What’s wrong with this story of customer contract-renewal behavior?

Developing a Better Model (II)

Consider the following story of customer behavior:

i) At the end of each period, an individual renews his contract with (constant and unobserved) probability $1 - \theta$.

ii) "Churn probabilities" vary across customers.

- Since we don't know any given customer's true value of θ, we treat it as a realization of a random variable (Θ).
- We need to specify a probability distribution that captures how θ varies across customers (by giving us the probability of each possible value of θ).

Developing a Better Model (II)

What is the probability that a randomly chosen new customer will cancel their contract at the end of period t?

i) If we knew their θ, it would simply be

$$P(T = t \mid \theta) = \theta(1 - \theta)^{t-1}.$$

ii) Since we only know the distribution of Θ across the population, we compute

$$P(T = t) = E_{\Theta}\big[P(T = t \mid \theta)\big],$$

i.e., we evaluate $P(T = t \mid \theta)$ for each possible value of θ, weighting it by the probability of a randomly chosen new customer having that value of θ.

Accounting for Heterogeneity (I)

- Suppose we have two (unobserved) segments:

$$\Theta = \begin{cases} \theta_1 & \text{with probability } \pi \\ \theta_2 & \text{with probability } 1 - \pi \end{cases}$$

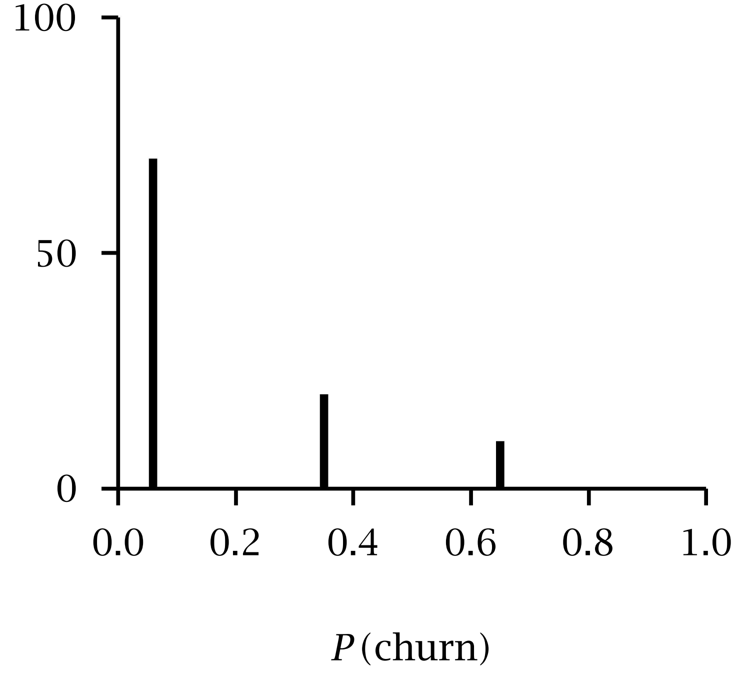
- We compute $E_\Theta\big[P(T = t \mid \theta)\big]$:

$$\begin{aligned} P(T = t \mid \theta_1, \theta_2, \pi) & \\ &= P(T = t \mid \Theta = \theta_1)P(\Theta = \theta_1) \\ &\qquad + P(T = t \mid \Theta = \theta_2)P(\Theta = \theta_2) \\ &= \theta_1(1 - \theta_1)^{t-1}\pi + \theta_2(1 - \theta_2)^{t-1}(1 - \pi) \end{aligned}$$

- Likewise for three or four segments ...

Vodafone Italia
Churn Clusters

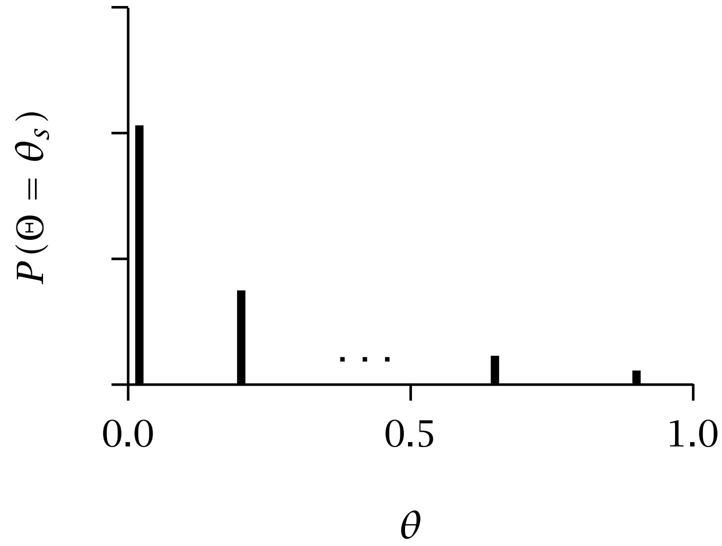
Cluster	P(churn)	%CB
Low risk	0.06	70
Medium risk	0.35	20
High risk	0.65	10



Source: “Vodafone Achievement and Challenges in Italy” presentation (2003-09-12)

As the Number of Segments $\longrightarrow \infty$

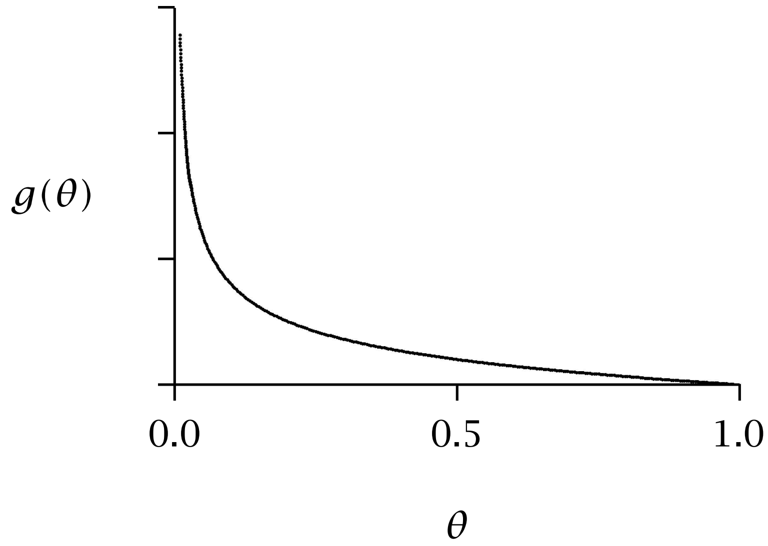
Discrete



$$\sum_{s=1}^{S} P(\Theta = \theta_s) = 1$$

$2S - 1$ parameters

Continuous



$$\int_0^1 g(\theta) d\theta = 1$$

k parameters

Accounting for Heterogeneity (II)

- We move from a finite number of segments (a *finite mixture* model) to an infinite number of segments (a *continuous mixture* model).

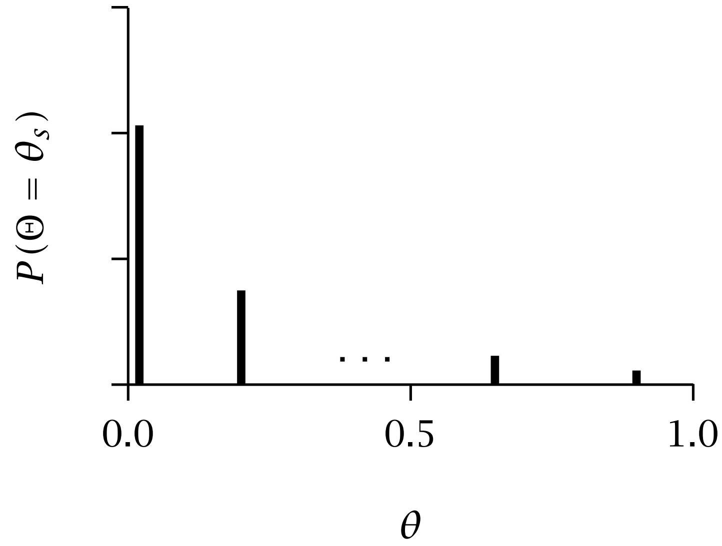
- We choose a continuous distribution for Θ, with probability density function (pdf) $g(\theta \mid \text{parameters})$.

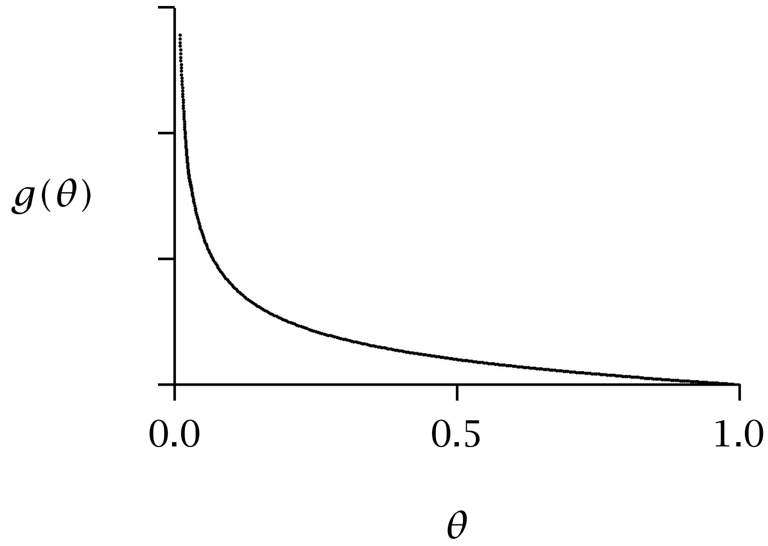
- We compute $E_{\Theta}[P(T = t \mid \theta)]$:

$$P(T = t \mid \text{parameters})$$

$$= \int_0^1 P(T = t \mid \Theta = \theta) g(\theta \mid \text{parameters})\, d\theta\,.$$

Accounting for Heterogeneity (II)





$$\sum_{s=1}^{S} P(\Theta = \theta_s) = 1 \qquad \int_0^1 g(\theta) d\theta = 1$$

$$P(T = t) = E_\Theta[P(T = t \mid \theta)]$$

$$\sum_{s=1}^{S} P(T = t \mid \Theta = \theta_s) P(\Theta = \theta_s) \qquad \int_0^1 P(T = t \mid \Theta = \theta) g(\theta)\, d\theta$$

The Beta Distribution

- The beta distribution is a flexible (and mathematically convenient) two-parameter distribution bounded between 0 and 1:

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

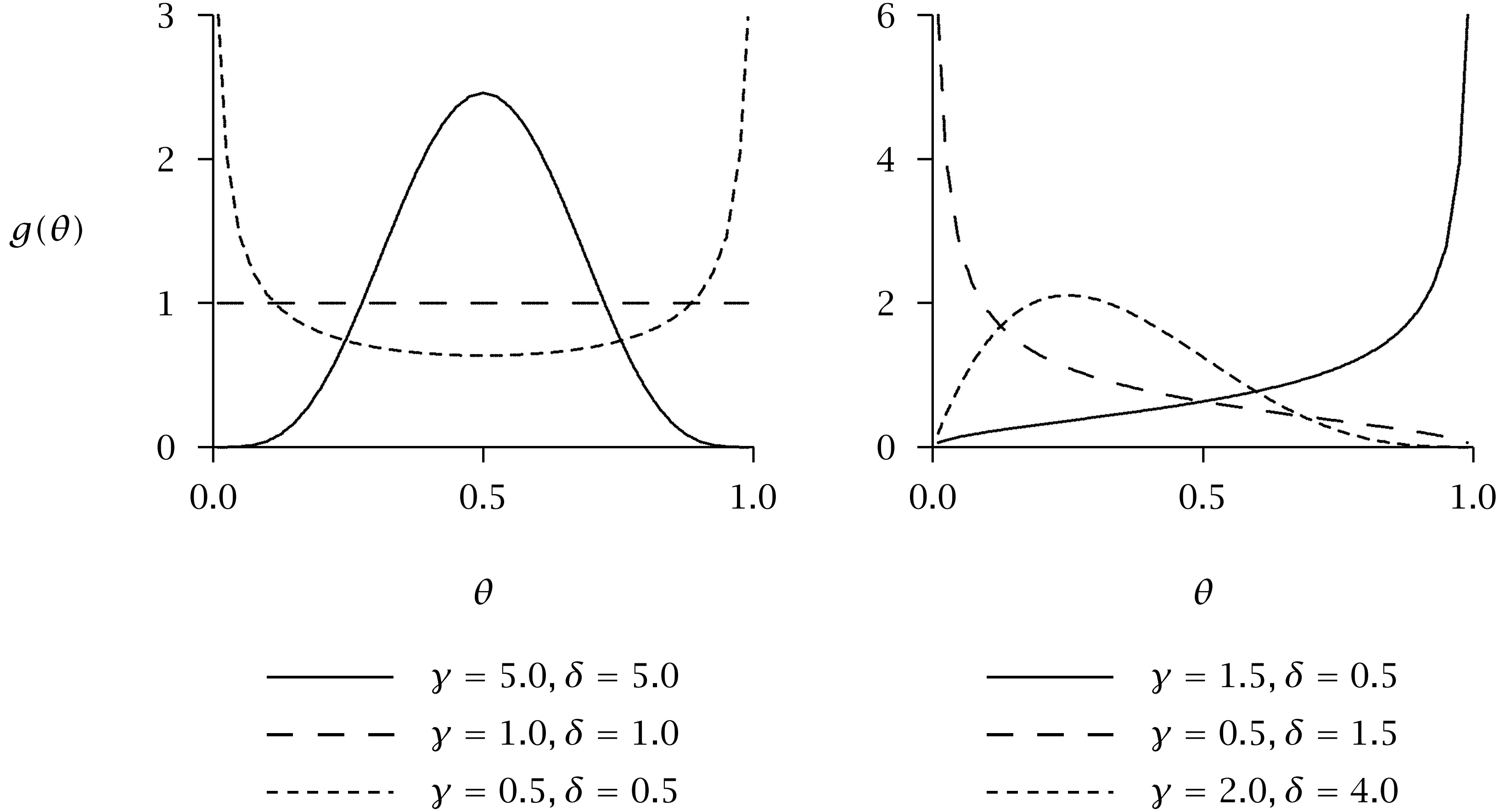
where $\gamma, \delta > 0$ and $B(\gamma, \delta)$ is the beta function.

- The mean and variance are given by

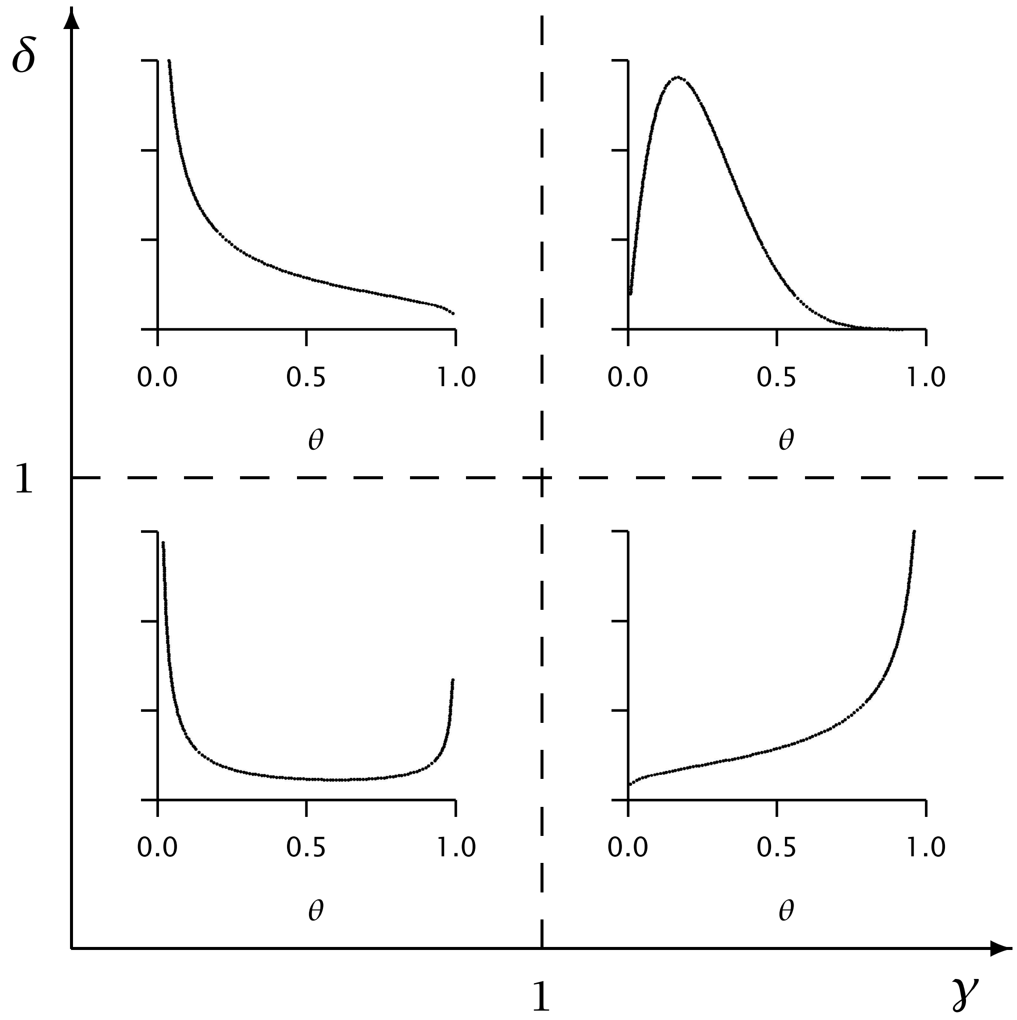
$$E(\Theta) = \frac{\gamma}{\gamma + \delta}$$

$$\mathrm{var}(\Theta) = \frac{\gamma\delta}{(\gamma + \delta)^2(\gamma + \delta + 1)}$$

Illustrative Beta Distributions



Five General Shapes of the Beta Distribution



The Beta Function

- The beta function $B(\gamma, \delta)$ is defined by the integral

$$B(\gamma, \delta) = \int_0^1 t^{\gamma-1}(1-t)^{\delta-1}dt, \; \gamma > 0, \delta > 0\,,$$

  and can be expressed in terms of gamma functions:

$$B(\gamma, \delta) = \frac{\Gamma(\gamma)\Gamma(\delta)}{\Gamma(\gamma+\delta)}\,.$$

- The gamma function $\Gamma(\gamma)$ is a generalized factorial, which has the recursive property $\Gamma(\gamma+1) = \gamma\Gamma(\gamma)$. Since $\Gamma(0) = 1$, $\Gamma(n) = (n-1)!$ for positive integer n.

Numerical Evaluation of the Beta Function

- Not all computing environments have a `beta` function (or even a `gamma` function).
- However, we typically have a function that evaluates $\ln\left(\Gamma(\cdot)\right)$, e.g., `gammaln`.
- In Excel,

$$\begin{aligned}\Gamma(\gamma) &= \texttt{exp(gammaln(}\gamma\texttt{))}\\ B(\gamma,\delta) &= \texttt{exp(gammaln(}\gamma\texttt{)+gammaln(}\delta\texttt{)}\\ &\qquad\qquad\texttt{-gammaln(}\gamma\texttt{+}\delta\texttt{))}\end{aligned}$$

Developing a Better Model (II)

For a randomly chosen individual,

$$
\begin{aligned}
P(T = t \mid \gamma, \delta) &= \int_0^1 P(T = t \mid \theta) g(\theta \mid \gamma, \delta)\, d\theta \\
&= \int_0^1 \theta (1 - \theta)^{t-1} \frac{\theta^{\gamma-1}(1-\theta)^{\delta-1}}{B(\gamma, \delta)}\, d\theta \\
&= \frac{1}{B(\gamma, \delta)} \int_0^1 \theta^{\gamma}(1-\theta)^{\delta+t-2}\, d\theta \\
&= \frac{B(\gamma+1, \delta+t-1)}{B(\gamma, \delta)} \,.
\end{aligned}
$$

Developing a Better Model (II)

Similarly,

$$\begin{aligned}
S(t \mid \gamma, \delta) &= \int_0^1 S(t \mid \theta) g(\theta \mid \gamma, \delta)\, d\theta \\
&= \int_0^1 (1-\theta)^t \frac{\theta^{\gamma-1}(1-\theta)^{\delta-1}}{B(\gamma, \delta)}\, d\theta \\
&= \frac{1}{B(\gamma, \delta)} \int_0^1 \theta^{\gamma-1}(1-\theta)^{\delta+t-1}\, d\theta \\
&= \frac{B(\gamma, \delta+t)}{B(\gamma, \delta)} .
\end{aligned}$$

We call this *continuous mixture* model the beta-geometric (BG) distribution.

Developing a Better Model (II)

We can compute BG probabilities using the following forward-recursion formula from $P(T = 1)$:

$$P(T = t) = \begin{cases} \dfrac{\gamma}{\gamma + \delta} & t = 1 \\[2ex] \dfrac{\delta + t - 2}{\gamma + \delta + t - 1} \times P(T = t - 1) & t = 2, 3, \ldots \end{cases}$$

Estimating Model Parameters

Assuming

i) the observed data were generated according to the heterogeneous "coin flipping" story of contract renewal, and

ii) we know γ and δ,

the probability of the observed pattern of renewals is:

$$[P(T=1\mid\gamma,\delta)]^{369}[P(T=2\mid\gamma,\delta)]^{163}[P(T=3\mid\gamma,\delta)]^{86}$$
$$\times[P(T=4\mid\gamma,\delta)]^{56}[S(4\mid\gamma,\delta)]^{326}$$

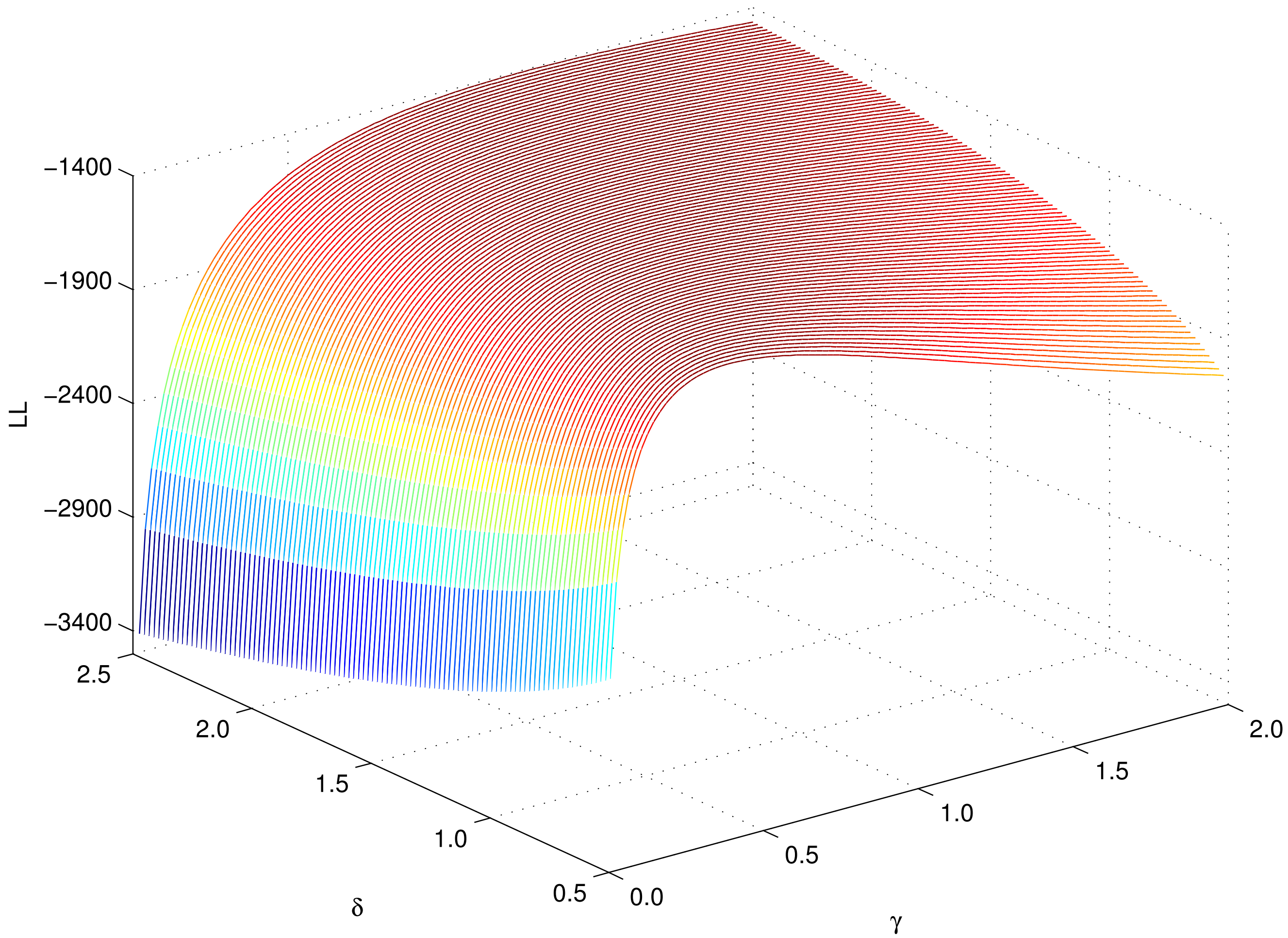
Estimating Model Parameters

The log-likelihood function is given by:

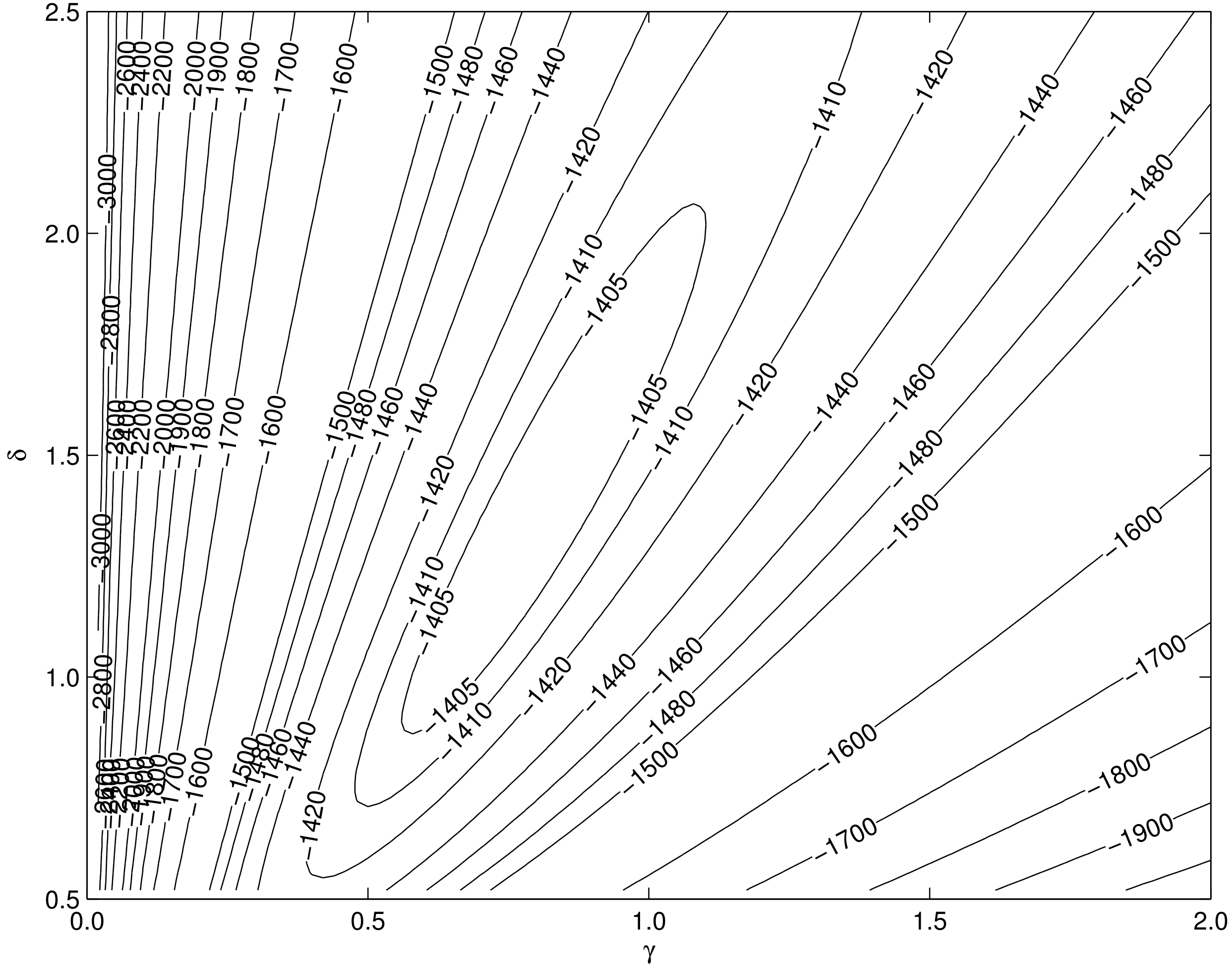
$$\begin{aligned} LL(\gamma, \delta \mid \text{data}) = {} & 369 \times \ln[P(T = 1 \mid \gamma, \delta)] + \\ & 163 \times \ln[P(T = 2 \mid \gamma, \delta)] + \\ & 86 \times \ln[P(T = 3 \mid \gamma, \delta)] + \\ & 56 \times \ln[P(T = 4 \mid \gamma, \delta)] + \\ & 326 \times \ln[S(4 \mid \gamma, \delta)] \end{aligned}$$

The maximum value of the log-likelihood function is $LL = -1401.6$, which occurs at $\hat{\gamma} = 0.764$ and $\hat{\delta} = 1.296$.

Surface Plot of BG LL Function



Contour Plot of BG LL Function



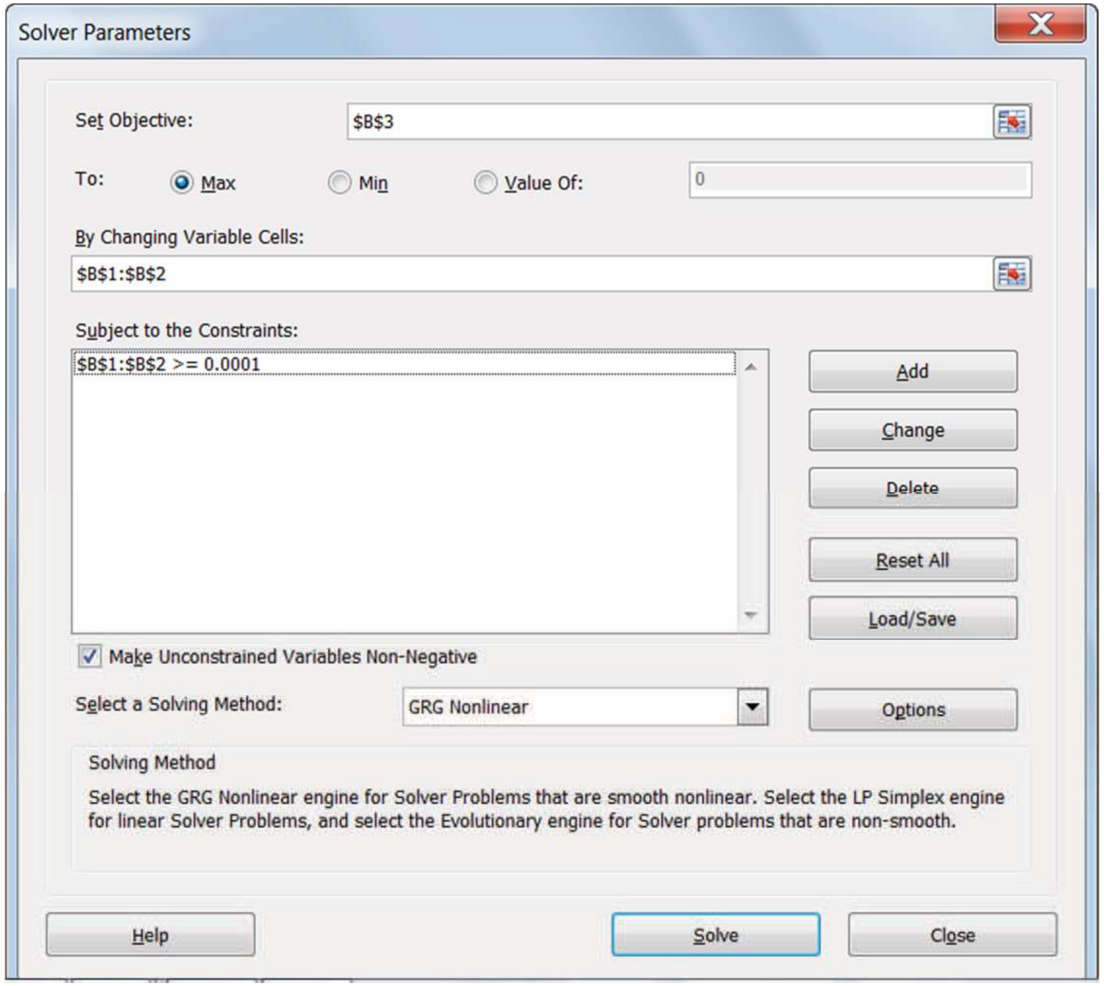
Estimating Model Parameters

	A	B	C	D	E	F
1	gamma	1.000				
2	delta	1.000				
3	LL	-1454.0				
4						
5	t	# Cust.	# Lost	P(die)	S(t)	
6	0	1000			1.0000	
7	1	[illegible]	[illegible]9	0.5000	0.5000	-255.77
8	2	468	163	0.1667	0.3333	-292.06
9	3	382	86	0.0833	0.2500	-213.70
10	[illegible]	[illegible]	[illegible]	[illegible]	[illegible].2000	-167.76
11						-524.68

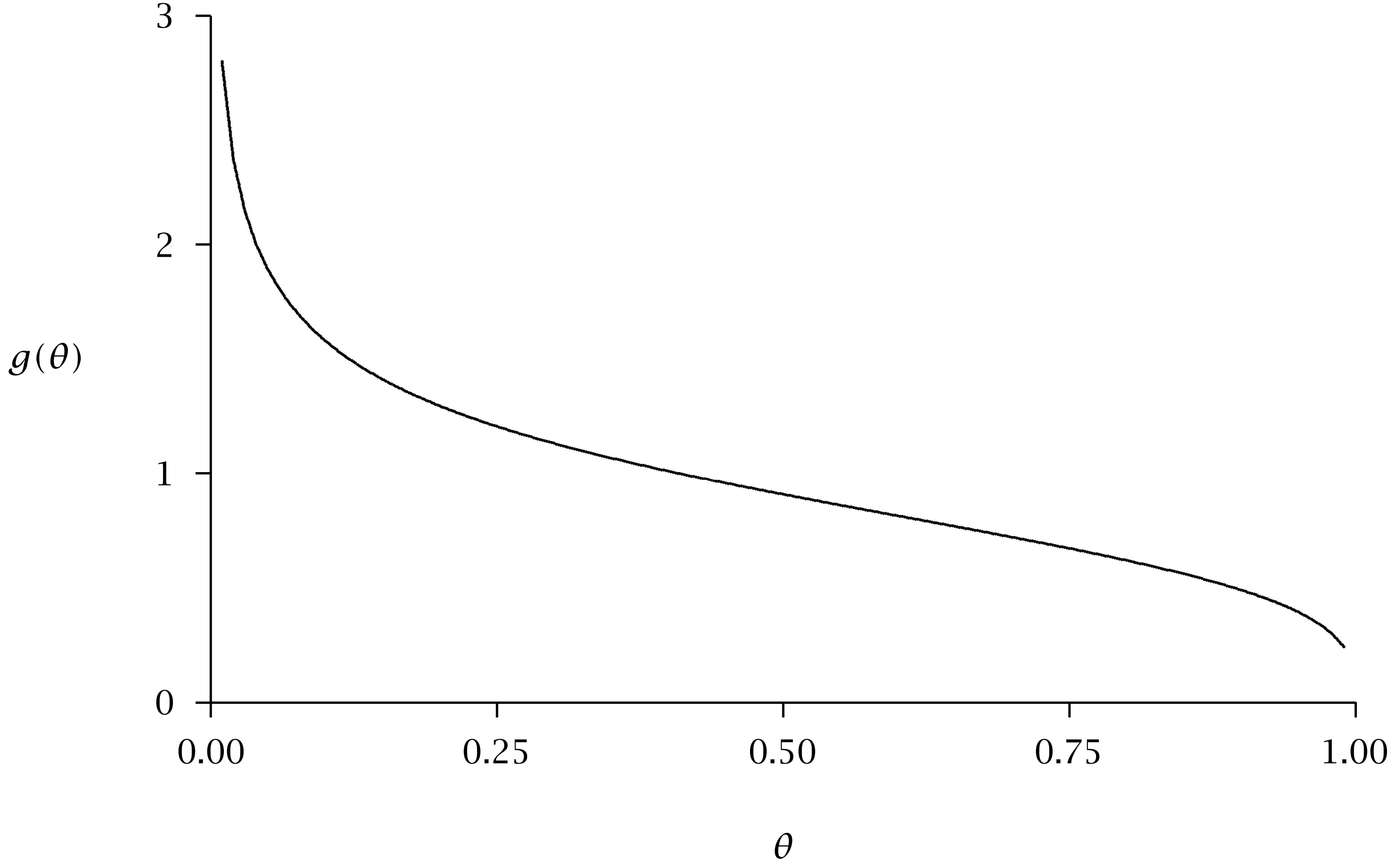
=B1/(B1+B2)

=D7*($B$2+A8-2)/($B$1+$B$2+A8-1)

Estimating Model Parameters

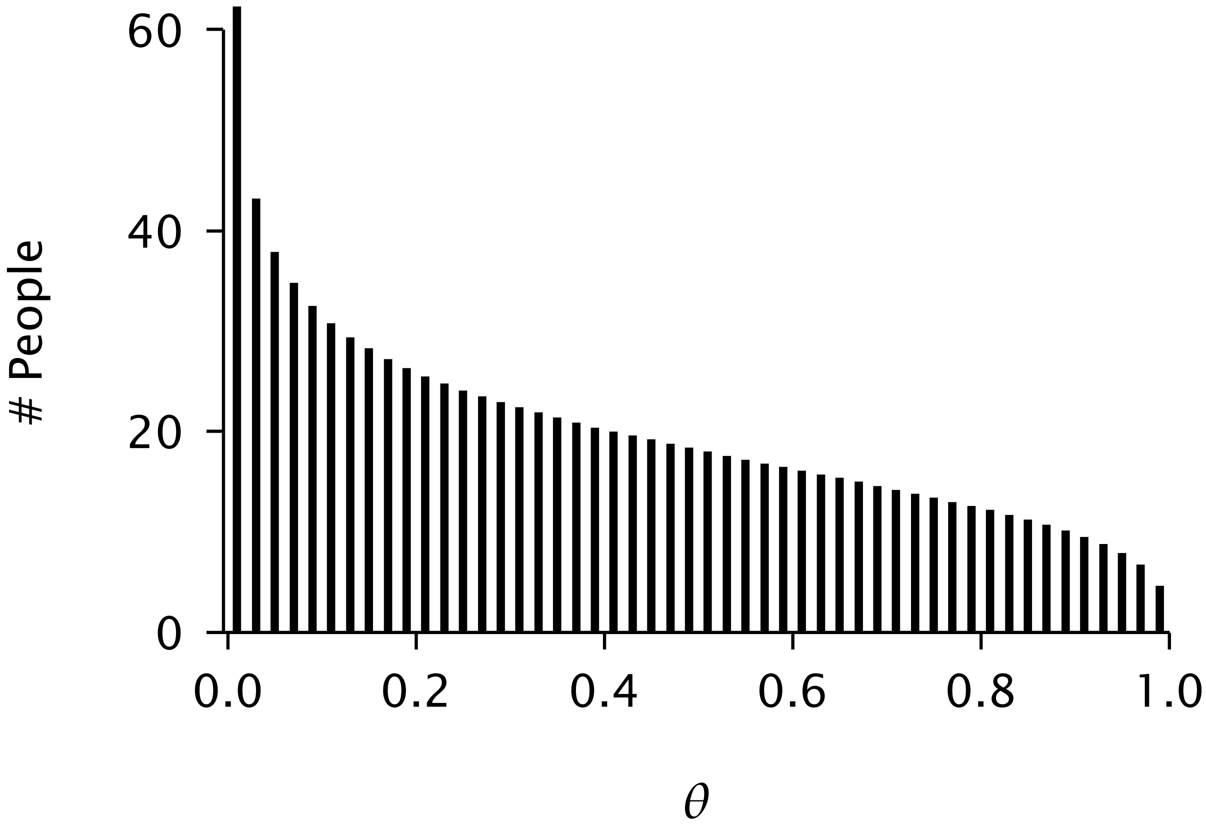


Estimated Distribution of Churn Probabilities



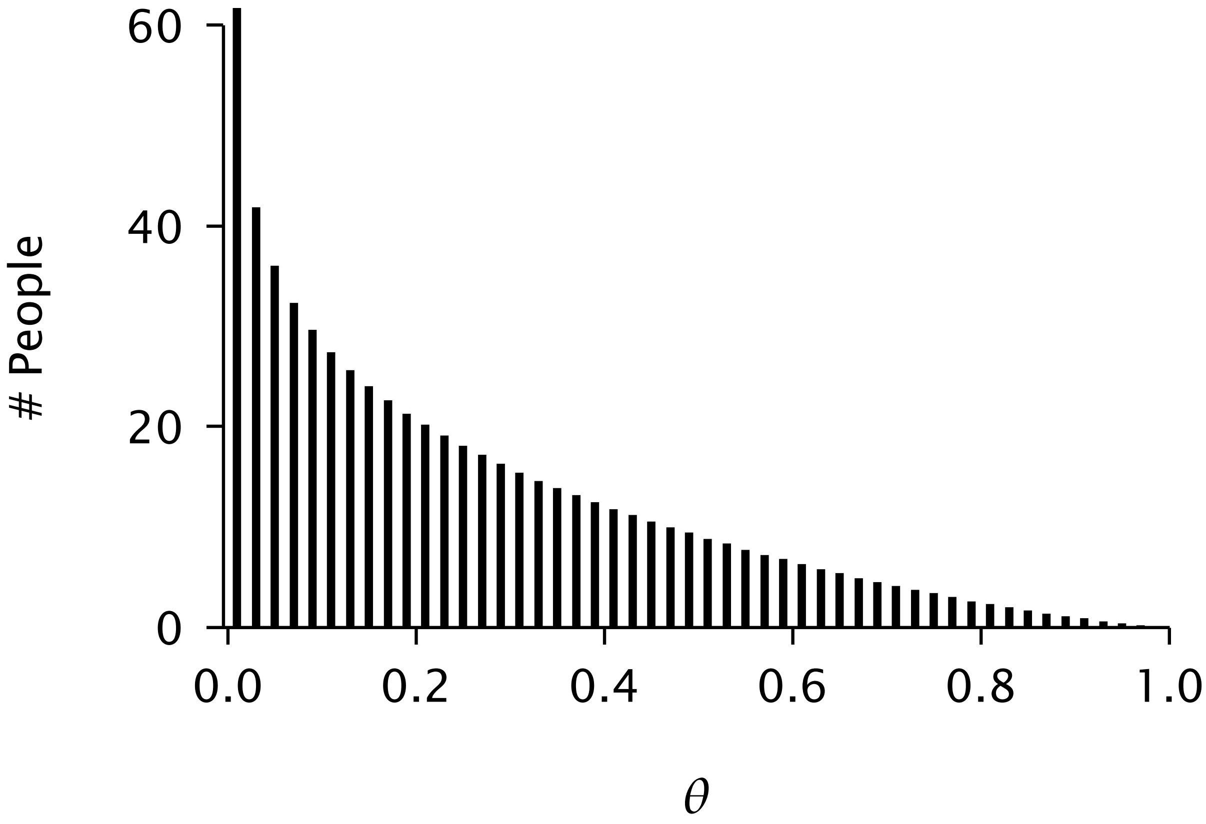
$$\hat{\gamma} = 0.764, \hat{\delta} = 1.296, \widehat{E(\Theta)} = 0.371$$

Year 1



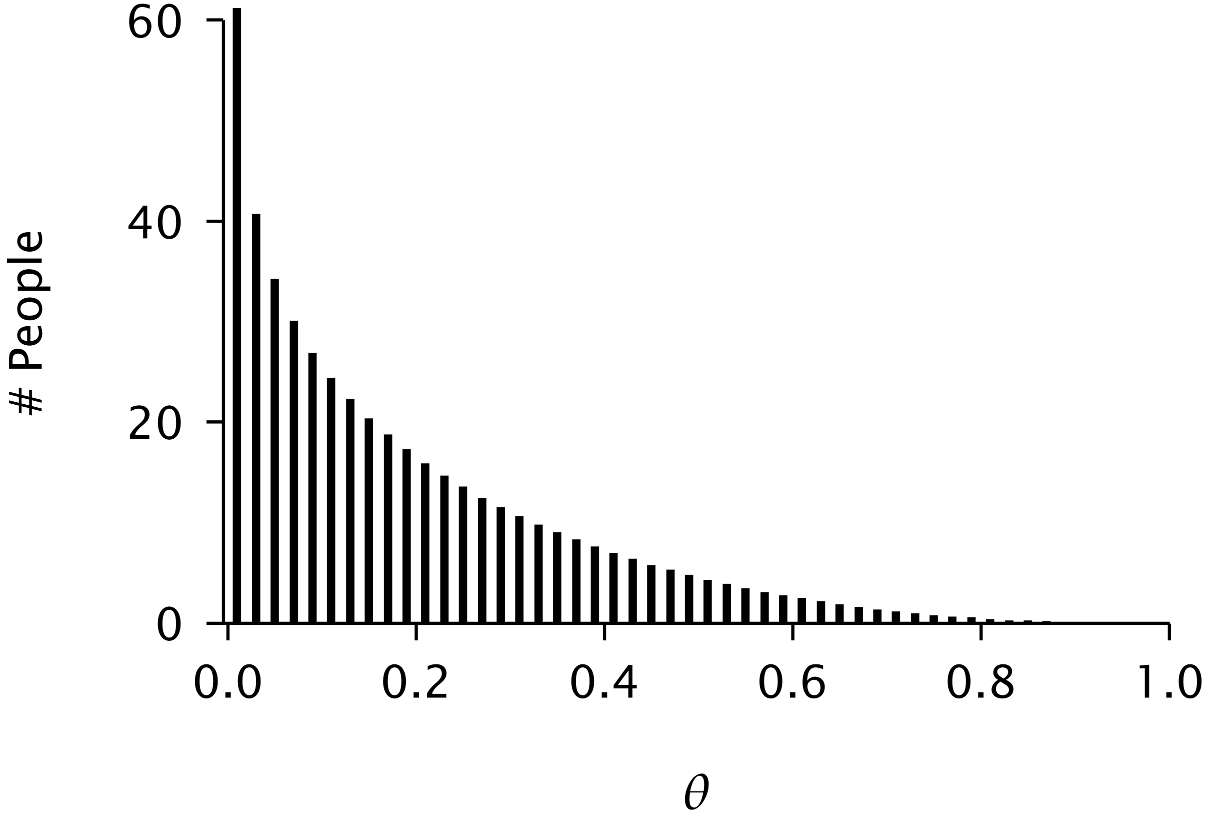
$E(\Theta) = 0.371 \longrightarrow$ expect $1000 \times (1 - 0.371) = 629$ customers to renew at the end of Year 1.

Year 2



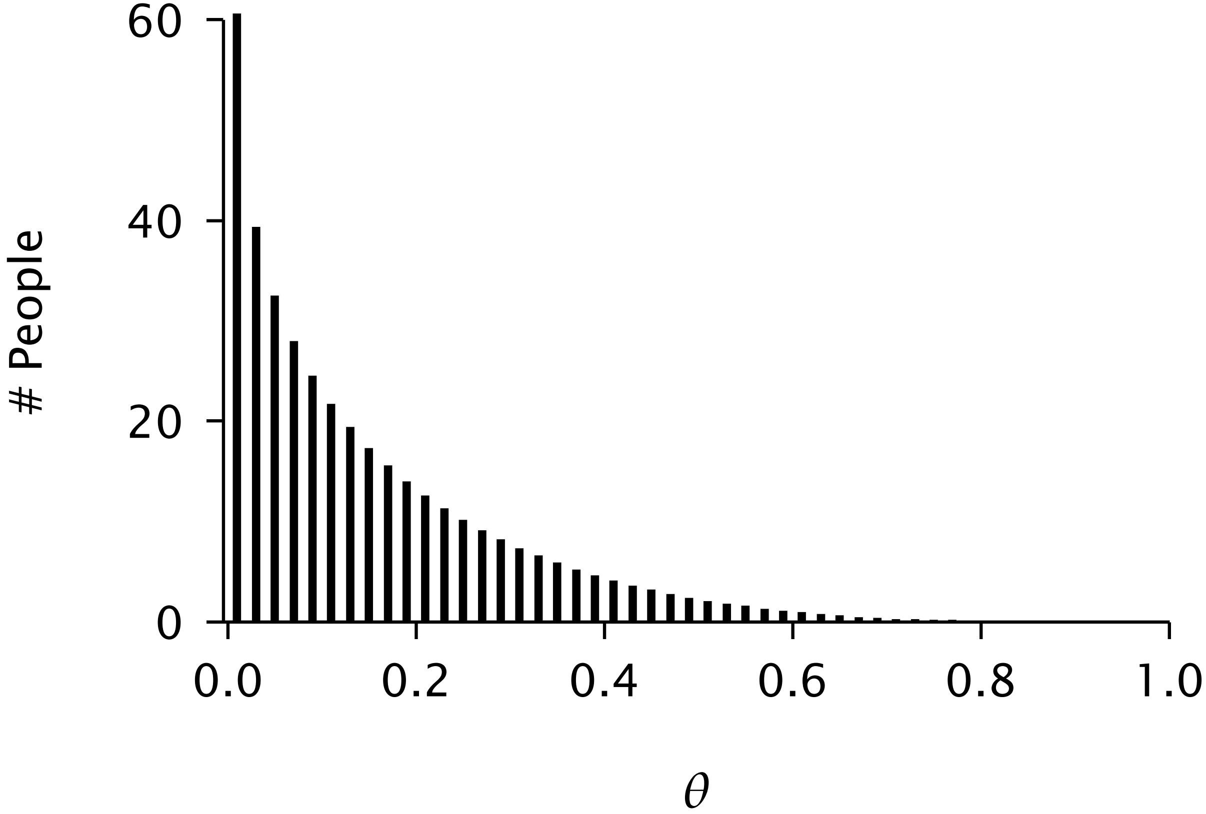
$E(\Theta) = 0.250 \longrightarrow$ expect $629 \times (1 - 0.250) = 472$ customers to renew at the end of Year 2.

Year 3



$E(\Theta) = 0.188 \longrightarrow$ expect $472 \times (1 - 0.188) = 383$ customers to renew at the end of Year 3.

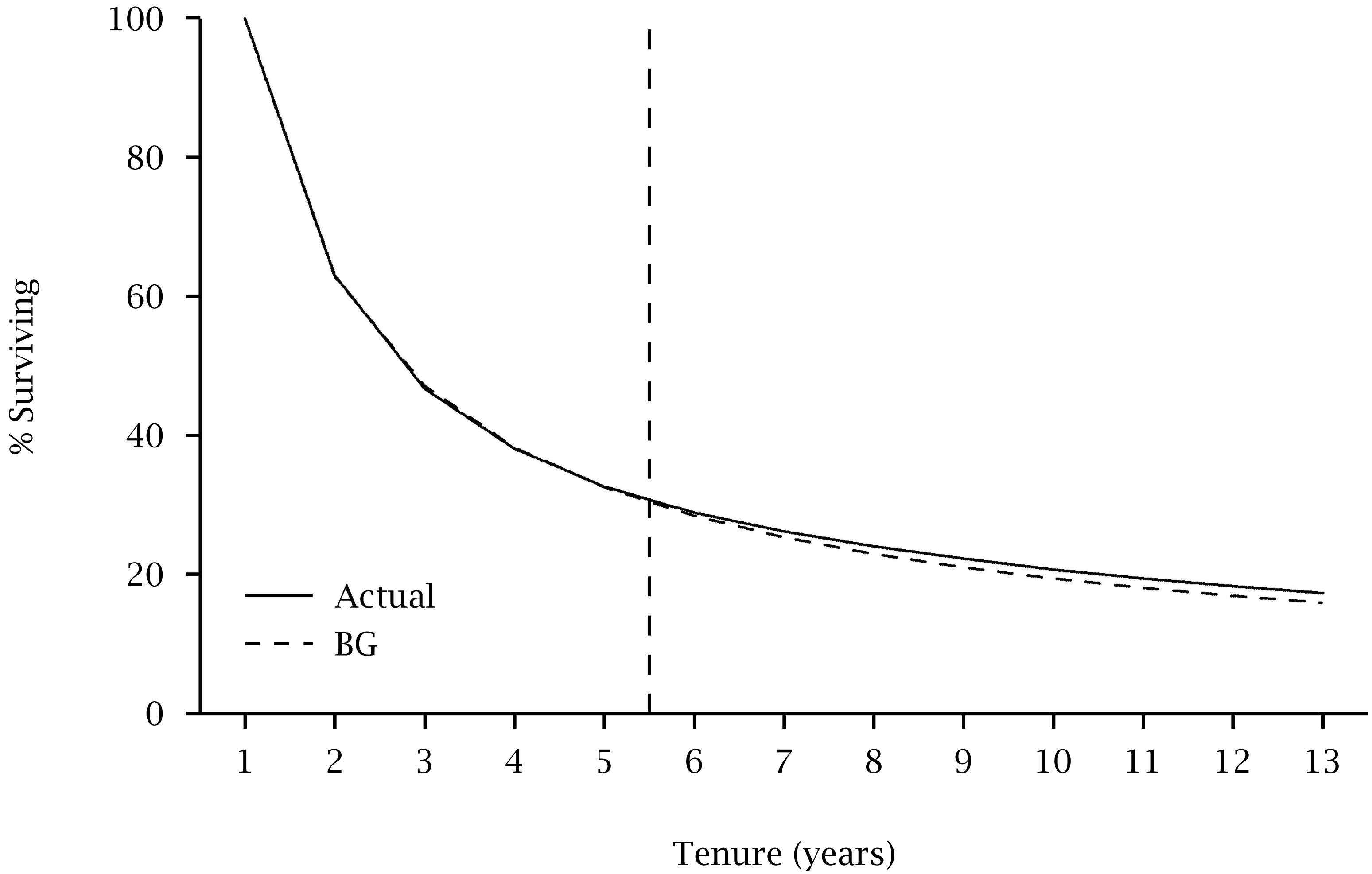
Year 4



$E(\Theta) = 0.151 \longrightarrow$ expect $383 \times (1 - 0.151) = 325$ customers to renew at the end of Year 4.

	A	B	C	D	E	F
1	gamma	0.764				
2	delta	1.296				
3	LL	-1401.6				
4						
5	t	# Cust.	# Lost	P(die)	S(t)	
6	0	1000			1.0000	
7	1	631	369	0.3708	0.6292	-366.08
8	2	468	163	0.1571	0.4721	-301.74
9	3	382	86	0.0888	0.3833	-208.22
10	4	326	56	0.0579	0.3255	-159.59
11	5			0.0410	0.2845	-365.93
12	6			0.0308	0.2537	
13	7			0.0240	0.2296	
14	8			0.0194	0.2103	
15	9			0.0160	0.1943	
16	10			0.0134	0.1809	
17	11			0.0115	0.1694	
18	12			0.0099	0.1595	

Survival Curve Projection



Implied Retention Rates

- The retention rate for period t is defined as the proportion of customers who had renewed their contract at the end of period $t-1$ who then renewed their contract at the end of period t.

- For any model of customer tenure with survivor function $S(t)$,

$$r_t = \frac{S(t)}{S(t-1)}$$

Implied Retention Rates

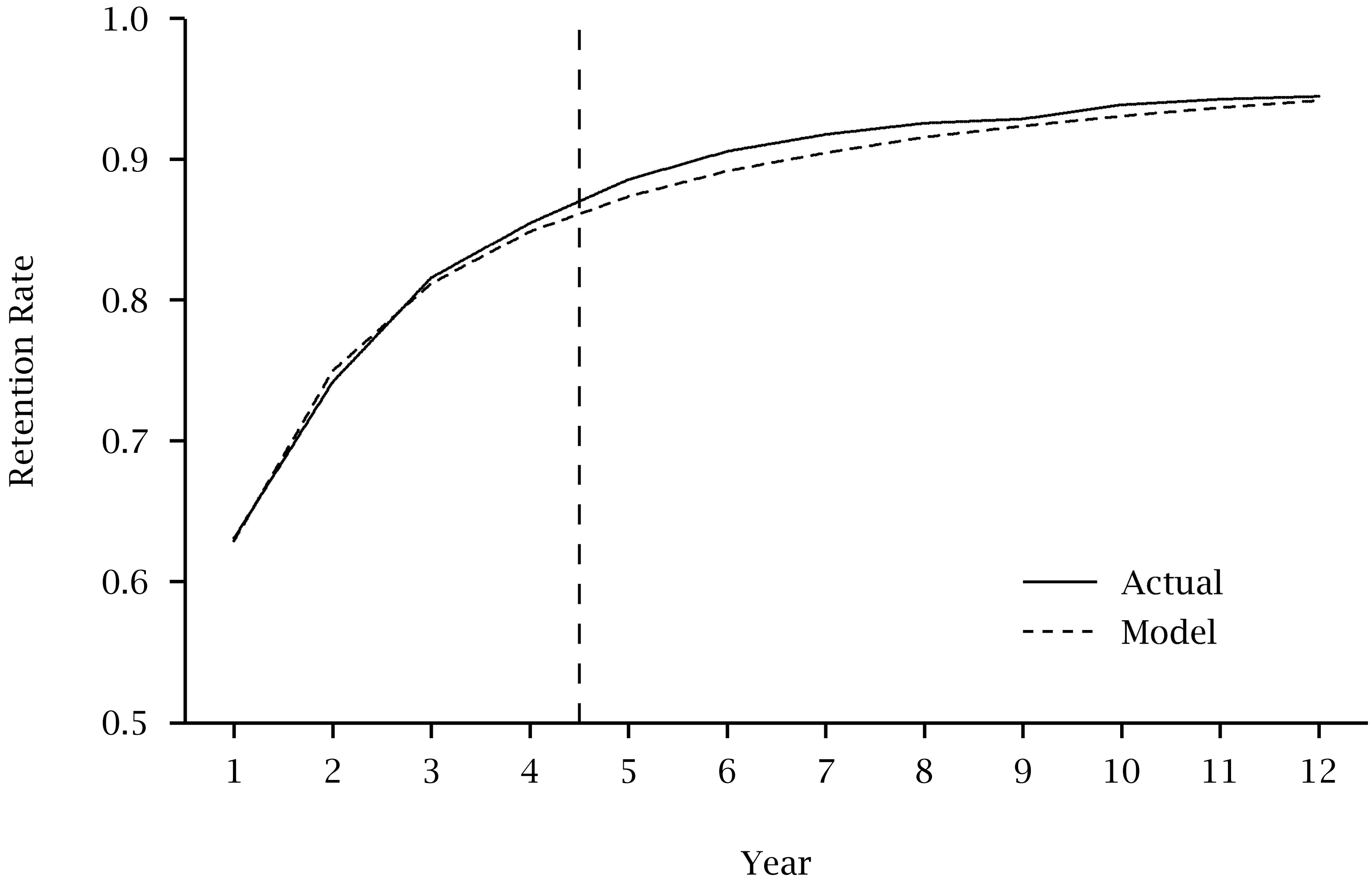
- For the BG model,

$$
\begin{aligned}
r_t &= \frac{B(\gamma, \delta + t)}{B(\gamma, \delta)} \bigg/ \frac{B(\gamma, \delta + t - 1)}{B(\gamma, \delta)} \\
&= \frac{B(\gamma, \delta + t)}{B(\gamma, \delta + t - 1)} \\
&= \frac{\delta + t - 1}{\gamma + \delta + t - 1}
\end{aligned}
$$

- An increasing function of time, even though the individual-level retention probability is constant.

- A sorting effect in a heterogeneous population.

Projecting Retention Rates



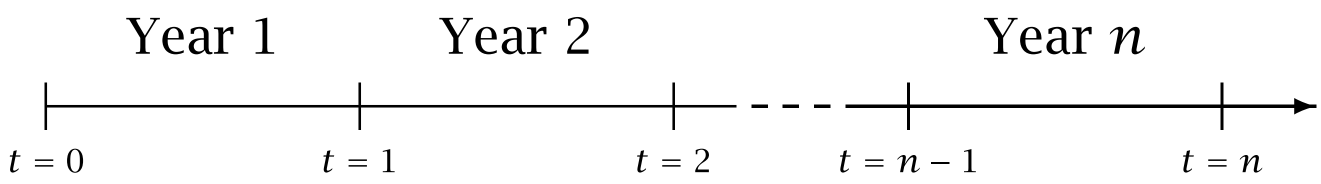
Back to the Motivating Problem

	Year 1		Year 2		Year 3		Year 4		Year 5	
$t = 0$	1000	$t = 1$	631	$t = 2$	468	$t = 3$	382	$t = 4$	326	$t = 5$

- What is the maximum amount you would spend to acquire a customer?

- What is the expected *residual value* of this group of customers at the end of Year 5?

Maximum Spend on Customer Acquisition



	Year 1	Year 2	Year n
Net CF	v_0	v_1	v_{n-1}
P(alive)	$S(0)$	$S(1)$	$S(n-1)$
discount	1	$\dfrac{1}{(1+d)}$	$\dfrac{1}{(1+d)^{n-1}}$

$$E(CLV) = \sum_{t=0}^{\infty} \frac{v_t\, S(t)}{(1+d)^t}$$

Maximum Spend on Customer Acquisition

- Assuming an individual's net cashflow per year is constant ($\bar{v}$),

$$E(CLV) = \bar{v} \underbrace{\sum_{t=0}^{\infty} \frac{S(t)}{(1+d)^t}}_{\text{discounted expected lifetime}}$$

- For the BG model,

$$DEL(\gamma, \delta, d) = \sum_{t=0}^{\infty} \frac{S(t \mid \gamma, \delta)}{(1+d)^t}$$

Computing DEL in Excel

	A	B	C	D	E	F
1	gamma	0.764		**DEL**	3.62	
2	delta	1.296				
3	d	0.1				
4						
5	t	P(T=t)	S(t)		disc.	
6	0		1.0000		1.0000	
7	1	0.3708	0.6292		0.9091	
8	2	0.1571	0.47		0.8264	
9	3	0.0888	0.3833		0.7513	
10	4	0.0579	0.3255		0.6830	
11	5	0.0410	0.2845		0.6209	
12	6	0.0308	0.2537		0.5645	
13	7	0.0240	0.2296		0.5132	
14	8	0.0194	0.2103		0.4665	
15	9	0.0160	0.1943		0.4241	
16	10	0.0134	0.1809		0.3855	
201	195	0.0001	0.0203		8.48E-09	
202	196	0.0001	0.0202		7.71E-09	
203	197	0.0001	0.0201		7.01E-09	
204	198	0.0001	0.0201		6.37E-09	
205	199	0.0001	0.0200		5.79E-09	
206	200	0.0001	0.0199		5.27E-09	

=SUMPRODUCT(C6:C206,E6:E206)

=1/(1+$B$3)^A6

Maximum Spend on Customer Acquisition

	$E(CLV)$
Naïve	\$247
BG	\$362

$\Longrightarrow$ Only considering the first five years of a customer's relationship with the firm would undervalue an "as-yet-to-be-acquired" customer by \$115 (32%).

Back to the Motivating Problem

Year 1	Year 2	Year 3	Year 4	Year 5
1000	631	468	382	326

$t = 0$ $t = 1$ $t = 2$ $t = 3$ $t = 4$ $t = 5$

- What is the maximum amount you would spend to acquire a customer?
- What is the expected *residual value* of this group of customers at the end of Year 5?

Residual Value of an Existing Customer

Standing at the end of Year n, what is the expected residual lifetime value of a customer?

	Year n	$t = n$	Year $n + 1$	$t = n + 1$	Year $n + 2$	$t = n + 2$
Net CF			v_n		v_{n+1}	
$P(\text{alive})$			$S(n \mid t > n - 1)$		$S(n + 1 \mid t > n - 1)$	
discount			1		$\frac{1}{(1 + d)}$	

$$E(RLV \mid n - 1 \text{ renewals}) = \sum_{t=n}^{\infty} \frac{v_t\, S(t \mid t > n - 1)}{(1 + d)^{t-n}}$$

$$= \sum_{t=n}^{\infty} \frac{v_t\, S(t)/S(n - 1)}{(1 + d)^{t-n}}$$

Residual Value of an Existing Customer

- Assuming an individual's net cashflow per year is constant ($\bar{v}$),

$$E(RLV \mid n - 1 \text{ renewals}) = \bar{v} \underbrace{\sum_{t=n}^{\infty} \frac{S(t \mid t > n - 1)}{(1 + d)^{t-n}}}_{\text{discounted expected residual lifetime}}$$

- For the BG model,

$$DERL(\gamma, \delta, d; n - 1 \text{ renewals})$$
$$= \sum_{t=n}^{\infty} \frac{S(t \mid \gamma, \delta)/S(n - 1 \mid \gamma, \delta)}{(1 + d)^{t}}$$

Computing DERL in Excel

	A	B	C	D	E	F	G
1	gamma	0.764			**DERL**	5.68	
2	delta	1.296					
3	d	0.1					
4					4 renewals (n=5)		
5	t	P(T=t)	S(t)		S(t|t>n-1)	disc.	
6	0		1.0000				
7	1	0.3708	0.6292				
8	2	0.1571	0.4721				
9	3	0.0888	0.3833				
10	4	0.0579	0.3255				
11	5	0.0410	0.2845		0.8740	1.0000	
12	6	0.0308	0.2537		0.7794	0.9091	
13	7	0.0240	[illegible]		0.7056	0.8264	
14	8	0.0194	[illegible]		[illegible]	0.7513	
15	9	0.0160	[illegible]		[illegible]	[illegible]	
16	10	0.0134	0.1809		0.5558	0.6209	
201	195	0.0001	0.0203		0.0623	1.37E-08	
202	196	0.0001	0.0202		0.0621	1.24E-08	
203	197	0.0001	0.0201		0.0619	1.13E-08	
204	198	0.0001	0.0201		0.0616	1.03E-08	
205	199	0.0001	0.0200		0.0614	9.33E-09	
206	200	0.0001	0.0199		0.0612	8.48E-09	

=SUMPRODUCT(E11:E206,F11:F206)

=C11/$C$10

=1/(1+$B$3)^(A13-5)

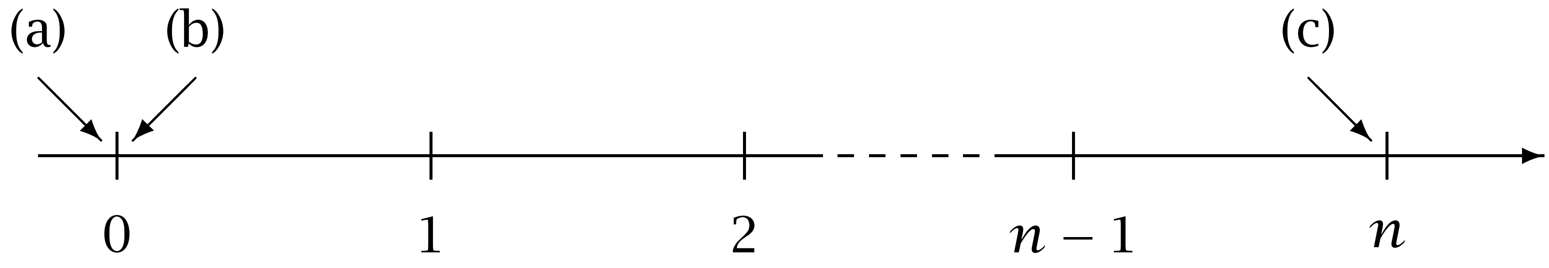
Residual Value of the Cohort

The expected residual lifetime value of a Year 5 customer is $\$100 \times 5.68 = \$568$.

$\Longrightarrow$ The collective expected residual value of those members of the cohort who are still active in Year 5 is $326 \times \$568 = \$185{,}168$.

Expressions for DE(R)L

Different points in time at which a customer's discounted expected (residual) lifetime can be computed:



(a) An "as-yet-to-be-acquired" customer

(b) A "just-acquired" customer

(c) In their nth period as a customer

Summary

- The notion of CLV
- Classifying business settings
- Understanding and projecting survival
- The beta-geometric (BG) distribution as a model of customer contract duration
- The right way to compute CLV in discrete-time contractual settings.
- The concepts of DEL and DERL for contractual settings, and their evaluation when contract durations are characterized by the BG.