Customer Lifetime Value Part II non-contractual settings

Customer Analytics

Lecture 7

Agenda

- Non-contractual settings
- Concept: latent attrition
- BG/BB model
- Other latent attrition models (Pareto/NBD model, BG/NBD)

Review:

Netflix

Netflix said it added 7.05 million subscribers during the fiscal 4th quarter. For the quarter, Netflix added 1.93 million memberships in the U.S. and 5.12 million internationally.

Bol

Bol.com opened its doors on 30 March 1999. Eighteen years later, the store has more than 7 million active customers* in the Netherlands and Belgium and an assortment of almost 15 million items.

* customers are considered active when they have placed an order during the preceding twelve-month period

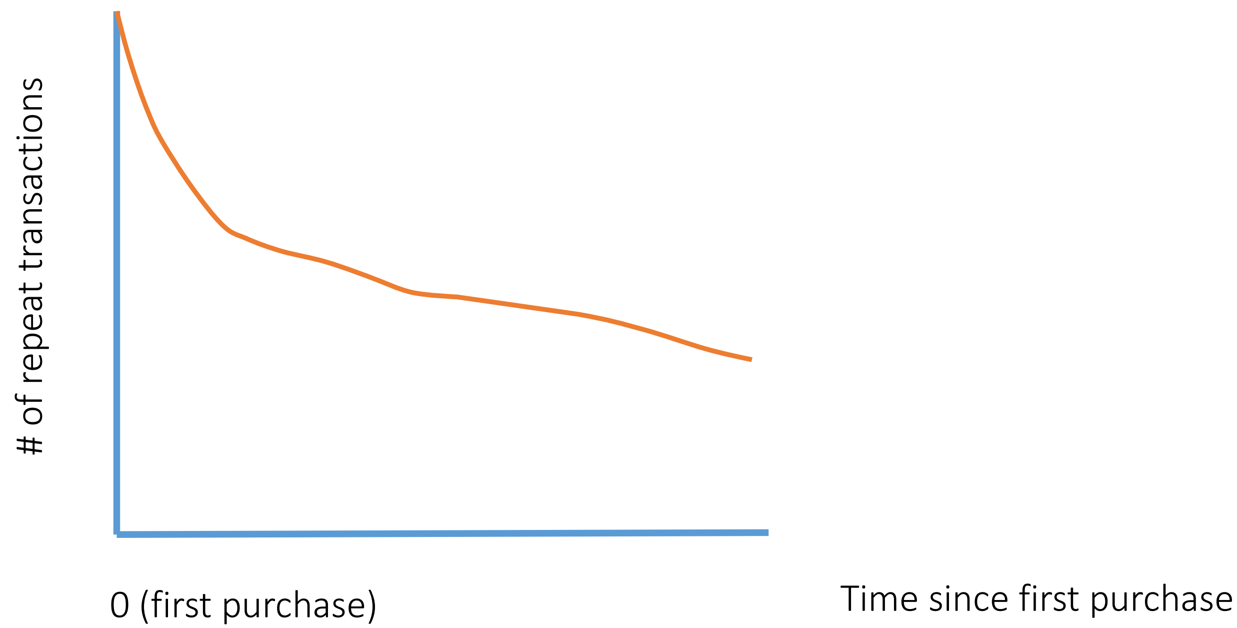
How would these numbers change if the threshold were 24 months instead of 12? Or 6 months?

What's the right number of active customers?

Review

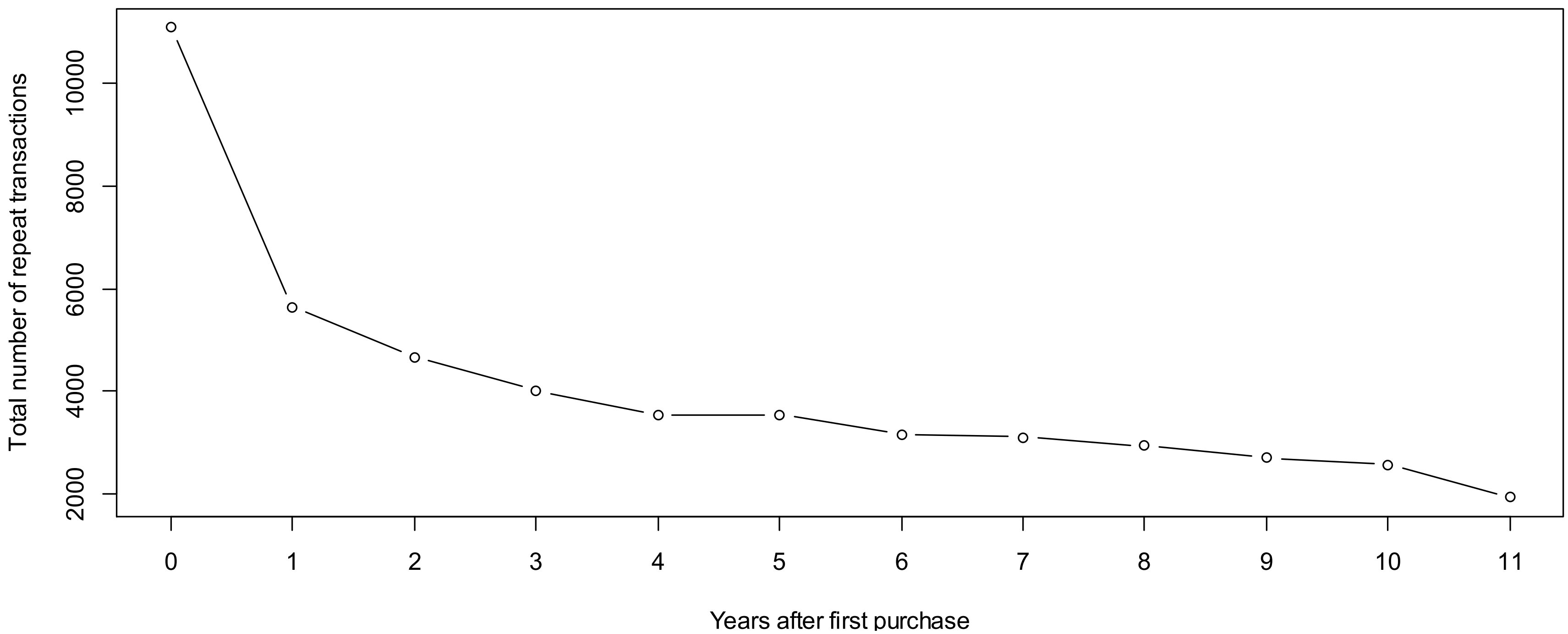
Contractual settings	Non-contractual (a.k.a. transactional) settings
• Minority of companies, but growing	• Majority of companies
• Customers notify the company to quit (ending contract)	• Customers silently leave
• Any subscription business model: gym membership, internet/cable, bank, insurance	• Grocery store, retailers, fast moving consumer goods (FMCG), hotels, airlines, media
• Focus is on retention rate, because quitting is observed.	• Focus is on repeat purchasing
Last class	This class

Repeat sales for a cohort of customers

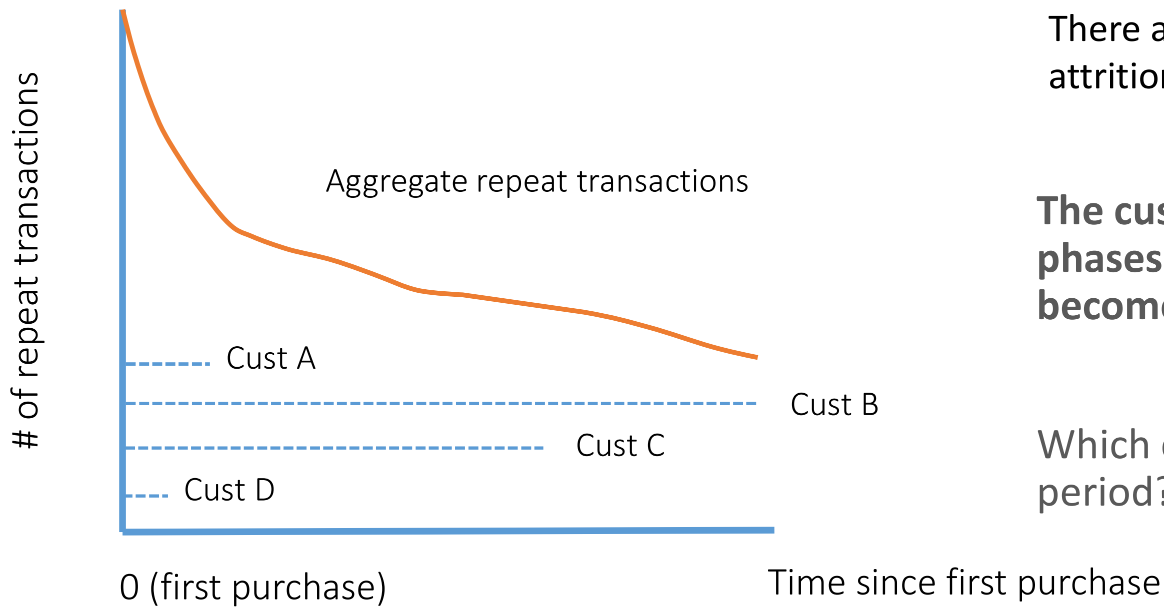


This is purchasing for an unchanging <u>cohort of customers</u> who started at the same time (year 0)

Declining cohort level repeat purchasing



Latent (unobserved) attrition models



Buy Til You Die

There are several models that fall under the "latent attrition" umbrella, but they all share this idea:

The customer's relationship with a firm has two phases: he is <u>alive</u> for some period of time, then becomes permanently inactive (<u>dead</u>)

Which customers remain alive until the end of the period?

Model development (in discrete time)

1. A customer's relationship with a firm has two phases: they are alive (A) and dead (D)

2. While alive, a customer makes a purchase with probability p each period. (While dead, a customer cannot make any purchase.)

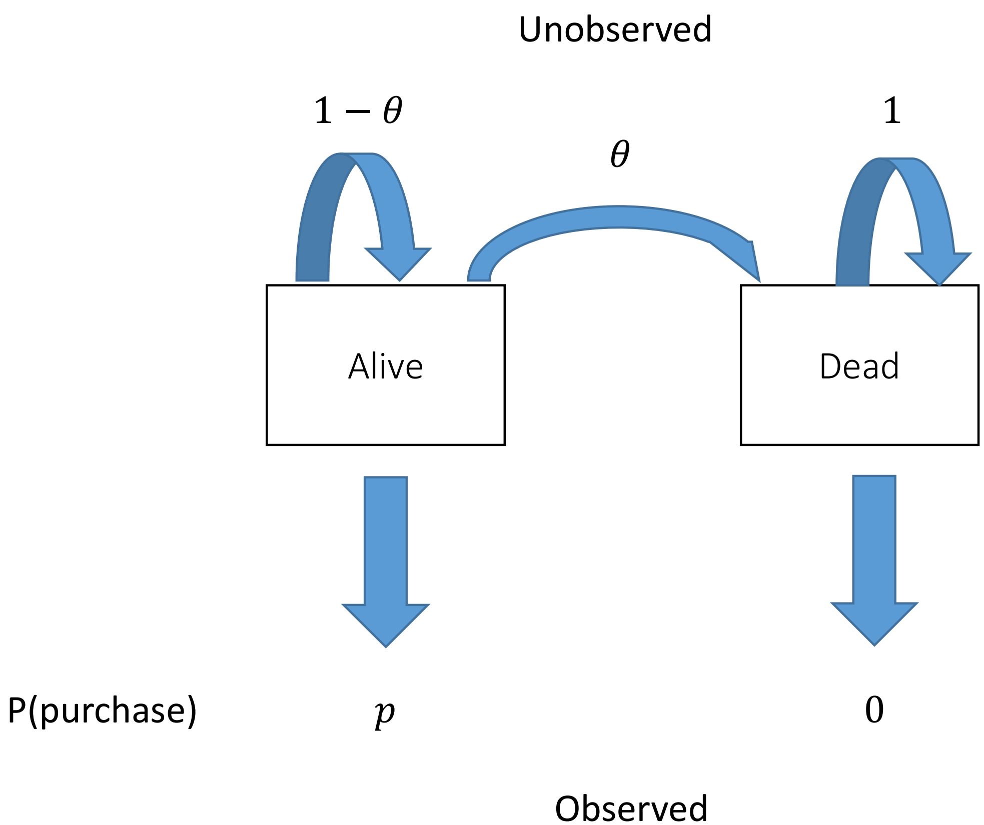
$$P(Y(t) = 1 | p, \text{alive at } t) = p$$

3. A living customer dies at the beginning of each period with probability θ

$$P(\text{alive at } t | q) = P(AA \ldots A | q) = (1 - \theta)^t$$

Fader, Hardie and Shang (2010), Customer-Base Analysis in a Discrete Time Noncontractual Setting

Latent attrition as a hidden Markov model



Transition matrix

Customer state at t-1 \ Customer state at t	Alive	Dead
Alive	$1-\theta$	θ
Dead	0	1

Death is an absorbing state:
no transitions back to Alive.

Since we don't observe death, how do we infer it?

We use probability to describe the uncertainty about whether a customer is gone temporarily or forever

1	2	3	4	5	6
1	0	1	1	0	0

Sample history

We use probability to describe the uncertainty about whether a customer is gone temporarily or forever

1	2	3	4	5	6
1	0	1	1	0	0
A	A	A	A	D	D
A	A	A	A	A	D
A	A	A	A	A	A

Sample history

We use probability to describe the uncertainty about whether a customer is gone temporarily or forever

1	**2**	**3**	**4**	**5**	**6**
1	0	1	1	0	0
A	A	A	A	D	D
A	A	A	A	A	D
A	A	A	A	A	A

$$
\begin{aligned}
L(101100|p,\theta) = {} & P(101100|p, AAAADD)\ P(AAAADD|\theta) \\
& + P(101100|p, AAAAAD)\ P(AAAAAD|\theta) \\
& + P(101100|p, AAAAAA)\ P(AAAAAA|\theta)
\end{aligned}
$$

Sample history

We use probability to describe the uncertainty about whether a customer is gone temporarily or forever

1	**2**	**3**	**4**	**5**	**6**
1	0	1	1	0	0
A	A	A	A	D	D
A	A	A	A	A	D
A	A	A	A	A	A

$$
\begin{aligned}
&= p(1-p)pp\,(1-\theta)^4\,\theta \\
&+ p(1-p)pp(1-p)\,(1-\theta)^5\,\theta \\
&+ p(1-p)pp(1-p)(1-p)\,(1-\theta)^6
\end{aligned}
$$

$$
\begin{aligned}
L(101100|p,\theta) &= P(101100|p,AAAADD)\;P(AAAADD|\theta) \\
&+ P(101100|p,AAAAAD)\;P(AAAAAD|\theta) \\
&+ P(101100|p,AAAAAA)\;P(AAAAAA|\theta)
\end{aligned}
$$

Sample history

We use probability to describe the uncertainty about whether a customer is gone temporarily or forever

1	2	3	4	5	6
1	0	1	1	0	0
A	A	A	A	D	D
A	A	A	A	A	D
A	A	A	A	A	A

$$\begin{aligned} &= p(1-p)pp\,(1-\theta)^4\,\theta \\ &+p(1-p)pp(1-p)\,(1-\theta)^5\,\theta \\ &+p(1-p)pp(1-p)(1-p)\,(1-\theta)^6 \end{aligned}$$

$$\begin{aligned} &= p^3(1-p)\,(1-\theta)^4\,\theta \\ &+p^3(1-p)^2(1-\theta)^5\,\theta \\ &+p^3(1-p)^3(1-\theta)^6 \end{aligned}$$

$$\begin{aligned} L(101100|p,\theta) &= P(101100|p, AAAADD)\ P(AAAADD|\theta) \\ &+P(101100|p, AAAAAD)\ P(AAAAAD|\theta) \\ &+P(101100|p, AAAAAA)\ P(AAAAAA|\theta) \end{aligned}$$

BG/BB Model: individual level likelihood

$$L(p,\theta|x,t_x,n) = p^x(1-p)^{n-x}(1-\theta)^n$$
$$+\sum_{j=0}^{n-t_x-1} p^x(1-p)^{t_x-x+j}\theta(1-\theta)^{t_x+j}$$

where

x = # of repeat purchases (frequency)
n = # of repeat purchase opportunities (periods)
t_x = period of last repeat purchase (recency*)

* Usually recency is $n-t_x$. If $x=0$, then $t_x=0$.

Fader, Hardie and Shang (2010), Customer-Base Analysis in a Discrete Time Noncontractual Setting

P(Alive)

There are may be several paths that account for a given customer's purchase history

The probability that he/she is alive is the probability of that path relative to all paths:

P(Alive) = P(Alive path) / P(All paths)

$$P(\text{alive at } n \mid x, t_x, n) = \frac{p^x (1-p)^{n-x} (1-\theta)^n}{L(p, \theta \mid x, t_x, n)}$$

Model development (heterogeneity)

4. p is distributed beta in population with parameters α and β :

$$f(p|\alpha,\beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}$$

5. θ is distributed beta as well, independently, with other parameters γ and δ

$$g(\theta|\gamma,\delta) = \frac{\theta^{\gamma-1}(1-\theta)^{\delta-1}}{B(\gamma,\delta)}$$

6. The transaction probability p and the dropout probability θ vary **independently** across customers

Fader, Hardie and Shang (2010), Customer-Base Analysis in a Discrete Time Noncontractual Setting

Beta-geometric/Beta-binomial (BGBB) model likelihood

Integrate the likelihood over the **2** prior distributions for p and θ

$$L(\alpha, \beta, \gamma, \delta | x, t_x, n) = \int_0^1 \int_0^1 L(p, \theta | x, t_x, n)\, f(p|\alpha, \beta)\, g(\theta|\gamma, \delta)\, dp\, d\theta$$

$$= \frac{B(\alpha + x, \beta + n - x)}{B(\alpha, \beta)} \frac{B(\gamma, \delta + n)}{B(\gamma, \delta)}$$

$$+ \sum_{j=0}^{n-t_x-1} \frac{B(\alpha + x, \beta + t_x - x + j)}{B(\alpha, \beta)} \frac{B(\gamma + 1, \delta + t_x + j)}{B(\gamma, \delta)}$$

Maximize log-likelihood to get parameter estimates $\{\alpha, \beta, \gamma, \delta\}$

Sufficient statistics

- Sufficient statistics are **recency** (t_x time of last transaction*) and **frequency** (x number of transactions) or **RF** of **RFM,** and number of transaction opportunities (n).

- The order of a given number of transactions prior to the last observed transaction doesn't matter.

- Sufficient statistics let us reduce the dimensionality of our data set

* Not exactly as in RFM, as we mentioned in slide 15

An example

Donations to a charity, of a group customers who made their first donation in 1995 (n=11104)

Over 1996-2001, cohort makes 24,615 repeat donations.

How many do they make in 2002-2006?

First purchase (all 1's)

Next six years of donation data we use to fit model

Following 5 years: we use to validate model

ID	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
100001	1	0	0	0	0	0	0	?	?	?	?	?
100002	1	0	0	0	0	0	0	?	?	?	?	?
100003	1	0	0	0	0	0	0	?	?	?	?	?
100004	1	0	1	0	1	1	1	?	?	?	?	?
100005	1	0	1	1	1	0	1	?	?	?	?	?
100006	1	1	1	1	0	1	0	?	?	?	?	?
100007	1	1	0	1	0	1	0	?	?	?	?	?
100008	1	1	1	1	1	1	1	?	?	?	?	?
100009	1	1	1	1	1	1	0	?	?	?	?	?
100010	1	0	0	0	0	0	0	?	?	?	?	?
⋮			⋮			⋮			⋮			⋮
111102	1	1	1	1	1	1	1	?	?	?	?	?
111103	1	0	1	1	0	1	1	?	?	?	?	?
111104	1	0	0	0	0	0	0	?	?	?	?	?

Sufficient statistics

	1996	1997	1998	1999	2000	2001
1	1	1	1	1	1	1
2	0	0	0	0	0	0
3	0	0	0	0	1	1
3	1	0	0	0	0	1

$\{x, t_x, n\}$

1 = {6, 6, 6}

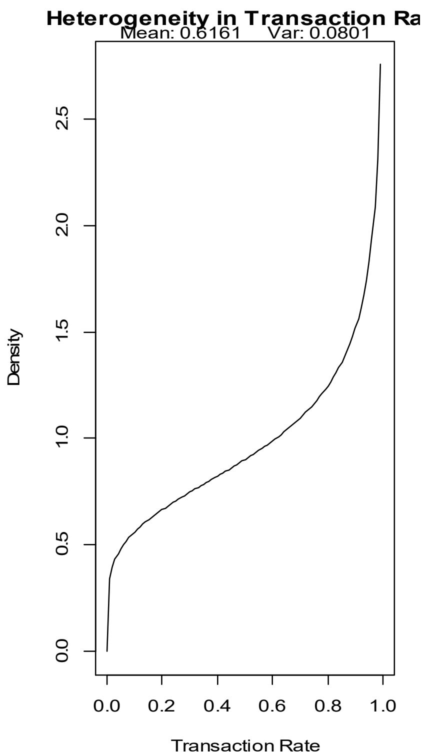
2 = {0, 0, 6}

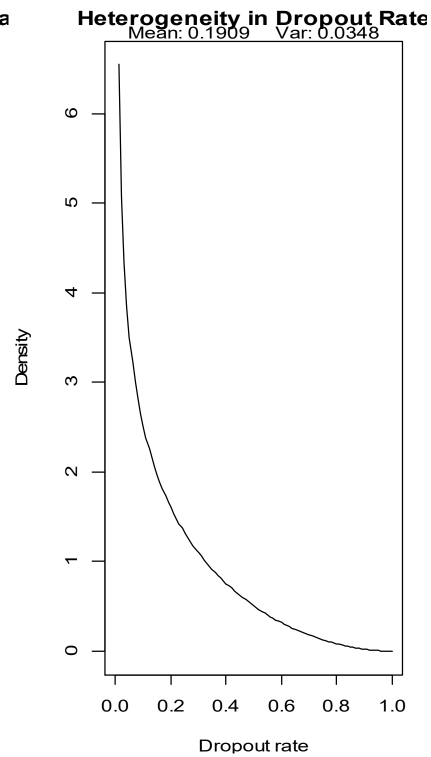
3 = {2, 6, 6}

	x (frequency)	t_x (recency)	n (years)	# donors
1	6	6	6	1203
	5	6	6	728
	4	6	6	512
	3	6	6	357
3	2	6	6	234
	1	6	6	129
	5	5	6	335
	4	5	6	284
	3	5	6	225
	2	5	6	173
	1	5	6	119
	4	4	6	240
	3	4	6	181
	2	4	6	155
	1	4	6	78
	3	3	6	322
	2	3	6	255
	1	3	6	129
	2	2	6	613
	1	2	6	277
	1	1	6	1091
2	0	0	6	3464

Likelihood

Distribution of transaction and dropout probabilities in the customer base



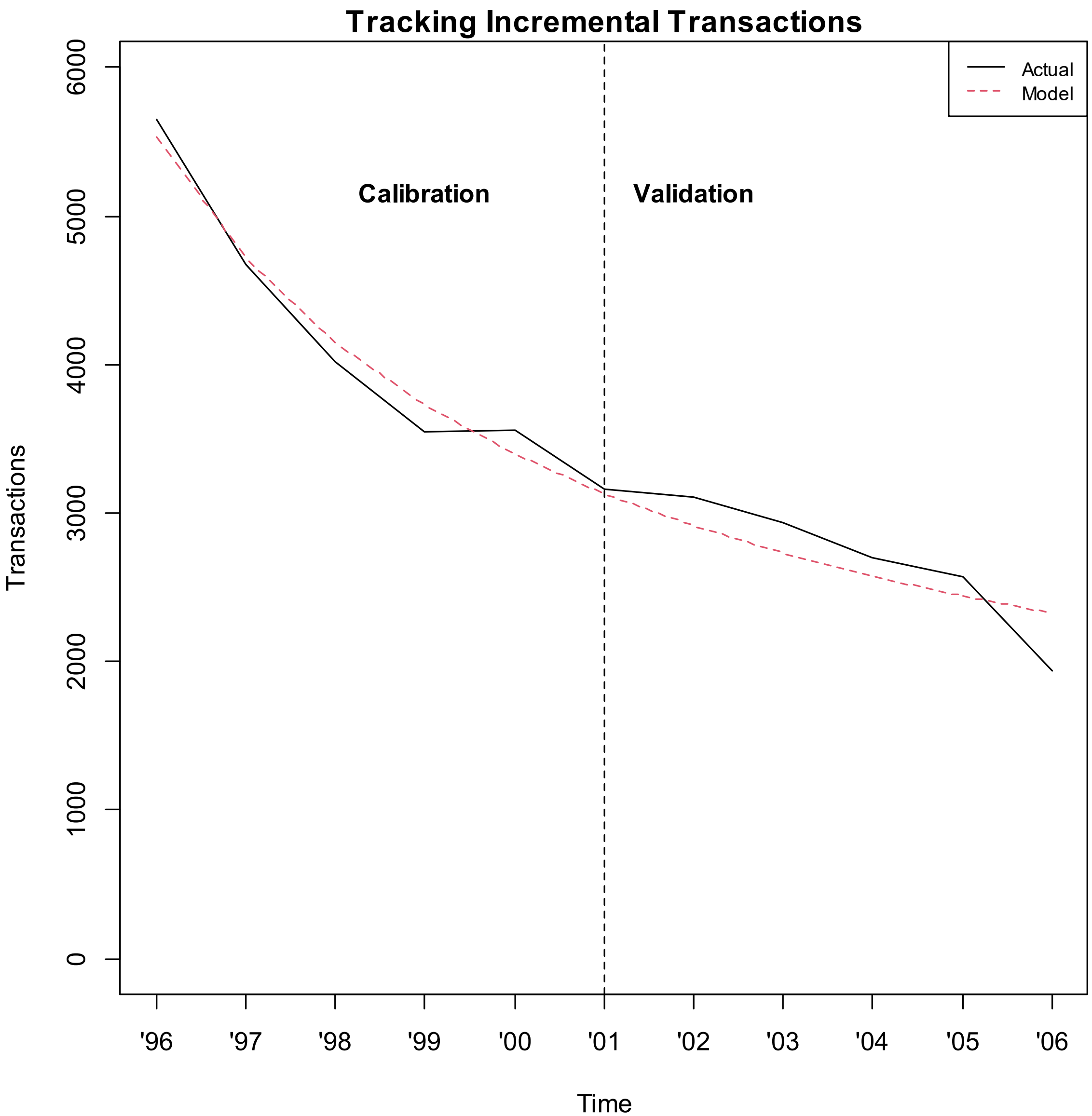


$\alpha = 1.204, \beta = 0.750$

$$E[p] = \frac{\alpha}{\alpha + \beta} = 0.616$$

$\gamma = 0.657, \delta = 2.783$

$$E[\theta] = \frac{\gamma}{\gamma + \delta} = 0.191$$



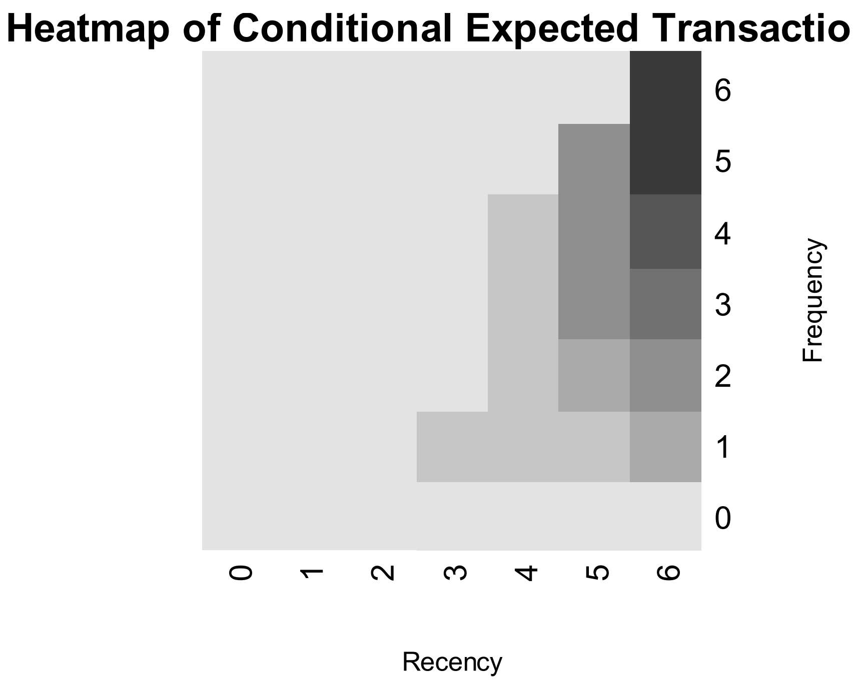
Aggregate forecasting

How many do they make in 2002-2006?

We can expect 13000 donations 2002-2006

Model predictions: Predictions by both R & F

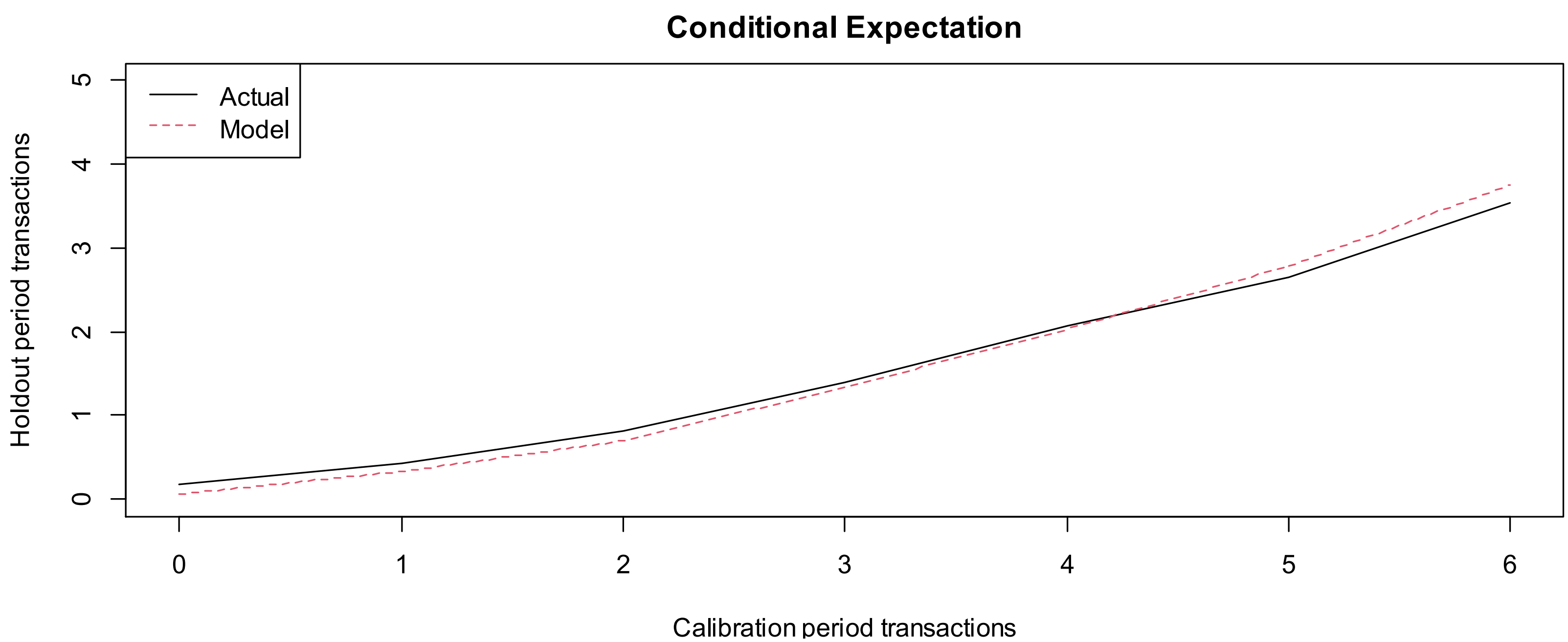
Using the model, expected number of transactions in the next 5 periods, conditional on R and F



\	0	1	2	3	4	5	6
6	0.00	0.00	0.00	0.00	0.00	0.00	3.75
5	0.00	0.00	0.00	0.00	0.00	1.81	3.23
4	0.00	0.00	0.00	0.00	0.58	2.03	2.71
3	0.00	0.00	0.00	0.22	1.03	1.80	2.19
2	0.00	0.00	0.12	0.54	1.06	1.44	1.67
1	0.00	0.09	0.31	0.59	0.84	1.02	1.15
0	0.07	0.00	0.00	0.00	0.00	0.00	0.00

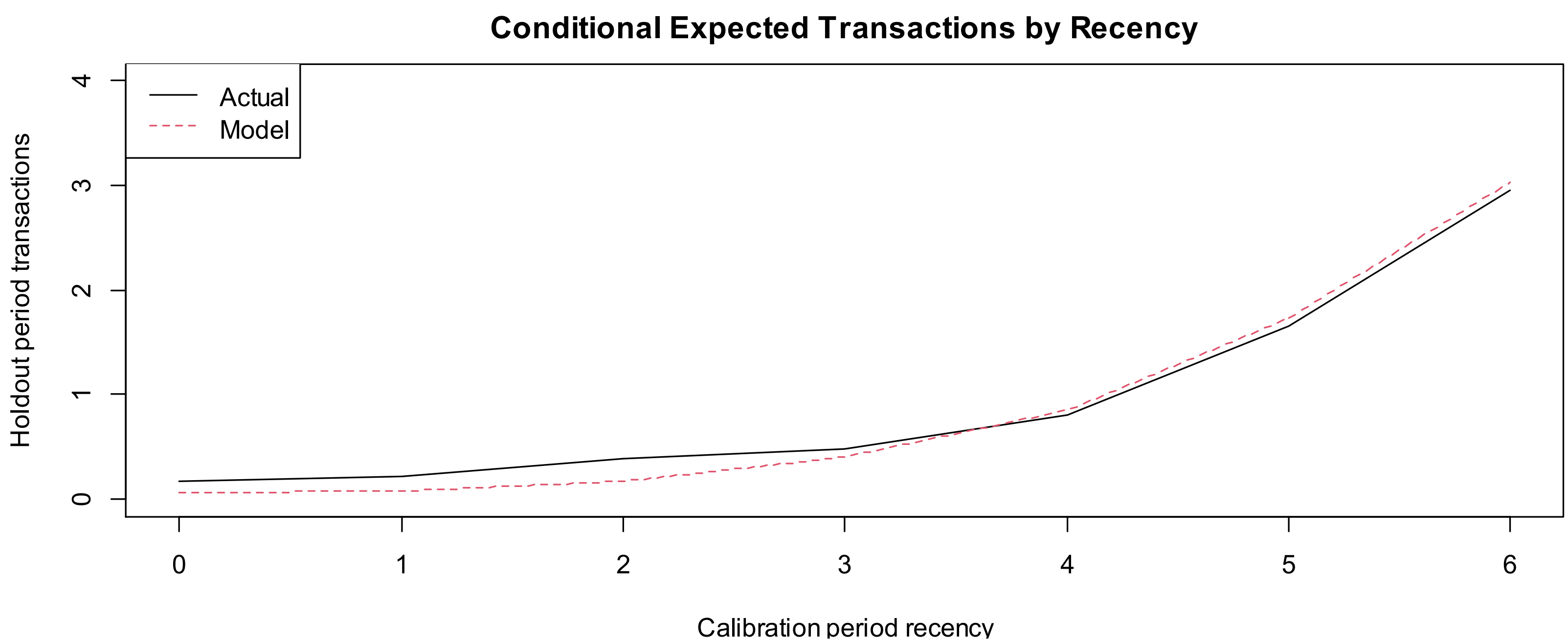
Model fit: Predictions by F groups

If we're interested in RLV, we should care about more individual level predictions: how many transactions from a customer with "6 for 6" do we predict? How close are we to the actual?



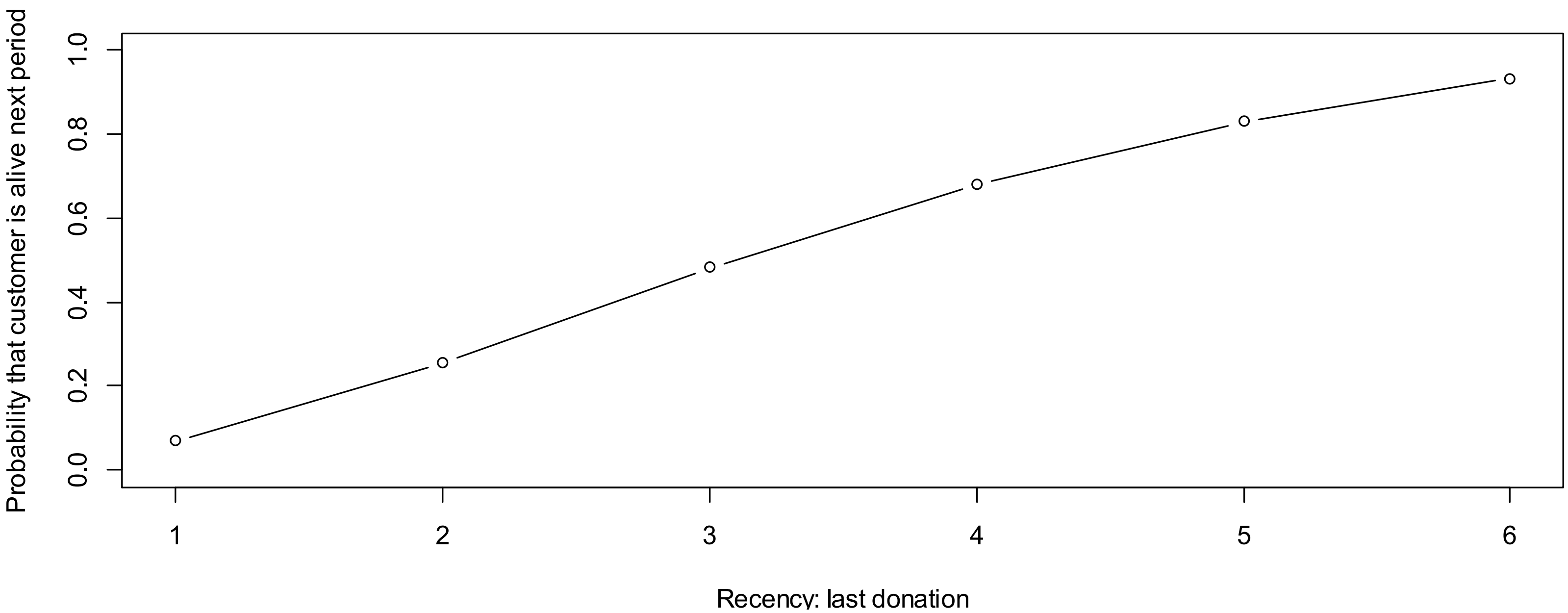
Model fit: Predictions by R groups

Same idea, but for recency: how much do we predict for a recent vs. not recent customer? How close is it to the truth?



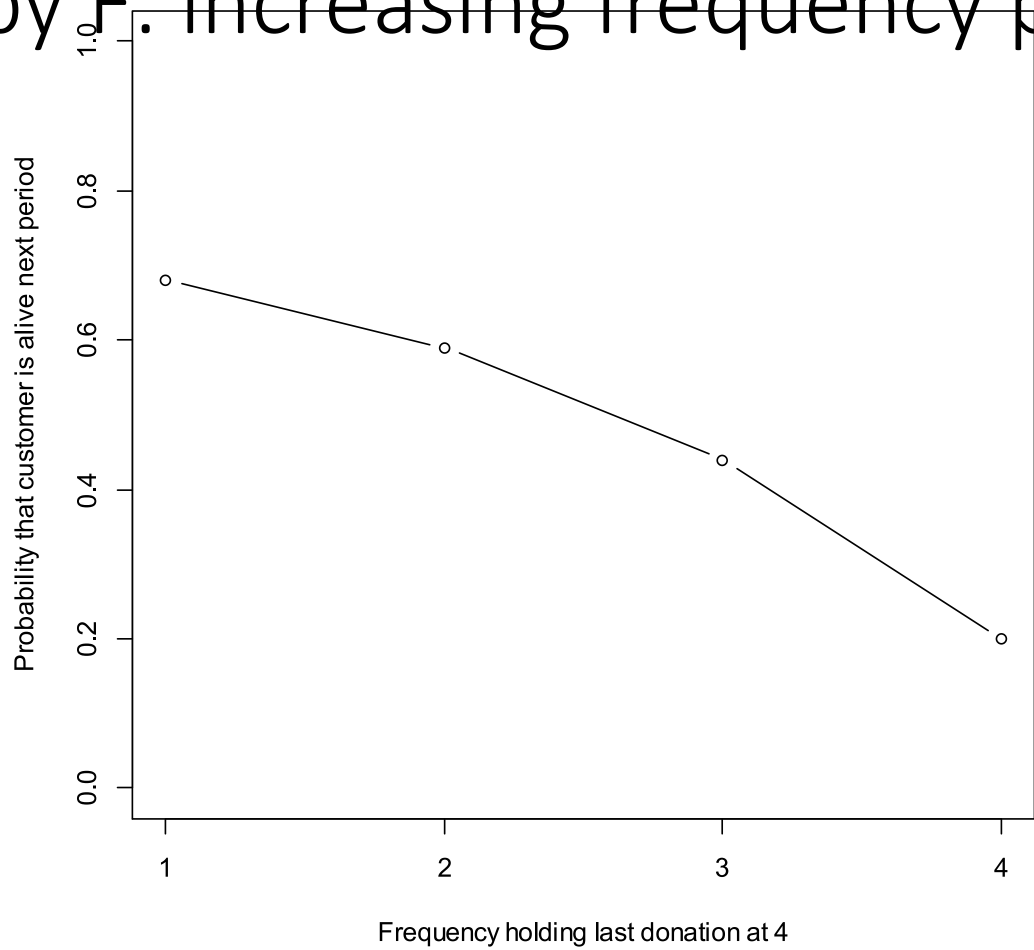
P(Alive at n+1) by R

The probability that a customer who was “1 for 6” with the single transaction occurring at 1, 2, .. 6 is alive at the next transaction



{x = 1, t_x =1, .. ,6, n}

P(Alive) by F: increasing frequency paradox



{x = 1,..,4, t_x =4 , n}

CLV

For a given individual, $P(Y(t) = 1|p, \theta) = p(1-\theta)^t$

$$P(Y(t) = 1|\alpha, \beta, \gamma, \delta) = \int_0^1 \int_0^1 P(Y(t) = 1|p, \theta)\, f(p|\alpha, \beta)\, g(\theta|\gamma, \delta)\, dp\, d\theta$$
$$= \left(\frac{\alpha}{\alpha + \beta}\right) \frac{B(\gamma, \delta + t)}{B(\gamma, \delta)}$$

CLV is the discounted sum of expected purchases (d discount rate and m margin per transaction):

$$E[\mathrm{CLV}] = m \left(1 + \sum_{t=1}^{\infty} \left(\frac{\alpha}{\alpha + \beta}\right) \frac{B(\gamma, \delta + t)}{B(\gamma, \delta)} \frac{1}{(1+d)^t}\right)$$

CLV=185, for m=50, d=.1

RLV

- For a customer, the Discounted Expected Residual Transactions (DERT) for someone with history {x, tx, n}

$$DERT = P(\text{alive at } n | x, t_x, n) \sum_{t=n+1}^{\infty} P(Y(t) = 1 | \text{alive at } t) \frac{P(\text{alive at } t)}{(1+d)^{t-n}}$$

-> integrate expression over joint posterior distribution of p and θ (Eq. 14)

Complicated result, but easy to calculate in R.

Fader, Hardie and Shang (2010), Customer-Base Analysis in a Discrete Time Noncontractual Setting

```
##        x t.x n.cal custs RLV
##   [1,] 6   6     6  1203 295.48
##   [2,] 5   6     6   728 254.46
##   [3,] 4   6     6   512 213.44
##   [4,] 3   6     6   357 172.42
##   [5,] 2   6     6   234 131.40
##   [6,] 1   6     6   129  90.38
##   [7,] 5   5     6   335 142.74
##   [8,] 4   5     6   284 159.84
##   [9,] 3   5     6   225 142.10
##  [10,] 2   5     6   173 113.62
##  [11,] 1   5     6   119  80.44
##  [12,] 4   4     6   240  45.92
##  [13,] 3   4     6   181  81.46
##  [14,] 2   4     6   155  83.27
##  [15,] 1   4     6    78  66.09
##  [16,] 3   3     6   322  17.60
##  [17,] 2   3     6   255  42.21
##  [18,] 1   3     6   129  46.76
##  [19,] 2   2     6   613   9.38
##  [20,] 1   2     6   277  24.74
##  [21,] 1   1     6  1091   6.75
##  [22,] 0   0     6  3464   5.74
```

m=50, d=.1

Discrete to continuous time

- Opportunities for transaction occur at discrete points in time
  - Concerts or conferences occur only one point in time
- Firms think of it as discrete:
  - Did the customer place an order in response to a quarterly mailing or campaign?
  - Next page
- Suppose we have a year of data from Amazon or Bol.
  - Should we define
    - 12 monthly transaction opportunities
    - 52 weekly
    - 365 daily

In response to a specific campaign

UNICEF helpt in een nieuw kamp de kinderen uit Moria

Bekijk hier de online versie

unicef
voor ieder kind

Beste meneer Knox,

Begin september verwoestte een brand kamp **Moria** op Lesbos. **12.000 kinderen en volwassenen raakten dakloos.** Inmiddels is op het eiland een nieuw kamp verrezen waarin de bewoners van Moria terechtkunnen.

Amir (10) heeft tien maanden in Moria gewoond. Hij had het er altijd koud en zijn vader stond eindeloos in de rij om voor zijn gezin eten, water of een douche te regelen. De jongen kan met

Classifying customer bases

Basic idea same: a purchase process,
a drop out process, heterogeneity

Opportunities for Transactions

	Non-contractual	**Contractual**
Continuous	• Grocery purchasing • Hotel stays • Prepaid cell phone	• Credit cards • Utilities • Airmiles, loyalty programs
Discrete	• Charity fund drives • Concert/conference attendance JUST NOW	• Media subscription • Insurance policies LAST LECTURE

Type of business setting

Pareto/NBD

1. While active, the number of transactions made by a customer in a time period of length t makes follows a Poisson distribution with mean λt.

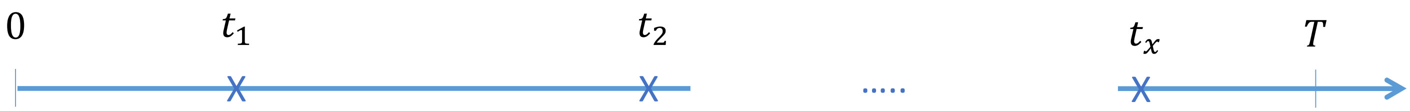
   → *continuous time version of binomial*

2. Each customer has an unobserved lifetime of length τ (after which he or she becomes inactive), which is distributed exponentially with dropout rate μ.

   → *continuous time version of geometric*

Schmittlein, Morrison and Colombo (1987); Fader, Hardie and Lee (2005)

Pareto/NBD



- The first purchase is at time 0.

- Each of a customers x repeat transactions occurred during the period $(0, T]$.

- Two possibilities:
  - The customer is still alive at T
  - The customer died sometime in $(t_x, T]$

Pareto/NBD

3. Heterogeneity in transaction rates across customers follows a *gamma* distribution with parameters r and α.

   → *similar to Beta, but range is* $[0, \infty)$

4. Heterogeneity in dropout rates across customers follows a *gamma* distribution with parameters s and β.

5. The transaction rate λ and dropout rate μ vary **independently** of across customers

Sufficient statistics and related model

As with BG/BB model, the <u>sufficient statistics</u> of the model are the number of repeat purchases (x), the recency of the last purchase (t_x) and the end of the observation period (T).

	x	t.x	T.cal
1	2	30.43	38.9
2	1	1.71	38.9
3	0	0.00	38.9
4	0	0.00	38.9
5	0	0.00	38.9
6	7	29.43	38.9

Here we are in weeks

BG/NBD

- The Pareto/NBD can be difficult to estimate: BG/NBD simpler model with similar behavioral story
- The BG/NBD, which assumes a discrete "death process" but a continuous transaction process. (Fader Hardie Lee 2005)
- 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 probability:

$$P\left(\text{inactive following } j^{th} \text{ purch}\right) = p(1-p)^{j-1} \qquad j = 1, 2, \ldots$$

Latent attrition models

Dynamics

- Customers go through two stages in their “lifetime” with a specific firm: they are alive for some period of time, then become permanently dead.

Heterogeneity

- Customers differ in their transaction/purchase rates and in their death rates

Customer-base analysis

- When you have a database of customers, generally you want to determine:
  - Which customers are most likely to be active in the future
  - The number of transactions to expect in the future for them, individually and collectively
  - CLV/RLV
- Note these are predictive, not descriptive statistics

Advantages relative to RFM and other next-period models

With the next-period models we discussed in class, it’s hard to make predictions beyond the next period, so CLV calculation is tough.

- How do you make predictions for the period after next, and so on.

RFM and similar models don’t really have a customer “story” or theory about them.

Different slices of the data yield different values

Extensions

- Incorporating other characteristics, marketing activities, competition
    - Complaints and recoveries (Knox and van Oest 2014)
    - Service quality (Braun, Schweidel and Stein 2015)
    - Direct marketing (Schweidel and Knox 2013)
    - Notes on extensions from Bruce Hardie's website
    - Contextual variables (Bachmann, Meierer & Näf 2021), check out their R package CLVTools
  - Have to be careful about sample selection (from e.g., targeting) and other issues
  - Need full datasets, no more sufficient statistics

- Relaxing independence of distributions of purchasing and dropout
  - Extensions allowing for correlation (Abe 2008)
  - Requires simulation