A Spreadsheet-Literate Non-Statistician's Guide to the Beta-Geometric Model

Peter S. Fader
www.petefader.com

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

December 2014

1 Introduction

The beta-geometric (BG) distribution is a robust simple model for characterizing and forecasting the length of a customer's relationship with a firm in a contractual setting. Fader and Hardie (2007a), hereafter FH, is a definitive reference, providing a detailed derivation of the key quantities of interest, as well as step-by-step details of how to implement the model in Excel. However the statistical concepts and notation used in that paper can be daunting for those analysts who do not have a strong statistics background, leading them to ignore the model when it should really be a basic tool in their analytics toolkit.

With spreadsheet-literate non-statisticians in mind as the target audience, the objective of this note is to describe the logic of the BG model in a non-technical manner and show how to implement it in Excel.

2 Motivating Problem

Consider a company with a subscription-based business model that acquired 1000 customers (on annual contracts) at the beginning of Year 1. Table 1 reports the pattern of renewals by this cohort over the subsequent four years.[1] We would like to predict how many members of this cohort will still be customers of the firm in Years 6, 7, ...

[1] A cohort is a group of customers acquired at the same time.

ID	Year 1	Year 2	Year 3	Year 4	Year 5
0001	1	1	0	0	0
0002	1	0	0	0	0
0003	1	1	1	0	0
0004	1	1	0	0	0
0005	1	1	1	1	1
⋮		⋮		⋮	
0998	1	0	0	0	0
0999	1	1	1	0	0
1000	1	0	0	0	0
	1000	631	468	382	326

Table 1: Pattern of year-on-year renewals for the cohort of 1000 customers acquired at the beginning of Year 1.

3 Notation and Terminology

We note that the proportion of the cohort "surviving" beyond the first renewal opportunity is $631/1000 = 0.631$. Similarly, the proportion of the cohort surviving beyond of the second renewal opportunity is $468/1000 = 0.468$. This notion of surviving beyond a particular point in time is captured by the *survivor function*.

More formally, the survivor function $S(t)$ is the probability that a randomly chosen member of the original cohort of customers survives beyond time t. With reference to Figure 1, a customer is "born" at $t = 0$ (the beginning of Year 1) and therefore, by definition, $S(0) = 1$. The empirical survivor function (i.e., the survivor function computed directly from the data) for this cohort is $S(1) = 0.631$, $S(2) = 0.468$, $S(3) = 0.382$ and $S(4) = 0.326$.

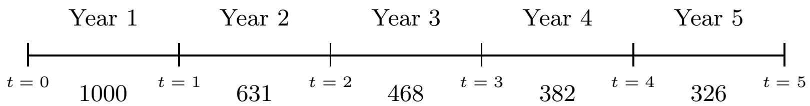


Figure 1: Summary of the cohort's subscription's renewal behavior.

We also note that the proportion of Year 1 customers who are customers in Year 2 is $631/1000 = 0.631$. Similarly, the proportion of Year 2 customers who are customers in Year 3 is $468/631 = 0.742$. This is captured by the notion of the *retention rate*, denoted by $r(t)$, which is the proportion of customers who survived beyond $t-1$ who also survive beyond t. The empirical retention rates for this cohort are $r(1) = 0.631$, $r(2) = 0.742$, $r(3) = 0.816$ and $r(4) = 0.853$. Retention rates can be computed directly from the data (as above) or via the survivor function using the following formula:

$$r(t) = \frac{S(t)}{S(t-1)}\,, \;\; t = 1, 2, 3, \ldots$$

Equivalently, given knowledge of the retention rates, we can compute the survivor function using the following *forward recursion*:

$$S(t) = \begin{cases} 1 & \text{if } t = 0 \\ r(t) \times S(t-1) & \text{if } t = 1, 2, 3, \ldots \end{cases} \tag{1}$$

4 The Beta-Geometric Model

The beta-geometric model is based on the following "as if" story of customer behavior:

i) At the end of each contract period, an individual customer decides whether or not to renew her contract by tossing a coin: "heads" (H) she renews her contract, "tails" (T) she cancels it. (Note that we are not assuming that this is a "fair" coin; that is, we are not assuming that there is a 50% chance of the coin coming up H when it is tossed.)

ii) For any given individual, the probability of her coin coming up T, Prob(T), does not change over time.

iii) The probability of a coin coming up T varies across customers.

The third element of this story should be not be controversial; after all, the notion of people being different (cross-sectional heterogeneity, to use technical terminology) is central to marketing, as manifest in the fundamental concept of segmentation.

On the other hand, the individual-level coin-flipping story may raise a few eyebrows. However, the thing to note is that we are not saying that people actually make their contract renewal decisions on the basis of coin flips. There are a thousand and one, if not a million and one, different reasons as to why a customer chooses to end their "relationship" with a firm. Even if the actual process were completely deterministic, it would be impossible to measure all the variables that determine an individual's behavior. We therefore claim ignorance and, from the perspective of an outside observer, view contract renewal as a chance (random) occurrence. The image of customers flipping coins is an "as if" story, and the fundamental question will be whether this so-called data-generating process captures (and, more fundamentally, predicts) the patterns of behavior we observe in the data.

Finally, the second element of this story may seem puzzling given that we typically observe increasing retention rates when we track the "survival" of a cohort of customers over time (as seen above). We will discuss this at a later stage.

In order for this verbal story of customer behavior to be of any use to the analyst wishing to generate estimates of survival beyond Year 5 for the data in Table 1, we need to translate it into the language of mathematics and then into Excel.

As we learn in many introductory probability courses, the first two elements of our story are equivalent to saying that the duration of an individual customer's relationship with the firm is characterized by the *geometric distribution* (i.e., the number of coin tosses before the coin comes up tails for the first time).

For anyone with a marketing background, it would appear that the simplest way of operationalizing the third element of our story of customer behavior is to assume the existence of two (or three, four, five, . . .) segments of customers, where the members of each segment all carry the same type of coin (i.e., with the same probability of coming up T) but the coins differ among segments. While intuitively appealing, such an approach does constrain each individual's probability of churning at the end of each contract period to one of a small number of specific values (i.e., those associated with each discrete segment). It turns out that a more parsimonious approach to capturing consumer heterogeneity is to assume that each individual's probability can take on any one of an infinite number of possible values between 0 and 1; this is achieved by assuming that variability in these probabilities across customers is captured by a continuous probability distribution. Whenever statisticians need a probability distribution to characterize something that can vary between between 0 and 1, they naturally turn to the *beta distribution*; they do so because it is both flexible (i.e., it can capture a lot of different patterns of heterogeneity) and easy to work with when performing various mathematical calculations. We follow this practice and operationalize the third element of our story by assuming that heterogeneity in Prob(T) is captured by a beta distribution.

Before going any further, let us briefly talk about the beta distribution. Going back to our introductory probability and statistics courses, discussions of the key parameters of any probability distribution focus on its mean and variance. In the case of the beta distribution, we will talk about γ and δ (gamma and delta for those unfamiliar with the Greek alphabet)[2], which are related to the mean and variance in the following manner:

$$\text{mean:} \quad \frac{\gamma}{\gamma+\delta}$$

$$\text{variance:} \quad \frac{\gamma\delta}{(\gamma+\delta)^2(\gamma+\delta+1)}$$

Why do statisticians "parameterize" the beta distribution in terms of γ and δ and not the mean and variance? First, it makes the math easier. Second γ and δ actually give us a better sense of the nature of the heterogeneity we are trying to capture (i.e., how Prob(T) varies across people). (See Appendix A for an examination of the shapes the beta distribution can take on.)

[2]FH use α and β (alpha and beta) instead of γ and δ.

Combining these two probability distribution gives us the beta-geometric (BG) distribution as a model of contract duration in a contractual setting. (We do not focus on the actual derivation here; the interested reader can find all the details in FH.) Of particular interest is the following expression for the retention rate under the BG model:

$$r(t \mid \gamma, \delta) = \frac{\delta + t - 1}{\gamma + \delta + t - 1}\,, \quad t = 1, 2, 3, \ldots \tag{2}$$

Given (2), we can easily compute the corresponding survivor function, $S(t)$, using the forward-recursion given in (1). Note that these two quantities are computed using only the four basic arithmetic operations (i.e., addition, subtraction, multiplication and division):

$$\begin{aligned} S(0) &= 1\,, \\ S(1) &= S(0) \times r(1) \\ &= \frac{\delta}{\gamma + \delta}\,, \\ S(2) &= S(1) \times r(2) \\ &= \frac{\delta}{\gamma + \delta} \times \frac{\delta + 1}{\gamma + \delta + 1}\,, \end{aligned}$$

and so on.

5 Fitting the Model to Data

In order to use the BG model to solve the motivating problem (i.e., generate accurate estimates of customer survival beyond Year 5), we need to know the numerical values of γ and δ that are mostly likely to have generated the pattern of renewals observed in Table 1.

FH used the method of maximum likelihood to arrive at estimates of γ and δ. While such an approach has some desirable statistical properties, the process of implementing it is not immediately obvious to the non-statistician. Instead, we will use a simpler regression-like approach that is much easier to understand and implement.

The basic approach we take is as follows. The observed retention rates are $r(1) = 0.631$, $r(2) = 0.742$, $r(3) = 0.816$, and $r(4) = 0.853$. We will find the values of γ and δ that make the model-based estimates of $r(1), \ldots, r(4)$, as computed using (2), as "close" as possible to the corresponding observed values.

The Excel worksheet we use to do this is shown in Figure 2 and is constructed in the following manner.

- We start by entering the observed data. The number of Year 1 customers (1000) is entered in cell `B6`, the number for Year 2 (631) is entered in cell `B7`, and so on down to 326 in cell `B10)` for Year 5.

	A	B	C	D	E
1	gamma	1.000			
2	delta	1.000			
3	SSE	3.00E-02			
4			Retention rate		
5	t	# Cust.	Actual	Model	Sq. Error
6	0	1000			
7	1	631	0.631	0.500	1.72E-02
8	2	468	0.742	0.667	5.63E-03
9	3	382	0.816	0.750	4.39E-03
10	4	326	0.853	0.800	2.85E-03

Figure 2: Screenshot of the Excel worksheet for parameter estimation.

- We enter the values of $t = 0, 1, \ldots, 4$ in cells `A6:A10` (corresponding to the beginning of Years 1–5, as in Figure 1).
- We compute the observed Year 1 retention rate in cell `C7` using the formula `=B7/B6`, and copy it down to cell `C10` to compute the observed retention rates for the other years.
- In order to enter the expression for $r(t)$ under the BG model without generating a `#DIV/0!` error message, we need some "starting values" for γ and δ. The exact values do not matter (provided they are greater than 0), so we start with 1 for both γ and δ, locating these parameter values in cells `B1:B2`.
- We compute the model-based $r(1)$ (for the values of γ and δ in cells `B1:B2`) by entering `=($B$2+A7-1)/($B$1+$B$2+A7-1)` in cell `D7`, and copy this formula down to cell `D10` to give us the model-based retention rates for the other years.
- Our objective is to find that values of γ and δ that make the numbers in cells `D7:D10` as "close" as possible to those in cells `C7:C10`. A natural way to assess closeness is examining (and ultimately minimizing) the *squared difference* between each pair of numbers. This is exactly what happens in an ordinary linear regression. Thus we seek the parameter values that minimize the sum of the squared differences between the actual and model-based estimates of the quantity of interest; these are called the least-squares estimates of the model parameters. (These differences are called "error" and so we seek to minimize the sum of squared errors, SSE.)
- We compute the squared error associated with the Year 1 retention-rate numbers by entering `=(C7-D7)^2` in cell `E7`. We copy this formula down to cell `E10` to give us the squared error numbers for the other years.
- We compute the sum of squared errors by entering `=SUM(E7:E10)` in cell `B3`;

this is the value of the SSE given the values for the two model parameters in cells `B1:B2`. (With starting values of 1 for both parameters, `SSE` $= 3.00E{-}02$.)

Our least-squares estimates of the two model parameters are those that minimize the value of the SSE function. (Strictly speaking, we are computing the nonlinear least-squares (NLS) estimates of the model parameters. We use the term nonlinear because (2) is a nonlinear function of t.) We do this using the Excel add-in Solver, available on the "Data" tab. The *target cell* is the value of the SSE, cell `B3`. We wish to *minimize* this by *changing* cells `B1:B2`. The *constraints* we place on the parameters are that γ and δ be greater than 0. As Solver only offers us a "greater than or equal to" constraint, we *add* the constraint that cells `B1:B2` are $\geq$ a small positive number (e.g., 0.0001)—see Figure 3.

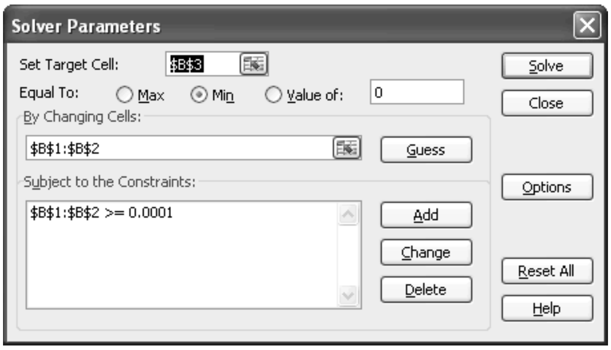


Figure 3: Setting up Solver to find the values of γ and δ that minimize SSE for the BG model.

Clicking the *Solve* button, Solver converges to a solution where the minimum value of the SSE is $1.16E{-}04$, associated with $\gamma = 0.760$ and $\delta = 1.286$. These are the NLS estimates of the model parameters.[3] (So as to be sure that we have actually found the minimum value of SSE, it is good practice to redo the optimization process using a completely different set of starting values. For example, using starting values of 0.01 and 0.01 (for which `SSE` $= 8.02E{-}02$), use Solver to find the minimum value of SSE. Are the corresponding values of the two model parameters equal to those given above? They should be.)

[3] While fitting the BG model to these same data using the method of maximum likelihood yields slightly different parameter estimates ($\gamma = 0.764$ and $\delta = 1.296$), the resulting estimates of the retention rates are effectively the same as those computed using our least-squares estimates (differing at the fourth decimal place before rounding).

6 Interpreting the Model Parameters

In fitting this model to the data, we are actually estimating the distribution of the underlying Prob(T) across the cohort of 1000 customers acquired at the beginning of Year 1. The distribution associated with parameter values $\gamma = 0.760$ and $\delta = 1.286$ is plotted in Figure 4. (See Appendix B for details of how to create this plot.)

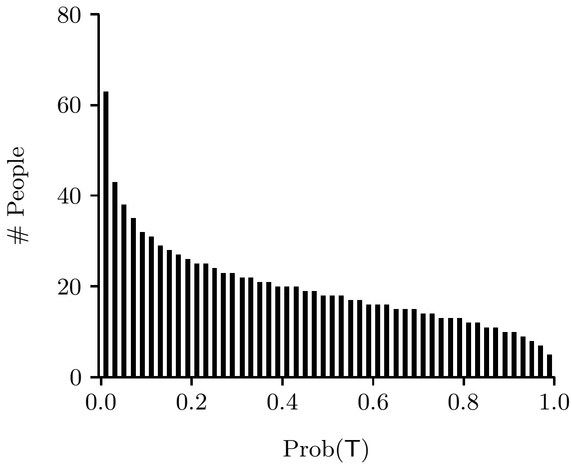


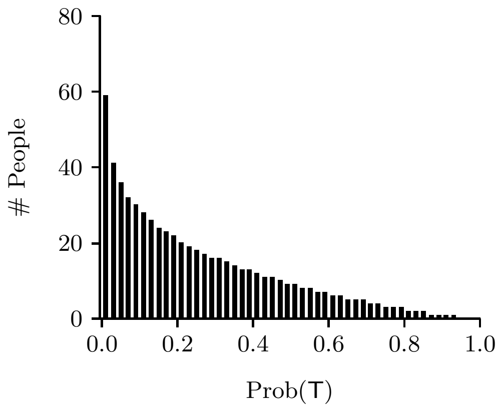
Figure 4: Estimated Distribution of Prob(T)

We see that 63 of the 1000 customers acquired at the beginning of Year 1 are deemed to have a coin for which Prob(T) is somewhere between 0.00 and 0.02, 43 with a coin for which Prob(T) is somewhere between 0.02 and 0.04, ..., right down to 5 with a coin for which Prob(T) is somewhere between 0.98 and 1.00.

At the end of the first year, all 1000 customers toss their coins (i.e., decide whether or not to renew their contracts). The average of Prob(T) across these 1000 customers is $0.760/(0.760 + 1.286) = 0.371$, which implies that 37.1% of the original 1000 cohort members will not renew their subscription at the end of Year 1, while 62.9% will renew. (This number is very close to the 63.1% we observe in the actual data.)

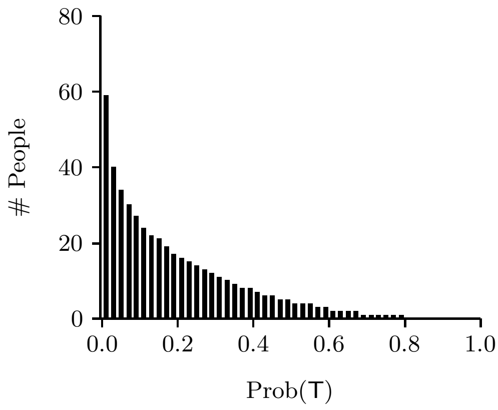
The distribution of Prob(T) across the 629 survivors is given in Figure 5a. Comparing this with Figure 4, we see that most of those customers with a high Prob(T) did indeed see their coins come up T and so did not renew their contract. However, most of the customers with a low Prob(T) successfully renewed. At the end of the second year, all 629 Year 1 renewers toss their coins. The average of Prob(T) across these customers is 0.249, which implies that 75.1% of them (472) will renew their subscription for a second time, while the remaining 24.9% will not renew.

The distribution of Prob(T) across the 472 survivors is given in Figure 5b. Comparing this with Figure 5a, we once again see that those customers with a high Prob(T) did indeed see their coins come up T and so did not renew their contract.



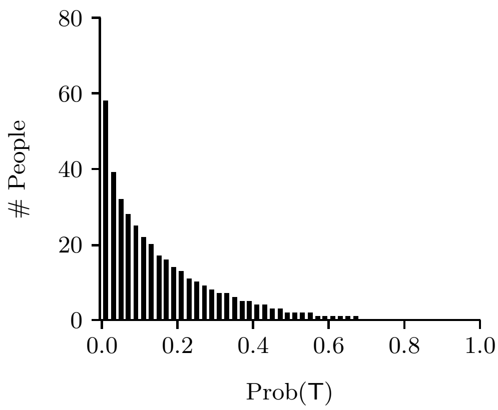
Renewed at the end of Year 1

(a)



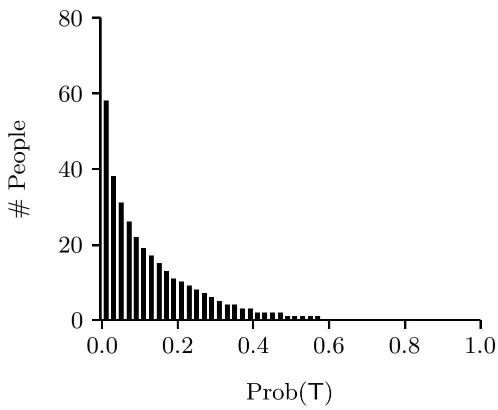
Renewed at the end of Year 2

(b)



Renewed at the end of Year 3

(c)



Renewed at the end of Year 4

(d)

Figure 5: Distribution of Prob(T) amongst surviving customers over time.

We note that while a few customers with a very high Prob(T) made it into the second year (Figure 5a), they did not make it into the third year (Figure 5b) — if you have a coin with Prob(T) = 0.95, the probability of getting HH, which is required to survive into Year 3, is very small, and we do not see this occurring amongst our cohort of 1000 customers.

The equivalent distributions for those that renew their subscriptions at the end of Years 3 and 4 are given in Figure 5c and Figure 5d. We note that, over time, those customers with coins that have higher Prob(T) are dropping out. This is reflected in the means the distribution in Figures 5b – 5d, which are, respectively, 0.188, 0.151, and 0.126. (This implies retention rates of 0.812, 0.849, and 0.874.) However, some individuals with lower values for Prob(T) are also disappearing. For example, after four coin tosses, we have lost about three-quarters of those original customers with Prob(T) somewhere between 0.28 and 0.30. (This is not surprising; the probability of seeing HHHH when Prob(T) $= 0.3$ is $(1-0.3)^4 = 0.24$.)

Note that the retention rates are increasing, something implied by (2), even though the second element of the "as if" story of customer behavior underpinning the BG model assumes no dynamics at the level of the individual customer. The observed phenomenon of retention rates increasing over time is simply an artifact of heterogeneity — those customers with coins that have higher Prob(T) are dropping out over time, leaving an ever-smaller pool of customers holding coins with lower Prob(T).

7 Generating Forecasts

Returning to our motivating problem, we have the actual renewal data for this cohort of customers for another eight years beyond those given in Table 1. This allows us to assess the predictive performance of the BG model.

We need to compute $S(t)$ out to $t = 12$ (i.e., surviving into Year 13). In order to do this, we compute $r(t)$ out to $t = 12$ using (2) and then use (1) to compute the survivor function. The Excel worksheet we use to do this is shown in Figure 6 and is constructed in the following manner.

- Our estimates of γ and δ are entered in cells `B1:B2`.
- We enter the values of $t = 0, 1, \ldots, 12$ in cells `A5:A17`.
- We compute the model-based estimate of $r(1)$ by entering `=($B$2+A6-1)/($B$1+$B$2+A6-1)` in cell `B6`, and copy this formula down to cell `B17` to compute the retention rates for the next 11 years.
- Given these retention rates, we compute the values of $S(t)$ using the forward-recursion formula given in (1):
  - By definition $S(0) = 1$, which we enter in cell `C5`.
  - We compute $S(1)$ by entering `=B6*C5` in cell `C6`.
  - We copy this formula down to `C17`.

In Figure 7, we compare the predicted retention rates with those actually observed over both the model calibration period and the longitudinal holdout (forecast) period. The predicted Year 12 retention rate is 0.942, while the actual proportion of those Year 12 subscribers who renewed their subscriptions at the end of Year 12 is 0.945.

	A	B	C
1	gamma	0.760	
2	delta	1.286	
3			
4	t	r(t)	S(t)
5	0		1.000
6	1	0.629	0.629
7	2	0.751	0.472
8	3	0.812	0.383
9	4	0.849	0.326
10	5	0.874	0.285
11	6	0.892	0.254
12	7	0.906	0.230
13	8	0.916	0.211
14	9	0.924	0.195
15	10	0.931	0.181
16	11	0.937	0.170
17	12	0.942	0.160

Figure 6: Screenshot of the Excel worksheet used to compute the survivor function.

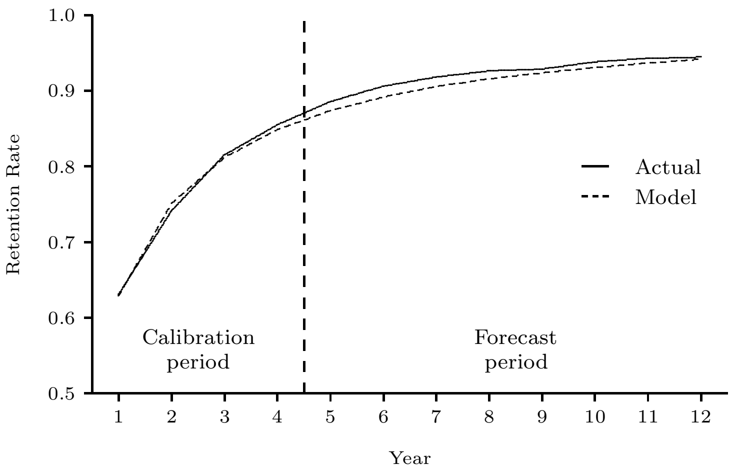


Figure 7: Actual vs. model-based estimates of the annual retention rates.

Our prediction of how many members of the original cohort that are still customers in any given year is computed by multiplying the BG estimates of $S(t)$ (column `C`) by 1000. In Figure 8, we compare these predictions with the actual numbers over both the model calibration period and the longitudinal holdout (forecast) period. The model predicts that 160 of the original 1000 will still be customers in Year 13; the actual number is 173. This is impressive given the forecasting horizon (relative to the length of the model calibration period) and the simplicity of the model.

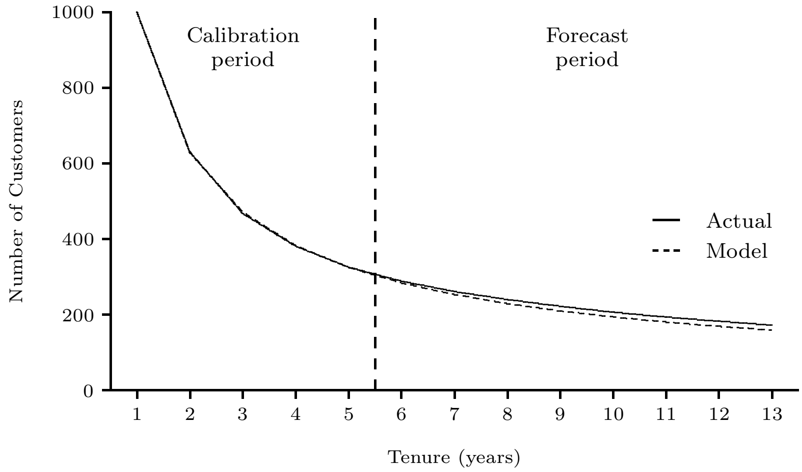


Figure 8: Actual vs. model-based estimates of the number of surviving customers.

Some readers may be thinking "Why bother with this "as if" story of customer behavior? Sure, the model works well, but do we really need this heterogeneous coin-flipping baggage? Why not just fit some flexible function of time directly to either the survival data or the retention data and use that to generate the required forecasts?"

We discourage such thinking for two reasons. First, as documented in FH and Fader and Hardie (2007b), the BG model is more robust than various flexible functions of time. Second, we actually find that it is easier to explain the logic of the BG model to the end-user (using the coin-flipping story given above) than it is to wave one's hands and talk about fitting flexible (but arbitrary) functions of time to the data. As this note has hopefully illustrated, it is easy for a spreadsheet-literate non-statistician to implement the BG model using a simple Excel spreadsheet.

References

Fader, Peter S. and Bruce G. S. Hardie (2007a), "How to Project Customer Retention," *Journal of Interactive Marketing*, **21** (Winter), 76–90.

Fader, Peter S. and Bruce G.S. Hardie (2007b), "How Not to Project Customer Retention." (`http://brucehardie.com/notes/016/`)

Appendix A: The Shape of the Beta Distribution

As we see in Figure A1, the shape of the beta distribution depends on the relative magnitude of γ and δ, and whether they are greater than or less than 1.

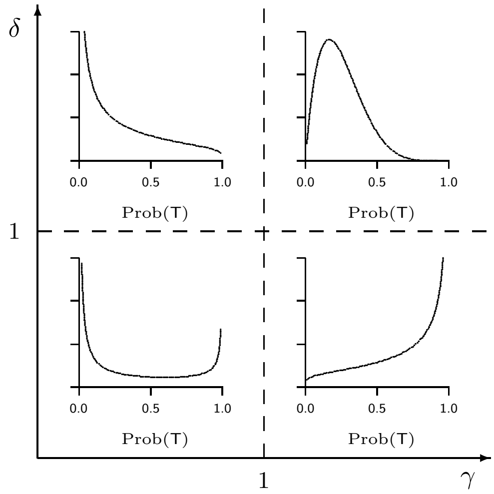


Figure A1: General shapes of the beta distribution as a function of γ and δ.

- When both γ and δ are less than 1 (bottom-left quadrant) we have a "U-shaped" distribution. In such a setting, this is both a large fraction of the population holding coins with Prob(T) close to 1 (who will churn at the first opportunity to do so), and a large fraction of the population holding coins with Prob(T) close to 0 (who will likely remain as customers for a very long time). As γ and δ get closer and closer to 0, the distribution becomes more and more polarized, with no one populating the middle area and everyone piling up at either 0 or 1.

- When both γ and δ are greater than 1 (top-right quadrant) we have an interior mode (the exact location of which depends on relative magnitude of γ vis-à-vis δ). As both γ and δ get larger and larger, there is less and less variability in Prob(T) across individuals. (Referring back to the expression for the variance of the beta distribution, the variance gets smaller and smaller.) In the limit, the distribution becomes a spike located at the mean (i.e., there is no heterogeneity in Prob(T)).

- When γ is greater than 1 and δ is less than 1 (bottom-right quadrant) we have a "J-shaped" distribution. As γ gets larger, and δ gets closer to 0, more and more of the population will pile up towards Prob(T) $= 1$.

- When γ is less than 1 and δ is greater than 1 (top-left quadrant) we have a "reverse-J-shaped" distribution. As δ gets larger, and γ gets closer to 0, more and more of the population will pile up towards $\text{Prob}(\mathsf{T}) = 0$.
- As a technical aside, when both γ and δ equal 1, the distribution is a flat line between 0 and 1; in other words, the beta distribution collapses to the uniform distribution.

As such, we can think of γ as trying to push the distribution towards $\text{Prob}(\mathsf{T}) = 1$ and δ as trying to push the distribution towards $\text{Prob}(\mathsf{T}) = 0$. A third force is "gravity" pushing down on the middle; when both γ and δ are greater than 1, we "break through" the force of gravity and have an interior mode.

Appendix B: Creating Figure 4 in Excel

The beta distribution is what statisticians call a continuous distribution. Rather than trying to interpret the "raw" plot of the beta distribution (which does not come naturally to the non-statistician), we find it better to create and present a discretized plot such as that given in Figure 4.

At the heart of this exercise is the Excel function `BETA.DIST`, which computes the probability that Prob(T) is less than or equal to a specific value. We use this in the Excel worksheet shown in Figure B1 to create Figure 4. This worksheet is constructed in the following manner.

	A	B	C	D
1	gamma	0.760		
2	delta	1.286		
3				
4	x	P(Prob(T) <= x)	P(x-0.02<Prob(T) <= x)	# people
5	0.00	0.0000		
6	0.02	0.0629	0.0629	63
7	0.04	0.1062	0.0433	43
8	0.06	0.1441	0.0379	38
9	0.08	0.1788	0.0347	35
10	0.10	0.2113	0.0325	32
53	0.96	0.9884	0.0080	8
54	0.98	0.9952	0.0069	7
55	1.00	1.0000	0.0048	5

Figure B1: Screenshot of the Excel worksheet used to create Figure 4.

- Our estimates of γ and δ are entered in cells `B1:B2`.
- We want to compute the number of people holding a coin whose Prob(T) falls in an interval of width 0.02, so we need a column containing $0.00, 0.02, 0.04, \ldots,$ $0.98, 1.00$. First we enter 0 in cell `A5`. Next we enter `=ROUND(A4+0.02,2)` in cell `A6` and copy this formula down to cell `A55`. We label this column `x` (cell `A4`).[i]
- We now want to compute the probability that the value of Prob(T) for a randomly chosen individual is less than or equal to `x`. We compute this in column `B` by entering `=BETA.DIST(A5,$B$1,$B$2,TRUE)` in cell `B5` and copying the formula down to `B55`. Since the possible values of Prob(T) are bounded between 0 and 1, it makes sense that the probability of Prob(T) being less than or equal to 0, $P(\text{Prob}(\mathsf{T}) \leq 0)$, is 0 and that the probability of Prob(T) being less than or equal to 1, $P(\text{Prob}(\mathsf{T}) \leq 1)$, is 1.

[i] In theory, we should not have to use the `ROUND` function. However, Excel is not a great environment for precise numerical computation and, if we do not use the `ROUND` function, the value of cell `A55` is not 1 but $1 + 4.44E{-16}$. This results in a `#NUM!` error in `B55` in the next step.

- With reference to cell `B6`, we see that $P(\text{Prob}(\mathsf{T}) \leq 0.02) = 0.0629$.
- Looking at cell `B7`, we see that $P(\text{Prob}(\mathsf{T}) \leq 0.04) = 0.1062$. However we are not really interested in this; rather, we want to know that probability that Prob(T) is between 0.02 and 0.04, $P(0.02 < \text{Prob}(\mathsf{T}) \leq 0.04)$.
- We recall from the basic rules of probability that $P(0.02 < \text{Prob}(\mathsf{T}) \leq 0.04) = P(\text{Prob}(\mathsf{T}) \leq 0.04) - P(\text{Prob}(\mathsf{T}) \leq 0.02)$, which in this specific instance equals 0.0433.
- More generally, we compute $P(\mathsf{x} - 0.02 < \text{Prob}(\mathsf{T}) \leq \mathsf{x})$ in column `C` by entering `=B6-B5` in cell `C6` and copying this formula down to cell `C55`.
- We see that the probability that a randomly chosen member of the cohort has a coin with Prob(T) between 0.00 and 0.02 is 0.0629. Similarly, the probability that a randomly chosen member of the cohort has a coin with Prob(T) between 0.02 and 0.04 is 0.0433. And so on.
- Figure 4 reports these probabilities in terms of the expected number of cohort members holding a coin with Prob(T) lying in the specified interval. We compute these numbers in column `D` by entering `=ROUND(1000*C6,0)` in cell `D6` and copying this formula down to cell `D55`. (Note that in this case, cells `D6:D55` sum to 999 due to the rounding to 0 decimal places. If the `ROUND` function is not used, these cells sum to 1000.)

The plots in Figure 5 are created in a similar manner, albeit with the following two modifications:

- The distribution of Prob(T) across those individuals who have made n renewals is captured by a beta distribution with parameters γ and $\delta + n$.[ii]
- The number of people who have made n renewals (for the cohort whose behavior is summarized in Table 1) is $1000 \times S(n)$.

(Clearly Figure 4 corresponds to the case of $n = 0$.) For example, Figure 5a gives us the distribution of Prob(T) across those members of the cohort that renewed at the end of Year 1. Here $n = 1$, so the value of cell `B2` is now 2.286 and the formulas in column `D` use 629 in place of 1000.

[ii] The derivation of this is given in

Fader, Peter S. and Bruce G. S. Hardie (2010), "Customer-Base Valuation in a Contractual Setting: The Perils of Ignoring Heterogeneity," *Marketing Science*, **29** (January–February), 85–93. <`http://brucehardie.com/papers/022/`>