambda(t_2 - t_1)} \cdots \lambda e^{-\lambda(t_x - t_{x-1})} e^{-\lambda(\tau - t_x)} \\ &= \lambda^x e^{-\lambda \tau}\,. \end{aligned}$$

Note that in both cases, information on when each of the x transactions occurred is not required; we can replace $t_1, \ldots, t_x, T$ with (x, t_x, T) where, by definition, $t_x = 0$ when $x = 0$. In other words, t_x and x are sufficient summaries of a customer's transaction history. (Using direct marketing terminology, t_x is recency and x is frequency.)

Removing the conditioning on τ yields the following expression for the individual-level likelihood function:

$$\begin{aligned} L(\lambda, \mu \,|\, x, t_x, T) = {} & L(\lambda \,|\, x, T, \tau > T) P(\tau > T \,|\, \mu) \\ & + \int_{t_x}^{T} L(\lambda \,|\, x, T, \text{inactive at } \tau \in (t_x, T]) f(\tau \,|\, \mu)\, d\tau \qquad (12) \\ = {} & \lambda^x e^{-\lambda T} e^{-\mu T} + \lambda^x \int_{t_x}^{T} e^{-\lambda \tau} \mu e^{-\mu \tau}\, d\tau \\ = {} & \lambda^x e^{-(\lambda+\mu)T} + \frac{\lambda^x \mu}{\lambda+\mu} e^{-(\lambda+\mu)t_x} - \frac{\lambda^x \mu}{\lambda+\mu} e^{-(\lambda+\mu)T} \qquad (13) \\ = {} & \frac{\lambda^x \mu}{\lambda+\mu} e^{-(\lambda+\mu)t_x} + \frac{\lambda^{x+1}}{\lambda+\mu} e^{-(\lambda+\mu)T}\,. \qquad (14) \end{aligned}$$

This is a new result, as SMC do not present an explicit expression for the model likelihood function.

3.2 Removing the Conditioning on λ and μ

We remove the conditioning on λ and μ by taking the expectation of $L(\lambda, \mu \,|\, x, t_x, T)$ over the distributions of λ and μ:

$$L(r, \alpha, s, \beta \,|\, x, t_x, T) = \int_0^\infty \int_0^\infty L(\lambda, \mu \,|\, x, t_x, T) g(\lambda \,|\, r, \alpha) g(\mu \,|\, s, \beta) d\lambda d\mu \,. \tag{15}$$

Substituting (13) in (15) gives us

$$L(r, \alpha, s, \beta \,|\, x, t_x, T) = \mathsf{A}_1 + \mathsf{A}_2 - \mathsf{A}_3 \tag{16}$$

where

$$\mathsf{A}_1 = \int_0^\infty \int_0^\infty \lambda^x e^{-(\lambda+\mu)T} g(\lambda \,|\, r, \alpha) g(\mu \,|\, s, \beta) d\lambda d\mu \tag{17}$$

and

$$\mathsf{A}_2 = \int_0^\infty \int_0^\infty \frac{\lambda^x \mu e^{-(\lambda+\mu)t_x}}{\lambda + \mu} g(\lambda \,|\, r, \alpha) g(\mu \,|\, s, \beta) d\lambda d\mu$$

and

$$\mathsf{A}_3 = \int_0^\infty \int_0^\infty \frac{\lambda^x \mu e^{-(\lambda+\mu)T}}{\lambda + \mu} g(\lambda \,|\, r, \alpha) g(\mu \,|\, s, \beta) d\lambda d\mu$$

Let us first consider A_1: substituting (2) and (3) in (17), we have

$$\begin{aligned} \mathsf{A}_1 &= \int_0^\infty \int_0^\infty \lambda^x e^{-(\lambda+\mu)T} \frac{\alpha^r \lambda^{r-1} e^{-\lambda\alpha}}{\Gamma(r)} \frac{\beta^s \mu^{s-1} e^{-\mu\beta}}{\Gamma(s)} d\lambda d\mu \\ &= \frac{\Gamma(r+x)\alpha^r \beta^s}{\Gamma(r)(\alpha+T)^{r+x}(\beta+T)^s} \end{aligned}$$

Looking closely at the expressions for A_2 and A_3, we see that they have the same form as that given in (7) with $\gamma = x$ and $\delta = 1$, and $t = t_x$ and $t = T$ respectively. Recalling the solutions given in (10) and (11), it follows that for $\alpha \geq \beta$,

$$\begin{aligned} \mathsf{A}_2 &= \frac{\alpha^r \beta^s}{(\alpha+t_x)^{r+s+x}} \frac{\Gamma(r+\gamma)}{\Gamma(r)} \left(\frac{s}{r+s+x}\right) {}_2F_1\big(r+s+x, s+1; r+s+x+1; \tfrac{\alpha-\beta}{\alpha+t_x}\big) \\ \mathsf{A}_3 &= \frac{\alpha^r \beta^s}{(\alpha+T)^{r+s+x}} \frac{\Gamma(r+x)}{\Gamma(r)} \left(\frac{s}{r+s+x}\right) {}_2F_1\big(r+s+x, s+1; r+s+x+1; \tfrac{\alpha-\beta}{\alpha+T}\big) \,, \end{aligned}$$

while for $\alpha \leq \beta$,

$$\begin{aligned} \mathsf{A}_2 &= \frac{\alpha^r \beta^s}{(\beta+t_x)^{r+s+x}} \frac{\Gamma(r+x)}{\Gamma(r)} \left(\frac{s}{r+s+x}\right) {}_2F_1\big(r+s+x, r+x; r+s+x+1; \tfrac{\beta-\alpha}{\beta+t_x}\big) \\ \mathsf{A}_3 &= \frac{\alpha^r \beta^s}{(\beta+T)^{r+s+x}} \frac{\Gamma(r+x)}{\Gamma(r)} \left(\frac{s}{r+s+x}\right) {}_2F_1\big(r+s+x, r+x; r+s+x+1; \tfrac{\beta-\alpha}{\beta+T}\big) \,. \end{aligned}$$

Substituting these expressions for A_1–A_3 into (16) and simplifying gives us the following expression for the likelihood function for a randomly-chosen individual with purchase history (x, t_x, T):

$$L(r, \alpha, s, \beta \,|\, x, t_x, T) = \frac{\Gamma(r+x)\alpha^r \beta^s}{\Gamma(r)} \left\{ \frac{1}{(\alpha+T)^{r+x}(\beta+T)^s} + \left(\frac{s}{r+s+x}\right) \mathsf{A}_0 \right\} \tag{18}$$

where for $\alpha \geq \beta$

$$\mathsf{A}_0 = \frac{{}_2F_1\big(r+s+x, s+1; r+s+x+1; \frac{\alpha-\beta}{\alpha+t_x}\big)}{(\alpha+t_x)^{r+s+x}} - \frac{{}_2F_1\big(r+s+x, s+1; r+s+x+1; \frac{\alpha-\beta}{\alpha+T}\big)}{(\alpha+T)^{r+s+x}} \tag{19}$$

while for $\alpha \leq \beta$

$$\mathsf{A}_0 = \frac{{}_2F_1\big(r+s+x, r+x; r+s+x+1; \frac{\beta-\alpha}{\beta+t_x}\big)}{(\beta+t_x)^{r+s+x}} - \frac{{}_2F_1\big(r+s+x, r+x; r+s+x+1; \frac{\beta-\alpha}{\beta+T}\big)}{(\beta+T)^{r+s+x}}\,. \tag{20}$$

(This expression for the model likelihood function is that used in Fader et al. (2005b).)

The four Pareto/NBD model parameters (r, α, s, β) can be estimated via the method of maximum likelihood in the following manner. Suppose we have a sample of N customers, where customer i had x_i transactions in the period $(0, T_i]$, with the last transaction occurring at t_{x_i}. The sample log-likelihood function is given by

$$LL(r, \alpha, s, \beta) = \sum_{i=1}^{N} \ln\big[L(r, \alpha, s, \beta \,|\, x_i, t_{x_i}, T_i)\big]\,.$$

This can be maximized using standard numerical optimization routines. (See Fader et al. (2005a) for details of a MATLAB-based implementation.)

A variant on the above derivation follows by changing the order of integration: we first integrate (12) over the distributions of λ and μ and then remove the conditioning on τ. Any reader who has followed our workings so far will realize that this gives us

$$\begin{aligned}
L(r, \alpha, s, \beta \,|\, x, t_x, T) &= L(r, \alpha \,|\, x, T)\, P(\tau > T \,|\, s, \beta) + \int_{t_x}^{T} L(r, \alpha \,|\, x, \tau)\, f(\tau \,|\, s, \beta)\, d\tau \\
&= \frac{\Gamma(r+x)}{\Gamma(r)} \left(\frac{\alpha}{\alpha+T}\right)^{r} \left(\frac{1}{\alpha+T}\right)^{x} \left(\frac{\beta}{\beta+T}\right)^{s} \\
&\quad + \int_{t_x}^{T} \frac{\Gamma(r+x)}{\Gamma(r)} \left(\frac{\alpha}{\alpha+\tau}\right)^{r} \left(\frac{1}{\alpha+\tau}\right)^{x} \frac{s}{\beta} \left(\frac{\beta}{\beta+\tau}\right)^{s+1} d\tau \\
&= \frac{\Gamma(r+x)\alpha^r\beta^s}{\Gamma(r)} \left[\frac{1}{(\alpha+T)^{r+x}(\beta+T)^s} + s\,\mathsf{C}\right]
\end{aligned} \tag{21}$$

where

$$\mathsf{C} = \int_{t_x}^{T} (\alpha+\tau)^{-(r+x)}(\beta+\tau)^{-(s+1)}\, d\tau\,.$$

i. For $\alpha \geq \beta$, we make the change of variable $y = \alpha + \tau$, giving us

$$\begin{aligned}
\mathsf{C} &= \int_{\alpha+t_x}^{\alpha+T} y^{-(r+x)}(\beta-\alpha+y)^{-(s+1)} dy \\
&= \int_{\alpha+t_x}^{\infty} y^{-(r+x)}(\beta-\alpha+y)^{-(s+1)} dy - \int_{\alpha+T}^{\infty} y^{-(r+x)}(\beta-\alpha+y)^{-(s+1)} dy
\end{aligned}$$

letting $z = (\alpha + t_x)/y$ in the first integral (which implies $dy = -dz(\alpha + t_x)z^{-2}$) and $z = (\alpha + T)/y$ in the second integral (which implies $dy = -dz(\alpha + T)z^{-2}$),

$$
\begin{aligned}
&= -\int_1^0 \left(\frac{\alpha + t_x}{z}\right)^{-(r+x)} \left(\frac{\alpha + t_x}{z^2}\right) \left[\beta - \alpha + \frac{\alpha + t_x}{z}\right]^{-(s+1)} dz \\
&\qquad + \int_1^0 \left(\frac{\alpha + T}{z}\right)^{-(r+x)} \left(\frac{\alpha + T}{z^2}\right) \left[\beta - \alpha + \frac{\alpha + T}{z}\right]^{-(s+1)} dz \\
&= (\alpha + t_x)^{-(r+s+x)} \int_0^1 z^{r+s+x-1} \big[1 - \big(\tfrac{\alpha-\beta}{\alpha+t_x}\big) z\big]^{-(s+1)} dz \\
&\qquad - (\alpha + T)^{-(r+s+x)} \int_0^1 z^{r+s+x-1} \big[1 - \big(\tfrac{\alpha-\beta}{\alpha+T}\big) z\big]^{-(s+1)} dz
\end{aligned}
$$

which recalling Euler's integral for the Gaussian hypergeometric function, (1),

$$
\begin{aligned}
= \frac{1}{r+s+x} \Bigg\{ & \frac{{}_2F_1\big(s+1, r+s+x; r+s+x+1; \tfrac{\alpha-\beta}{\alpha+t_x}\big)}{(\alpha + t_x)^{r+s+x}} \\
& - \frac{{}_2F_1\big(s+1, r+s+x; r+s+x+1; \tfrac{\alpha-\beta}{\alpha+T}\big)}{(\alpha + T)^{r+s+x}} \Bigg\}.
\end{aligned}
$$

Substituting this expression for C in (21), and recalling the symmetry of Gaussian hypergeometric function in its upper parameters (i.e., ${}_2F_1(a, b; c; z) = {}_2F_1(b, a; c; z)$), yields (18) and (19).

ii. For $\alpha \leq \beta$, we make the change of variable $y = \beta + \tau$, giving us

$$
\begin{aligned}
\mathsf{C} &= \int_{\beta+t_x}^{\beta+T} y^{-(s+1)} (\alpha - \beta + y)^{-(r+x)} dy \\
&= \int_{\alpha+t_x}^{\infty} y^{-(s+1)} (\alpha - \beta + y)^{-(r+x)} dy - \int_{\alpha+T}^{\infty} y^{-(s+1)} (\alpha - \beta + y)^{-(r+x)} dy
\end{aligned}
$$

letting $z = (\beta + t_x)/y$ in the first integral (which implies $dy = -dz(\beta + t_x)z^{-2}$) and $z = (\beta + T)/y$ in the second integral (which implies $dy = -dz(\beta + T)z^{-2}$),

$$
\begin{aligned}
&= -\int_1^0 \left(\frac{\beta + t_x}{z}\right)^{-(s+1)} \left(\frac{\beta + t_x}{z^2}\right) \left[\alpha - \beta + \frac{\beta + t_x}{z}\right]^{-(r+x)} dz \\
&\qquad + \int_1^0 \left(\frac{\beta + T}{z}\right)^{-(s+1)} \left(\frac{\beta + T}{z^2}\right) \left[\alpha - \beta + \frac{\beta + T}{z}\right]^{-(r+x)} dz \\
&= (\beta + t_x)^{-(r+s+x)} \int_0^1 z^{r+s+x-1} \big[1 - \big(\tfrac{\beta-\alpha}{\beta+t_x}\big) z\big]^{-(r+x)} dz \\
&\qquad - (\beta + T)^{-(r+s+x)} \int_0^1 z^{r+s+x-1} \big[1 - \big(\tfrac{\beta-\alpha}{\beta+T}\big) z\big]^{-(r+x)} dz \\
&= \frac{1}{r+s+x} \Bigg[\frac{1}{(\beta + t_x)^{r+s+x}} {}_2F_1\big(r+x, r+s+x; r+s+x+1; \tfrac{\beta-\alpha}{\beta+t_x}\big) \\
&\qquad - \frac{1}{(\beta + T)^{r+s+x}} {}_2F_1\big(r+x, r+s+x; r+s+x+1; \tfrac{\beta-\alpha}{\beta+T}\big) \Bigg].
\end{aligned}
$$

Substituting this expression for C in (21), and recalling the symmetry of Gaussian hypergeometric function in its upper parameters, yields (18) and (20).

3.3 An Alternative Expression for the Model Likelihood Function

An alternative expression for the model likelihood function can be obtained by substituting (14), instead of (13), in (15), giving us

$$L(r, \alpha, s, \beta \,|\, x, t_x, T) = \mathsf{A}_2 + \mathsf{A}_4$$

where A_2 is defined as above and

$$\mathsf{A}_4 = \int_0^\infty \int_0^\infty \frac{\lambda^{x+1} e^{-(\lambda+\mu)T}}{\lambda + \mu} g(\lambda \,|\, r, \alpha) g(\mu \,|\, s, \beta) d\lambda d\mu$$

Looking closely at this expression for A_4, we see that it has the same form as that given in (7) with $\gamma = x + 1$, $\delta = 0$, and $t = T$. Recalling the solutions given in (10) and (11), it follows that for $\alpha \geq \beta$,

$$\mathsf{A}_4 = \frac{\alpha^r \beta^s}{(\alpha + T)^{r+s+x}} \frac{\Gamma(r + x)}{\Gamma(r)} \left(\frac{r + x}{r + s + x} \right) {}_2F_1\big(r + s + x, s; r + s + x + 1; \tfrac{\alpha-\beta}{\alpha+T}\big) \,,$$

while for $\alpha \leq \beta$,

$$\mathsf{A}_4 = \frac{\alpha^r \beta^s}{(\beta + T)^{r+s+x}} \frac{\Gamma(r + x)}{\Gamma(r)} \left(\frac{r + x}{r + s + x} \right) {}_2F_1\big(r + s + x, r + x + 1; r + s + x + 1; \tfrac{\beta-\alpha}{\beta+T}\big) \,.$$

Therefore, for $\alpha \geq \beta$,

$$\begin{aligned} L(r, \alpha, s, \beta \,|\, x, t_x, T) = \frac{\Gamma(r + x)\alpha^r \beta^s}{\Gamma(r)} \Bigg\{ \left(\frac{s}{r + s + x} \right) & \frac{{}_2F_1\big(r + s + x, s + 1; r + s + x + 1; \frac{\alpha-\beta}{\alpha+t_x}\big)}{(\alpha + t_x)^{r+s+x}} \\ + \left(\frac{r + x}{r + s + x} \right) & \frac{{}_2F_1\big(r + s + x, s; r + s + x + 1; \frac{\alpha-\beta}{\alpha+T}\big)}{(\alpha + T)^{r+s+x}} \Bigg\} \,, \end{aligned} \tag{22}$$

while for $\alpha \leq \beta$,

$$\begin{aligned} L(r, \alpha, s, \beta \,|\, x, t_x, T) = \frac{\Gamma(r + x)\alpha^r \beta^s}{\Gamma(r)} \Bigg\{ \left(\frac{s}{r + s + x} \right) & \frac{{}_2F_1\big(r + s + x, r + x; r + s + x + 1; \frac{\beta-\alpha}{\beta+t_x}\big)}{(\beta + t_x)^{r+s+x}} \\ + \left(\frac{r + x}{r + s + x} \right) & \frac{{}_2F_1\big(r + s + x, r + x + 1; r + s + x + 1; \frac{\beta-\alpha}{\beta+T}\big)}{(\beta + T)^{r+s+x}} \Bigg\} \,. \end{aligned} \tag{23}$$

Are (18)–(20) equivalent to (22) and (23)? While the indirect equivalence is obvious (given the equivalence of (13) and (14)), the direct equivalence is not immediately obvious.

Direct equivalence implies that for $\alpha \geq \beta$,

$$\begin{aligned} \left(\frac{\alpha + T}{\beta + T} \right)^s - \left(\frac{s}{r + s + x} \right) {}_2F_1\big(r + s + x, s + 1; r + s + x + 1; \tfrac{\alpha-\beta}{\alpha+T}\big) & \\ = \left(\frac{r + x}{r + s + x} \right) {}_2F_1\big(r + s + x, s; r + s + x + 1; \tfrac{\beta-\alpha}{\beta+T}\big) & \end{aligned} \tag{24}$$

while for $\alpha \leq \beta$,

$$\begin{aligned} \left(\frac{\beta + T}{\alpha + T} \right)^{r+x} - \left(\frac{s}{r + s + x} \right) {}_2F_1\big(r + s + x, r + x; r + s + x + 1; \tfrac{\beta-\alpha}{\beta+T}\big) & \\ = \left(\frac{r + x}{r + s + x} \right) {}_2F_1\big(r + s + x, r + x + 1; r + s + x + 1; \tfrac{\beta-\alpha}{\beta+T}\big) & \end{aligned} \tag{25}$$

Looking at the so-called Gauss' relations for contiguous functions, we have the following result (Abramowitz and Stegun 1972, equation 15.2.24):

$$(c - b - 1){}_2F_1(a, b; c; z) + b\,{}_2F_1(a, b + 1; c; z) - (c - 1){}_2F_1(a, b; c - 1; z) = 0 \,.$$

i. For the case of $\alpha \geq \beta$, let $a = r+s+x$, $b = s$ and $c = r+s+x+1$. This gives us

$$\begin{aligned}(r+x)\,{}_2F_1\big(r+s+x, s; r+s+x+1; \tfrac{\alpha-\beta}{\alpha+T}\big) + s\,{}_2F_1\big(r+s+x, s+1; r+s+x+1; \tfrac{\alpha-\beta}{\alpha+T}\big)\\ - (r+s+x)\,{}_2F_1\big(r+s+x, s; r+s+x; \tfrac{\alpha-\beta}{\alpha+T}\big) = 0\end{aligned}$$

Noting that ${}_2F_1(a, b; b; z) = (1-z)^{-a}$,

$${}_2F_1\big(r+s+x, s; r+s+x; \tfrac{\alpha-\beta}{\alpha+T}\big) = \left(\frac{\alpha+T}{\beta+T}\right)^s$$

and therefore

$$\begin{aligned}(r+x)\,{}_2F_1\big(r+s+x, s; r+s+x+1; \tfrac{\alpha-\beta}{\alpha+T}\big) + s\,{}_2F_1\big(r+s+x, s+1; r+s+x+1; \tfrac{\alpha-\beta}{\alpha+T}\big)\\ - (r+s+x)\left(\frac{\alpha+T}{\beta+T}\right)^s = 0\,,\end{aligned}$$

which is clearly equivalent to (24).

ii. For the case of $\alpha \leq \beta$, let $a = r+s+x$, $b = r+x$ and $c = r+s+x+1$. This gives us

$$\begin{aligned}s\,{}_2F_1\big(r+s+x, r+x; r+s+x+1; \tfrac{\beta-\alpha}{\beta+T}\big) + (r+x)\,{}_2F_1\big(r+s+x, r+x+1; r+s+x+1; \tfrac{\beta-\alpha}{\beta+T}\big)\\ - (r+s+x)\,{}_2F_1\big(r+s+x, r+x; r+s+x; \tfrac{\beta-\alpha}{\beta+T}\big) = 0\end{aligned}$$

The result that ${}_2F_1(a, b; b; z) = (1-z)^{-a}$ implies that

$${}_2F_1\big(r+s+x, r+x; r+s+x; \tfrac{\beta-\alpha}{\beta+T}\big) = \left(\frac{\beta+T}{\alpha+T}\right)^{r+x}$$

and therefore

$$\begin{aligned}s\,{}_2F_1\big(r+s+x, r+x; r+s+x+1; \tfrac{\beta-\alpha}{\beta+T}\big) + (r+x)\,{}_2F_1\big(r+s+x, r+x+1; r+s+x+1; \tfrac{\beta-\alpha}{\beta+T}\big)\\ - (r+s+x)\left(\frac{\beta+T}{\alpha+T}\right)^{r+x} = 0\,,\end{aligned}$$

which is clearly equivalent to (25).

4 Mean and Variance of the Pareto/NBD Model

Given that the number of transactions follows a Poisson process while the customer is alive,

i. if τ, the (unobserved) time at which the customer becomes inactive, is greater than t, the expected number of transactions is simply λt.

ii. if $\tau \leq t$, the expected number of transactions in the interval $(0, \tau]$ is $\lambda\tau$.

Removing the conditioning on the time at which the customer becomes inactive, it follows that the expected number of transactions in the time interval $(0, t]$, conditional on λ and μ, is

$$\begin{aligned}E[X(t)\,|\,\lambda,\mu] &= \lambda t P(\tau > t\,|\,\mu) + \int_0^t \lambda\tau f(\tau\,|\,\mu)\,d\tau\\ &= \lambda t e^{-\mu t} + \lambda\int_0^t \mu\tau e^{-\mu\tau}\,d\tau\\ &= \lambda t e^{-\mu t} + \frac{\lambda}{\mu}\int_0^t \mu^2\tau e^{-\mu\tau}\,d\tau\end{aligned}$$

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

$$
\begin{aligned}
&= \lambda t e^{-\mu t} + \frac{\lambda}{\mu}\Big\{1 - e^{-\mu t} - \mu t e^{-\mu t}\Big\} \\
&= \frac{\lambda}{\mu} - \frac{\lambda}{\mu} e^{-\mu t} \,. \qquad (26)
\end{aligned}
$$

To arrive at an expression for $E[X(t)]$ for a randomly-chosen individual, we take the expectation of (26) over the distributions of λ and μ:

$$
\begin{aligned}
E[X(t) \,|\, r, \alpha, s, \beta] &= \int_0^\infty \int_0^\infty E[X(t) \,|\, \lambda, \mu] g(\lambda \,|\, r, \alpha) g(\mu \,|\, s, \beta)\, d\lambda d\mu \\
&= \frac{r\beta}{\alpha(s-1)} - \frac{r\beta^s}{\alpha(s-1)(\beta+t)^{s-1}} \\
&= \frac{r\beta}{\alpha(s-1)} \left[1 - \left(\frac{\beta}{\beta+t} \right)^{s-1} \right] , \qquad (27)
\end{aligned}
$$

which is the expression reported in SMC, equation (17).

In order to derive the variance of the Pareto/NBD, we recall the defining relationship for the variance of a random variable:

$$
\operatorname{var}(X) = E(X^2) - E(X)^2 \,. \qquad (28)
$$

Having derived an expression for $E(X)$, we now need to derive an expression for $E(X^2)$.

Given that the number of transactions follows a Poisson process while the customer is alive, it follows that $E[(X(t)^2 \,|\, \lambda] = \lambda t + (\lambda t)^2$ if $\tau > t$, and $E[(X(\tau)^2 \,|\, \lambda] = \lambda\tau + (\lambda\tau)^2$ if $\tau \leq t$. Removing the conditioning on the time at which the customer becomes inactive, we have

$$
\begin{aligned}
E[X(t)^2 \,|\, \lambda, \mu] &= \{\lambda t + (\lambda t)^2\} P(\tau > t \,|\, \mu) + \int_0^t \{\lambda\tau + (\lambda\tau)^2\} f(\tau \,|\, \mu)\, d\tau \\
&= E[X(t) \,|\, \lambda, \mu] + (\lambda t)^2 e^{-\mu t} + \lambda^2 \int_0^t \mu\tau^2 e^{-\mu\tau}\, d\tau \\
&= E[X(t) \,|\, \lambda, \mu] + (\lambda t)^2 e^{-\mu t} + \frac{2\lambda^2}{\mu^2} \int_0^t \frac{\mu^3 \tau^2 e^{-\mu\tau}}{2}\, d\tau
\end{aligned}
$$

which, noting that the integrand is an Erlang-3 pdf,

$$
\begin{aligned}
&= E[X(t) \,|\, \lambda, \mu] + (\lambda t)^2 e^{-\mu t} + \frac{2\lambda^2}{\mu^2} \left\{ 1 - e^{-\mu t} - \mu t e^{-\mu t} - \frac{(\mu t)^2 e^{-\mu t}}{2} \right\} \\
&= \lambda \left\{ \frac{1}{\mu} - \frac{1}{\mu} e^{-\mu t} \right\} + 2\lambda^2 \left\{ \frac{1}{\mu^2} - \frac{e^{-\mu t}}{\mu^2} - \frac{t e^{-\mu t}}{\mu} \right\} . \qquad (29)
\end{aligned}
$$

To arrive at an expression for $E[X(t)^2]$ for a randomly-chosen individual, we take the expectation of (29) over the distributions of λ and μ:

$$
\begin{aligned}
E[X(t)^2 \,|\, r, \alpha, s, \beta] &= \int_0^\infty \int_0^\infty E[X(t)^2 \,|\, \lambda, \mu] g(\lambda \,|\, r, \alpha) g(\mu \,|\, s, \beta)\, d\lambda d\mu \\
&= \frac{r\beta}{\alpha(s-1)} \left[1 - \left(\frac{\beta}{\beta+t} \right)^{s-1} \right] \\
&\qquad + \frac{2r(r+1)\beta}{\alpha^2(s-1)} \left[\frac{\beta}{s-2} - \frac{\beta}{s-2} \left(\frac{\beta}{\beta+t} \right)^{s-2} - t \left(\frac{\beta}{\beta+t} \right)^{s-1} \right] \qquad (30)
\end{aligned}
$$

Our expression for $\operatorname{var}[X(t) \,|\, r, \alpha, s, \beta]$ is obtained by substituting (27) and (30) in (28); this is equivalent to the expression reported in SMC, equation (19).

5 Derivation of $P(\text{alive}\,|\,x, t_x, T)$

The probability that a customer with purchase history (x, t_x, T) is "alive" at time T is the probability that the (unobserved) time at which he becomes inactive (τ) occurs after T, $P(\tau > T)$. Referring back to our derivation of the individual-level likelihood function (i.e., (12)), the application of Bayes' theorem gives us

$$\begin{aligned} P(\tau > T\,|\,\lambda, \mu, x, t_x, T) &= \frac{L(\lambda\,|\,x, T, \tau > T)P(\tau > T\,|\,\mu)}{L(\lambda, \mu\,|\,x, t_x, T)} \\ &= \frac{\lambda^x e^{-(\lambda+\mu)T}}{L(\lambda, \mu\,|\,x, t_x, T)}\,. \end{aligned} \tag{31}$$

Substituting (13) in (31), we have

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

which is the expression reported in SMC, equation (A10).

As the transaction rate λ and death rate μ are unobserved, we compute $P(\text{alive}\,|\,x, t_x, T)$ for a randomly-chosen individual by taking the expectation of (31) over the distribution of λ and μ, updated to take account of the information (x, t_x, T):

$$\begin{aligned} &P(\text{alive}\,|\,r, \alpha, s, \beta, x, t_x, T) \\ &\qquad = \int_0^\infty \int_0^\infty P(\tau > T\,|\,\lambda, \mu, x, t_x, T) g(\lambda, \mu\,|\,r, \alpha, s, \beta, x, t_x, T)\, d\lambda d\mu \end{aligned} \tag{32}$$

By Bayes' theorem, the joint posterior distribution of λ and μ is

$$g(\lambda, \mu\,|\,r, \alpha, s, \beta, x, t_x, T) = \frac{L(\lambda, \mu\,|\,x, t_x, T) g(\lambda\,|\,r, \alpha) g(\mu\,|\,s, \beta)}{L(r, \alpha, s, \beta\,|\,x, t_x, T)}\,. \tag{33}$$

Substituting (31) and (33) in (32), we get

$$\begin{aligned} &P(\text{alive}\,|\,r, \alpha, s, \beta, x, t_x, T) \\ &\qquad = \int_0^\infty \int_0^\infty \lambda^x e^{-(\lambda+\mu)T} g(\lambda\,|\,r, \alpha) g(\mu\,|\,s, \beta)\, d\lambda d\mu \Big/ L(r, \alpha, s, \beta\,|\,x, t_x, T) \\ &\qquad = \frac{\Gamma(r+x)\alpha^r \beta^s}{\Gamma(r)(\alpha+T)^{r+x}(\beta+T)^s} \Big/ L(r, \alpha, s, \beta\,|\,x, t_x, T)\,. \end{aligned} \tag{34}$$

Substituting (18) on (34) gives us

$$P(\text{alive}\,|\,r, \alpha, s, \beta, x, t_x, T) = \left\{1 + \left(\frac{s}{r+s+x}\right)(\alpha+T)^{r+x}(\beta+T)^s \mathsf{A}_0\right\}^{-1}\,. \tag{35}$$

where A_0 is defined in (19) and (20).

An alternative derivation follows from the derivation of the Pareto/NBD likelihood function given in (21). Applying Bayes' theorem,

$$\begin{aligned} P(\tau > T\,|\,r, \alpha, s, \beta, x, t_x, T) &= \frac{L(r, \alpha\,|\,x, T)\, P(\tau > T\,|\,s, \beta)}{L(r, \alpha, s, \beta\,|\,x, t_x, T)} \\ &= \frac{\Gamma(r+x)}{\Gamma(r)}\left(\frac{\alpha}{\alpha+T}\right)^r \left(\frac{1}{\alpha+T}\right)^x \left(\frac{\beta}{\beta+T}\right)^s \Big/ L(r, \alpha, s, \beta\,|\,x, t_x, T)\,, \end{aligned}$$

which is the expression given in (34).

5.1 Equivalence with SMC Expressions

i. Substituting (19) in (35), it follows that for $\alpha \geq \beta$,

$$\begin{aligned} P(\text{alive} \mid r, \alpha, s, \beta, x, t_x, T) = \Bigg\{ 1 + \frac{s}{r+x+s} \\ \times \Bigg[\left(\frac{\alpha+T}{\alpha+t_x}\right)^{r+x} \left(\frac{\beta+T}{\alpha+t_x}\right)^{s} {}_2F_1\big(r+s+x, s+1; r+s+x+1; \tfrac{\alpha-\beta}{\alpha+t_x}\big) \\ - \left(\frac{\beta+T}{\alpha+T}\right)^{s} {}_2F_1\big(r+s+x, s+1; r+s+x+1; \tfrac{\alpha-\beta}{\alpha+T}\big)\Bigg]\Bigg\}^{-1} \end{aligned} \tag{36}$$

which is the expression reported in SMC, equation (11). (Note the error in SMC, equation (A25).)

ii. Substituting (20) in (35), it follows that for $\alpha \leq \beta$,

$$\begin{aligned} P(\text{alive} \mid r, \alpha, s, \beta, x, t_x, T) = \Bigg\{ 1 + \frac{s}{r+x+s} \\ \times \Bigg[\left(\frac{\alpha+T}{\beta+t_x}\right)^{r+x} \left(\frac{\beta+T}{\beta+t_x}\right)^{s} {}_2F_1\big(r+s+x, r+x; r+s+x+1; \tfrac{\beta-\alpha}{\beta+t_x}\big) \\ - \left(\frac{\alpha+T}{\beta+T}\right)^{r+x} {}_2F_1\big(r+s+x, r+x; r+s+x+1; \tfrac{\beta-\alpha}{\beta+T}\big)\Bigg]\Bigg\}^{-1} \end{aligned} \tag{37}$$

which is the expression reported in SMC, equation (12).

iii. Noting that ${}_2F_1(a, b; c; 0) = 1$ for $c > b$, (36) and (37) reduce to

$$P(\text{alive} \mid r, \alpha, s, \beta, x, t_x, T) = \left\{ 1 + \frac{s}{r+s+x} \left[\left(\frac{\alpha+T}{\alpha+t_x}\right)^{r+x+s} - 1 \right] \right\}^{-1}$$

when $\alpha = \beta$, which is the expression reported in SMC, equation (13).

6 Derivation of the Conditional Expectation

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

If the customer is active at T, it follows from our derivation of an expression for $E[X(t)]$ that

$$E[Y(t) \mid \lambda, \mu, \text{alive at } T] = \lambda t P(\tau > T+t \mid \mu, \tau > T) + \int_T^{T+t} \lambda\tau f(\tau \mid \mu, \tau > T)\, d\tau$$

which, given the memoryless property of the exponential distribution associated with τ,

$$\begin{aligned} &= \lambda t e^{-\mu t} + \lambda \int_0^t \mu\tau e^{-\mu\tau}\, d\tau \\ &= \frac{\lambda}{\mu} - \frac{\lambda}{\mu} e^{-\mu t}. \end{aligned} \tag{38}$$

Of course we don't know whether a customer is alive at T; therefore

$$E[Y(t) \mid \lambda, \mu, x, t_x, T] = E[Y(t) \mid \lambda, \mu, \text{alive at } T] P(\tau > T \mid \lambda, \mu, x, t_x, T) \tag{39}$$

As the transaction rate λ and death rate μ are unobserved, we compute $E[Y(t)\,|\,x, t_x, T]$ for a randomly-chosen individual by taking the expectation of (39) over the joint posterior distribution of λ and μ, (33):

$$E[Y(t)\,|\,r, \alpha, s, \beta, x, t_x, T] = \int_0^\infty \int_0^\infty \Big\{ E[Y(t)\,|\,\lambda, \mu, \text{alive at } T] P(\tau > T\,|\,\lambda, \mu, x, t_x, T) \\ g(\lambda, \mu\,|\,r, \alpha, s, \beta, x, t_x, T) \Big\} \, d\lambda d\mu \tag{40}$$

Substituting (31), (33), and (38) in (40), and solving the associated double integral gives us

$$E[Y(t)\,|\,r, \alpha, s, \beta, x, t_x, T] = \frac{\Gamma(r+x+1)}{\Gamma(r)(s-1)} \frac{\alpha^r \beta^s}{(\alpha+t)^{r+x+1}} \\ \times \left[\frac{1}{(\beta+T)^{s-1}} - \frac{1}{(\beta+T+t)^{s-1}} \right] \Big/ L(r, \alpha, s, \beta\,|\,x, t_x, T)\,. \tag{41}$$

Rearranging terms gives us

$$E[Y(t)\,|\,r, \alpha, s, \beta, x, t_x, T] = \left\{ \frac{\Gamma(r+x)\alpha^r \beta^s}{\Gamma(r)(\alpha+T)^{r+x}(\beta+T)^s} \Big/ L(r, \alpha, s, \beta\,|\,x, t_x, T) \right\} \\ \times \frac{(r+x)(\beta+T)}{(\alpha+T)(s-1)} \left[1 - \left(\frac{\beta+T}{\beta+T+t} \right)^{s-1} \right] .$$

The bracketed term is our expression for $P(\text{alive}\,|\,x, t_x, T)$, (34), while the rest of the expression is mean of the Pareto/NBD, (27), with "updated" parameters that reflect the individual's behavior up to time T (assuming no "death" in $(0, T]$); this is the expression reported in SMC, equation (22).

References

Abramowitz, Milton and Irene A. Stegun (eds.) (1972), *Handbook of Mathematical Functions*, New York: Dover Publications.

Andrews, George E., Richard Askey, and Ranjan Roy (1999), *Special Functions*, Cambridge: Cambridge University Press.

Casella, George, and Roger L. Berger (2002), *Statistical Inference*, 2nd edition, Pacific Grove, CA: Duxbury.

Gradshteyn, I. S. and I. M. Ryzhik (1994), *Table of Integrals, Series, and Products*, 5th edition, San Diego, CA: Academic Press.

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005a), "A Note on Implementing the Pareto/NBD Model in MATLAB." <http://brucehardie.com/notes/008/>

Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee (2005b), "RFM and CLV: Using Iso-Value Curves for Customer Base Analysis," *Journal of Marketing Research*, **42** (November), 415–430.

Mood, Alexander M., Franklin A. Graybill, and Duane C. Boes (1974), *Introduction to the Theory of Statistics*, 3rd edition, New York: McGraw-Hill Publishing Company.

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.