Deriving the Likelihood Expression for Holdout Data for the Pareto/NBD Model

Peter S. Fader
www.petefader.com

Bruce G. S. Hardie
www.brucehardie.com

Kinshuk Jerath*

February 2007

Schmittlein et al. (1987) and Fader and Hardie (2005) derive expressions for the likelihood (denoted by $L(\lambda, \mu | x, t_x, T)$) that a specific customer has an observed transaction history, as implied by the assumptions of the Pareto/NBD model. Further, they derive the corresponding expression for the likelihood for a randomly chosen customer (denoted by $L(r, \alpha, s, \beta | x, t_x, T)$). This expression can be used to recover the values of the underlying parameters for a cohort of customers, using calibration data. In this note, we derive the expression for the likelihood of observing customer transaction activity in the holdout period, for a randomly chosen customer, given his calibration period transaction history.

The time line for a customer is shown in figure 1. We start tracking the customer at time 0 (when he is assumed to be alive). In the calibration period (that extends till time T), the customer makes x purchases at times $t_1, t_2, \ldots, t_x$ (if $x = 0$ we assume $t_x = 0$). In the holdout period (that extends till time T^h), the customer makes x^h purchases at times $t_1^h, t_2^h, \ldots, t_{x^h}^h$ (if $x^h = 0$ we assume $t_{x^h}^h = T$). We denote the likelihood for the holdout data for a randomly chosen customer by $L^h(r, \alpha, s, \beta | x, t_x, T, x^h, t_x^h, T^h)$.

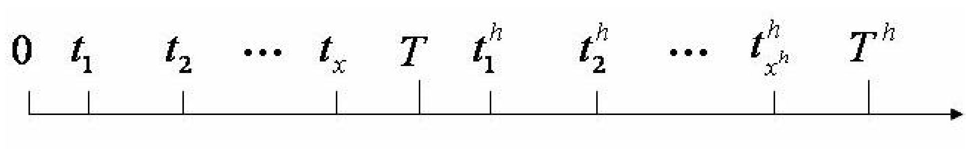

Figure 1: Time line for a customer

We carry forth the following notation from Fader and Hardie (2005). The detailed derivations and expressions for these quantities can be found in that note.

- $L(\lambda, \mu | x, t_x, T)$ and $L(r, \alpha, s, \beta | x, t_x, T)$ denotes the likelihood expressions for a specific and randomly chosen customer respectively for the calibration data.
- $\Pr(\text{alive at } T | x, t_x, T) = \frac{\lambda^x e^{-(\lambda+\mu)T}}{L(\lambda, \mu | x, t_x, T)}$ denotes the expression for the probability that the customer is alive at T (end of calibration period) given his history.

- $g(\lambda|r,\alpha)$ and $g(\mu|s,\beta)$ denotes the prior distributions of λ and μ respectively.
- $g(\lambda,\mu|r,\alpha,s,\beta,x,t_x,T) = \frac{L(\lambda,\mu|x,t_x,T)g(\lambda|r,\alpha)g(\mu|s,\beta)}{L(r,\alpha,s,\beta|x,t_x,T)}$ denotes the joint posterior distribution of λ and μ.

Let $\mathcal{L}(\lambda,\mu|x^h,t^h_{x^h},T^h,T)$ denote the likelihood of observing the holdout data given the customer is alive at T. Then, closely following the derivation for the calibration likelihood, we have:

$$\begin{aligned}\mathcal{L}(\lambda,\mu|x^h,t^h_{x^h},T^h,T) &= \lambda^{x^h}e^{-\lambda(T^h-T)}e^{-\mu(T^h-T)} + \lambda^{x^h}\int_{t^h_{x^h}}^{T^h} e^{-\lambda(\tau-T)}\mu e^{-\mu(\tau-T)}d\tau \\ &= \frac{\lambda^{x^h}\mu}{\lambda+\mu}e^{-(\lambda+\mu)(t^h_{x^h}-T)} + \frac{\lambda^{x^h+1}}{\lambda+\mu}e^{-(\lambda+\mu)(T^h-T)}\end{aligned}$$

Using the above, the holdout likelihood for a specific customer is:

$$\begin{aligned}L^h(\lambda,\mu|x,t_x,T,x^h,t^h_x,T^h) &= \delta_{x^h>0}\,\mathrm{Pr}(\text{alive at } T|x,t_x,T)\mathcal{L}(\lambda,\mu|x^h,t^h_{x^h},T^h,T) \\ &+ \delta_{x^h=0}\left[\mathrm{Pr}(\text{alive at } T|x,t_x,T)\mathcal{L}(\lambda,\mu|x^h,t^h_{x^h},T^h,T) + \mathrm{Pr}(\text{dead at } T|x,t_x,T)\right] \\ &= \mathrm{Pr}(\text{alive at } T|x,t_x,T)\mathcal{L}(\lambda,\mu|x^h,t^h_{x^h},T^h,T) + \delta_{x^h=0}\,\mathrm{Pr}(\text{dead at } T|x,t_x,T)\end{aligned}$$

For a randomly chosen customer, we obtain the likelihood expression by integrating $L^h(\lambda,\mu|x,t_x,T,x^h,t^h_x,T^h)$ over λ and μ using their joint posterior distribution, $g(\lambda,\mu|r,\alpha,s,\beta,x,t_x,T)$

$$\begin{aligned}&L^h(r,\alpha,s,\beta|x,t_x,T,x^h,t^h_x,T^h) \\ &\quad= \int_0^\infty\int_0^\infty L^h(\lambda,\mu|x,t_x,T,x^h,t^h_x,T^h)g(\lambda,\mu|r,\alpha,s,\beta,x,t_x,T)d\lambda d\mu \\ &\quad= \int_0^\infty\int_0^\infty \mathrm{Pr}(\text{alive at } T|x,t_x,T)\mathcal{L}(\lambda,\mu|x^h,t^h_{x^h},T^h,T)g(\lambda,\mu|r,\alpha,s,\beta,x,t_x,T)d\lambda d\mu \\ &\qquad+ \delta_{x^h=0}\int_0^\infty\int_0^\infty \mathrm{Pr}(\text{dead at } T|x,t_x,T)g(\lambda,\mu|r,\alpha,s,\beta,x,t_x,T)d\lambda d\mu \\ &\quad= \mathsf{A} + \delta_{x^h=0}\mathsf{B}\end{aligned} \tag{1}$$

To solve A:

$$\begin{aligned}\mathsf{A} &= \int_0^\infty\int_0^\infty \mathrm{Pr}(\text{alive at } T|x,t_x,T)\mathcal{L}(\lambda,\mu|x^h,t^h_{x^h},T^h,T)g(\lambda,\mu|r,\alpha,s,\beta,x,t_x,T)d\lambda d\mu \\ &= \int_0^\infty\int_0^\infty \frac{\lambda^x e^{-(\lambda+\mu)T}}{L(\lambda,\mu|x,t_x,T)}\mathcal{L}(\lambda,\mu|x^h,t^h_{x^h},T^h,T)\frac{L(\lambda,\mu|x,t_x,T)g(\lambda|r,\alpha)g(\mu|s,\beta)}{L(r,\alpha,s,\beta|x,t_x,T)}d\lambda d\mu \\ &= \int_0^\infty\int_0^\infty \lambda^x e^{-(\lambda+\mu)T}\left(\frac{\lambda^{x^h}\mu}{\lambda+\mu}e^{-(\lambda+\mu)(t^h_{x^h}-T)} + \frac{\lambda^{x^h+1}}{\lambda+\mu}e^{-(\lambda+\mu)(T^h-T)}\right)\frac{g(\lambda|r,\alpha)g(\mu|s,\beta)}{L(r,\alpha,s,\beta|x,t_x,T)}d\lambda d\mu \\ &= \frac{1}{L(r,\alpha,s,\beta|x,t_x,T)}\int_0^\infty\int_0^\infty\left(\frac{\lambda^{x+x^h}\mu}{\lambda+\mu}e^{-(\lambda+\mu)t^h_{x^h}} + \frac{\lambda^{x+x^h+1}}{\lambda+\mu}e^{-(\lambda+\mu)T^h}\right)g(\lambda|r,\alpha)g(\mu|s,\beta)d\lambda d\mu \\ &= \frac{1}{L(r,\alpha,s,\beta|x,t_x,T)}\int_0^\infty\int_0^\infty L(\lambda,\mu|x+x^h,t^h_{x^h},T^h)g(\lambda|r,\alpha)g(\mu|s,\beta)d\lambda d\mu \\ &= \frac{L(r,\alpha,s,\beta|x+x^h,t^h_{x^h},T^h)}{L(r,\alpha,s,\beta|x,t_x,T)}\end{aligned}$$

where $t^h_{x^h} = T$ if $x^h = 0$ and the last step follows from observing equation (15) in Fader and Hardie (2005).

To solve B:

$$\begin{aligned}
\mathsf{B} &= \int_0^\infty \int_0^\infty \Pr(\text{dead at } T|x, t_x, T) g(\lambda, \mu|r, \alpha, s, \beta, x, t_x, T) d\lambda d\mu \\
&= \int_0^\infty \int_0^\infty \left(1 - \Pr(\text{alive at } T|x, t_x, T)\right) g(\lambda, \mu|r, \alpha, s, \beta, x, t_x, T) d\lambda d\mu \\
&= 1 - \int_0^\infty \int_0^\infty \Pr(\text{alive at } T|x, t_x, T) g(\lambda, \mu|r, \alpha, s, \beta, x, t_x, T) d\lambda d\mu \\
&= 1 - \frac{\Gamma(r+x)\alpha^r\beta^s}{\Gamma(r)(\alpha+T)^{r+x}(\beta+T)^s} \bigg/ L(r, \alpha, s, \beta|x, t_x, T)
\end{aligned}$$

where the last step is from equation (34) in Fader and Hardie (2005).

Substituting for A and B in (1), we obtain:

$$\begin{aligned}
L^h(r, \alpha, s, \beta|x, t_x, T, x^h, t^h_x, T^h) = &\frac{L(r, \alpha, s, \beta|x + x^h, t^h_{x^h}, T^h)}{L(r, \alpha, s, \beta|x, t_x, T)} \\
&+ \delta_{x^h=0} \left(1 - \frac{\Gamma(r+x)\alpha^r\beta^s}{\Gamma(r)(\alpha+T)^{r+x}(\beta+T)^s} \bigg/ L(r, \alpha, s, \beta|x, t_x, T)\right)
\end{aligned}$$

that can be written in a more intuitive form as

$$\begin{aligned}
L^h(r, \alpha, s, \beta|x, t_x, T, x^h, t^h_x, T^h) = &\frac{L(r, \alpha, s, \beta|x + x^h, t^h_{x^h}, T^h)}{L(r, \alpha, s, \beta|x, t_x, T)} \\
&+ \delta_{x^h=0} \left(1 - \Pr(\text{alive at } T|r, \alpha, s, \beta, x, t_x, T)\right) \qquad (2)
\end{aligned}$$

where $t^h_{x^h} = T$ if $x^h = 0$ and $t_x = 0$ if $x = 0$.

References

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/`>

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.