Review of Probability Models

The Logic of Probability Models

- Many researchers attempt to describe/predict behavior using observed variables.
- However, they still use random components in recognition that not all factors are included in the model.
- We treat behavior as if it were “random” (probabilistic, stochastic).
- We propose a model of individual-level behavior which is “summed” across heterogeneous individuals to obtain a model of aggregate behavior.

Building a Probability Model

(i) Determine the marketing decision problem/ information needed.

(ii) Identify the *observable* individual-level behavior of interest.

- We denote this by x.

(iii) Select a probability distribution that characterizes this individual-level behavior.

- This is denoted by $f(x|\theta)$.
- We view the parameters of this distribution as individual-level *latent traits*.

Building a Probability Model

(iv) Specify a distribution to characterize the distribution of the latent trait variable(s) across the population.

- We denote this by $g(\theta)$.
- This is often called the *mixing distribution*.

(v) Derive the corresponding *aggregate* or *observed* distribution for the behavior of interest:

$$f(x) = \int f(x|\theta)g(\theta)\,d\theta$$

Building a Probability Model

(vi) Estimate the parameters (of the mixing distribution) by fitting the aggregate distribution to the observed data.

(vii) Use the model to solve the marketing decision problem/provide the required information.

“Classes” of Models

- We focus on three fundamental behavioral processes:
  - Timing ⟶ “when / how long”
  - Counting ⟶ “how many”
  - “Choice” ⟶ “whether / which”
- Our toolkit contains simple models for each behavioral process.
- More complex behavioral phenomena can be captured by combining models from each of these processes.

Individual-level Building Blocks

Count data arise from asking the question, “How many?”. As such, they are non-negative integers with no upper limit.

Let the random variable X be a count variable:

X is distributed Poisson with mean λ if

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

Individual-level Building Blocks

Timing (or duration) data are generated by answering "when" and "how long" questions, asked with regards to a specific event of interest.

The models we develop for timing data are also used to model other non-negative continuous quantities (e.g., transaction value).

Let the random variable T be a timing variable:

T is distributed exponential with rate parameter λ if

$$F(t \mid \lambda) = P(T \leq t \mid \lambda) = 1 - e^{-\lambda t}, \; t > 0.$$

Individual-level Building Blocks

A Bernoulli trial is a probabilistic experiment in which there are two possible outcomes, 'success' (or '1') and 'failure' (or '0'), where θ is the probability of success.

Repeated Bernoulli trials lead to the *geometric* and *binomial* distributions.

Individual-level Building Blocks

Let the random variable X be the number of independent and identically distributed Bernoulli trials required until the first success:

X is a geometric random variable, where

$$P(X = x \mid \theta) = \theta(1-\theta)^{x-1}, \; x = 1, 2, 3, \ldots$$

The geometric distribution can be used to model *either* omitted-zero class count data *or* discrete-time timing data.

Individual-level Building Blocks

Let the random variable X be the total number of successes occurring in n independent and identically distributed Bernoulli trials:

X is distributed binomial with parameter θ, where

$$P(X = x \mid n, \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \; x = 0, 1, 2, \ldots, n.$$

We use the binomial distribution to model repeated choice data — answers to the question, "How many times did a particular outcome occur in a fixed number of events?"

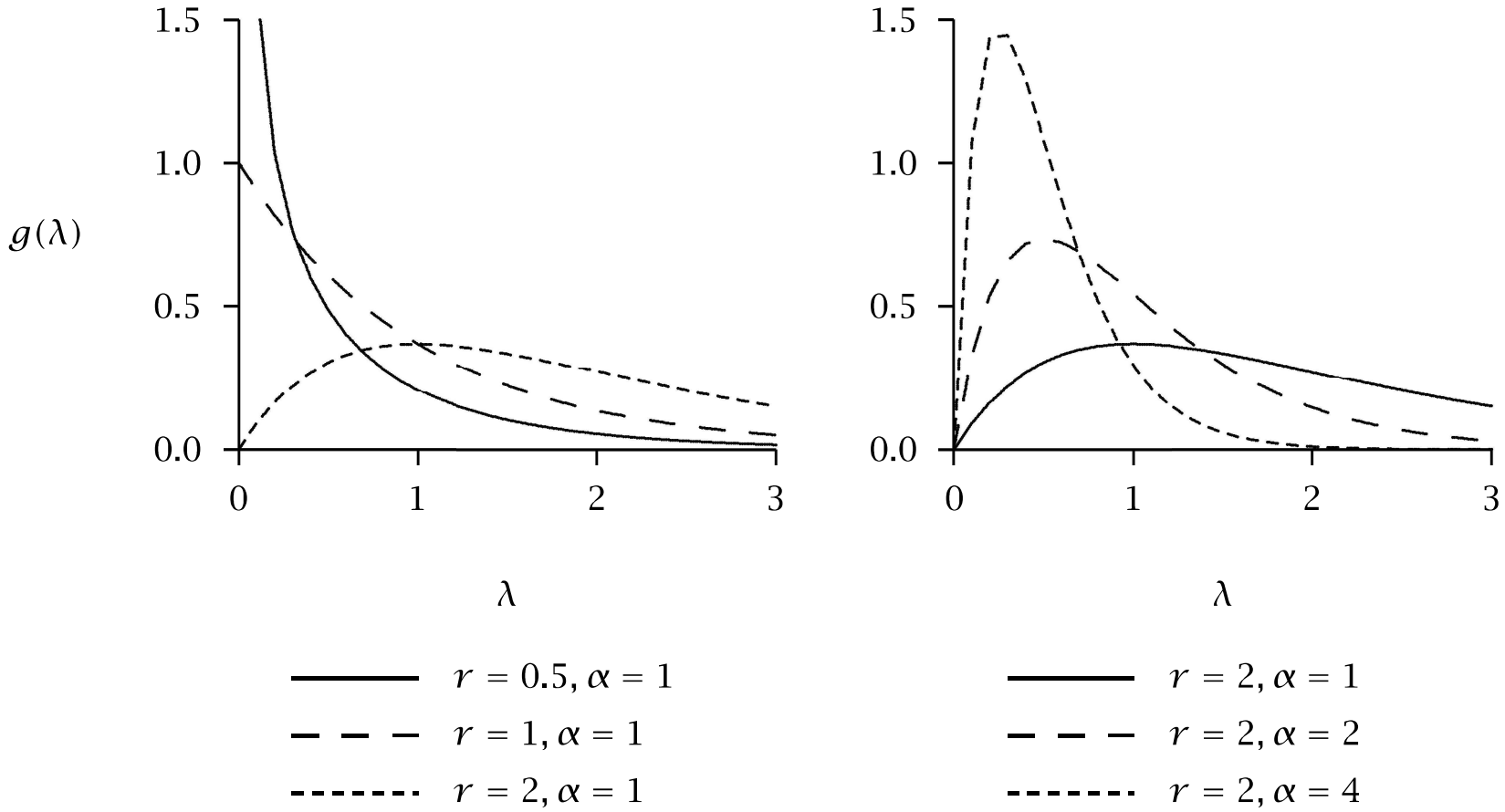
Capturing Heterogeneity in Latent Traits

The gamma distribution:

$$g(\lambda \mid r, \alpha) = \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)}, \ \lambda > 0$$

- $\Gamma(\cdot)$ is the gamma function
- r is the "shape" parameter and α is the "scale" parameter
- The gamma distribution is a flexible (unimodal) distribution ... and is mathematically convenient.

Illustrative Gamma Density Functions



Capturing Heterogeneity in Latent Traits

The beta distribution:

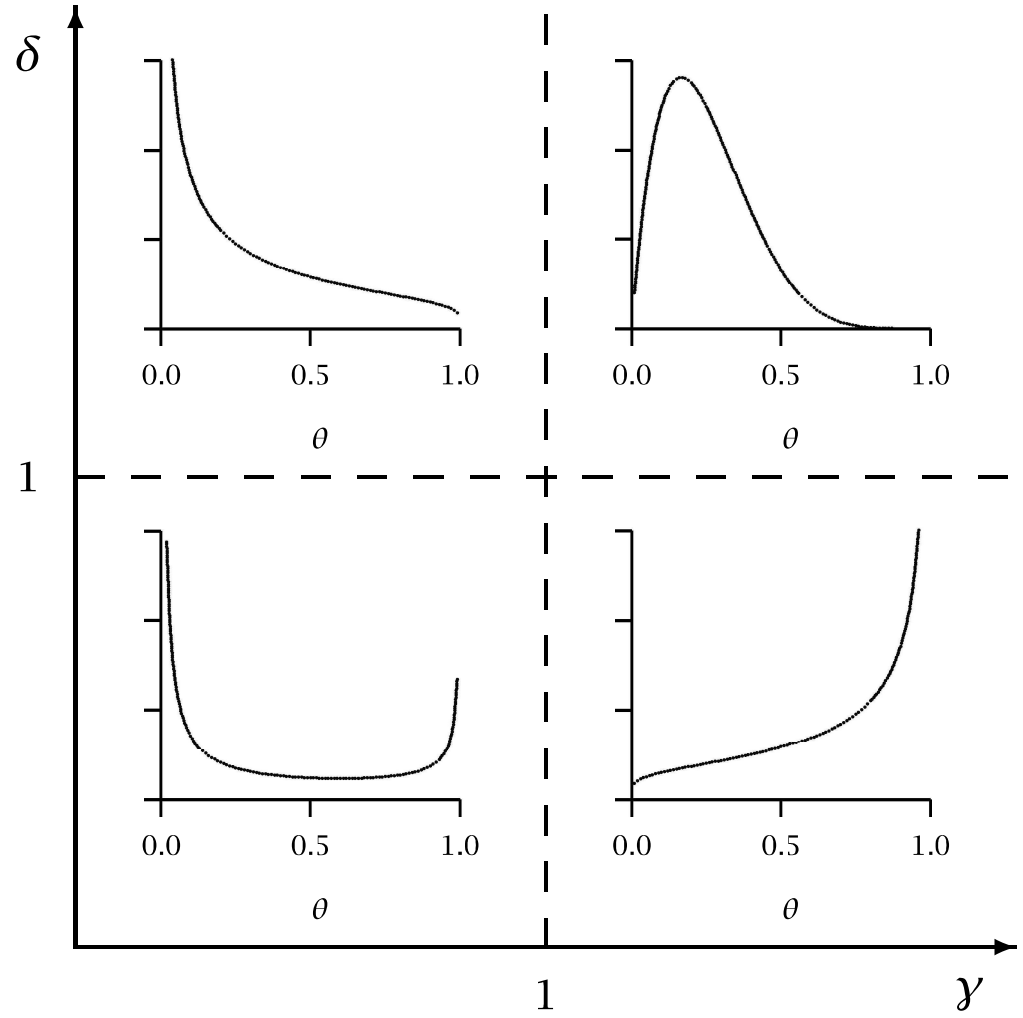
$$g(\theta \mid \gamma, \delta) = \frac{\theta^{\gamma-1}(1-\theta)^{\delta-1}}{B(\gamma,\delta)}\,,\ 0 < \theta < 1\,.$$

- $B(\gamma,\delta)$ is the beta function, which can be expressed in terms of gamma functions:

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

- The beta distribution is a flexible distribution ... and is mathematically convenient

Five General Shapes of the Beta Distribution



The Negative Binomial Distribution (NBD)

- The individual-level behavior of interest can be characterized by the Poisson distribution when the mean λ is known.
- We do not observe an individual's λ but assume it is distributed across the population according to a gamma distribution.

$$
\begin{aligned}
P(X = x \mid r, \alpha) &= \int_0^\infty P(X = x \mid \lambda) g(\lambda \mid r, \alpha) d\lambda \\
&= \frac{\Gamma(r + x)}{\Gamma(r) x!} \left(\frac{\alpha}{\alpha + 1}\right)^r \left(\frac{1}{\alpha + 1}\right)^x .
\end{aligned}
$$

The Pareto Distribution of the Second Kind

- The individual-level behavior of interest can be characterized by the exponential distribution when the rate parameter λ is known.
- We do not observe an individual's λ but assume it is distributed across the population according to a gamma distribution.

$$\begin{aligned} F(t \mid r, \alpha) &= \int_0^\infty F(t \mid \lambda)\, g(\lambda \mid r, \alpha)\, d\lambda \\ &= 1 - \left(\frac{\alpha}{\alpha + t}\right)^r . \end{aligned}$$

The Beta-Geometric Model

- The individual-level behavior of interest can be characterized by the geometric distribution when the parameter θ is known.
- We do not observe an individual's θ but assume it is distributed across the population according to a beta distribution.

$$P(X = x \mid \gamma, \delta) = \int_0^1 P(X = x \mid \theta) g(\theta \mid \gamma, \delta)\, d\theta$$
$$= \frac{B(\gamma + 1, \delta + x - 1)}{B(\gamma, \delta)}.$$

The Beta-Binomial Distribution

- The individual-level behavior of interest can be characterized by the binomial distribution when the parameter θ is known.
- We do not observe an individual's θ but assume it is distributed across the population according to a beta distribution.

$$P(X = x \mid n, \alpha, \beta) = \int_0^1 P(X = x \mid n, \theta) g(\theta \mid \alpha, \beta)\, d\theta$$
$$= \binom{n}{x} \frac{B(\alpha + x, \beta + n - x)}{B(\alpha, \beta)}.$$

Summary of Probability Models

Phenomenon	Individual-level	Heterogeneity	Model
Counting	Poisson	gamma	NBD
Timing	exponential	gamma	Pareto II
Discrete timing (or counting)	geometric	beta	BG
Choice	binomial	beta	BB

Customer Lifetime Value

Customer lifetime value is *the present value of the future cash flows associated with the customer*.

- A forward-looking concept
- Not to be confused with (historic) customer profitability

Two key questions:

- How long will the customer remain “alive”?
- What is the net cashflow per period while “alive”?

Q: How long will the customer remain “alive”?

A: It depends on the business setting . . .

Classifying Business Settings

Consider the following two statements regarding the size of a company's customer base:

- Based on numbers presented in a news release that reported Vodafone Group Plc's results for the six months ended 30 September 2013, we see that Vodafone UK has 11.3 million "pay monthly" customers at the end of that period.
- In his "Q3 2013 Earnings Conference Call" the CFO of Amazon made the comment that "[a]ctive customer accounts exceeded 224 million," where customers are considered active when they have placed an order during the preceding twelve-month period.

Classifying Business Settings

- It is important to distinguish between contractual and noncontractual settings:
  - In a *contractual* setting, we observe the time at which a customer ended their relationship with the firm.
  - In a *noncontractual* setting, the time at which a customer "dies" is unobserved (i.e., attrition is latent).
- The challenge of noncontractual markets:

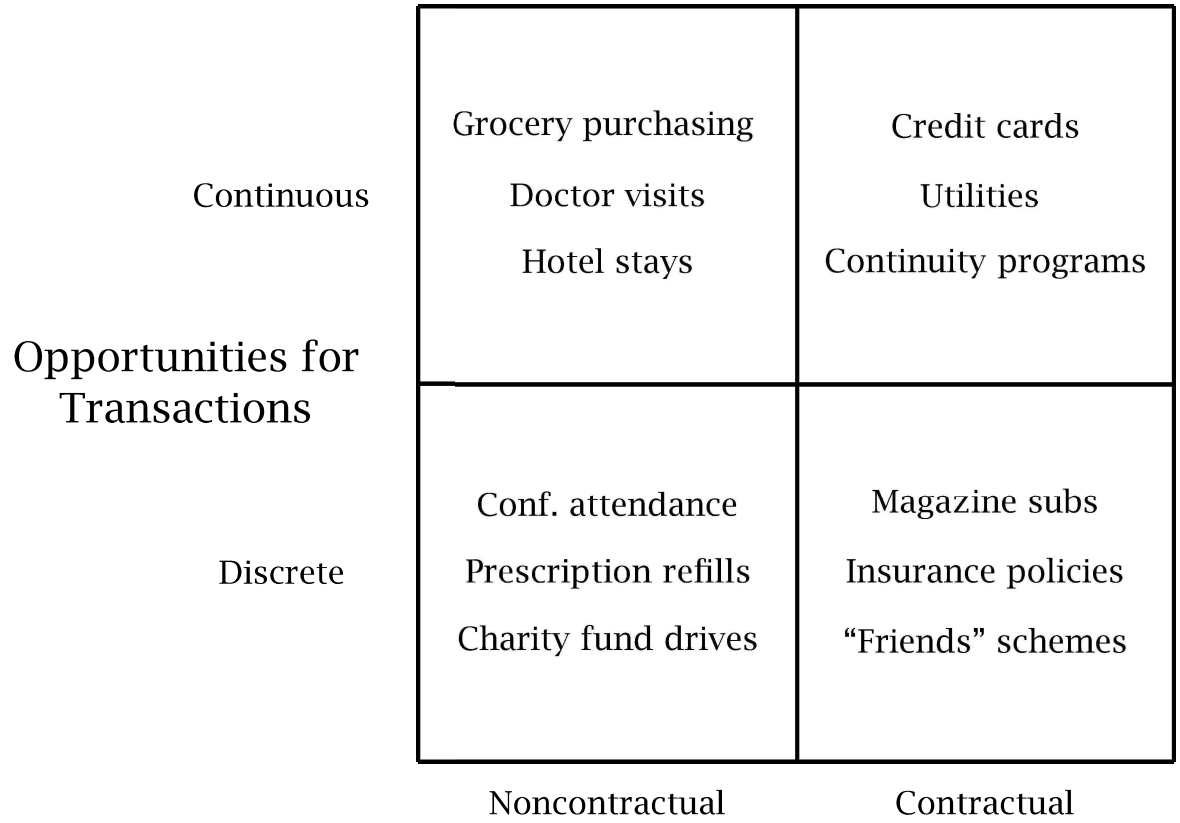
  How do we differentiate between those customers who have ended their relationship with the firm versus those who are simply in the midst of a long hiatus between transactions?

Classifying Business Settings

Consider the following four specific business settings:

- Airline lounges
- Electrical utilities
- Academic conferences
- Mail-order clothing companies.

Classifying Customer Bases



Adapted from: 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.

Calculating CLV

Standard classroom formula:

$$CLV = \sum_{t=0}^{T} m \frac{r^t}{(1+d)^t}$$

where m = net cash flow per period (if alive)

r = retention rate

d = discount rate

T = horizon for calculation

Calculating $E(CLV)$

A more correct starting point:

$$E(CLV) = \int_0^\infty E[v(t)]S(t)d(t)dt$$

where $E[v(t)]$ = expected value (or net cashflow) of the customer at time t (if alive)

$S(t)$ = the probability that the customer is alive beyond time t

$d(t)$ = discount factor that reflects the present value of money received at time t

Calculating $E(CLV)$

- Definitional; of little use by itself.
- We must operationalize $E[v(t)]$, $S(t)$, and $d(t)$ in a specific business setting . . . then solve the integral.
- Important distinctions:
  - Expected lifetime value of an as-yet-to-be-acquired customer
  - Expected lifetime value of a just-acquired customer
  - Expected *residual* lifetime value, $E(RLV)$, of an existing customer

Calculating $E(CLV)$

- The expected lifetime value of an as-yet-to-be-acquired customer is given by

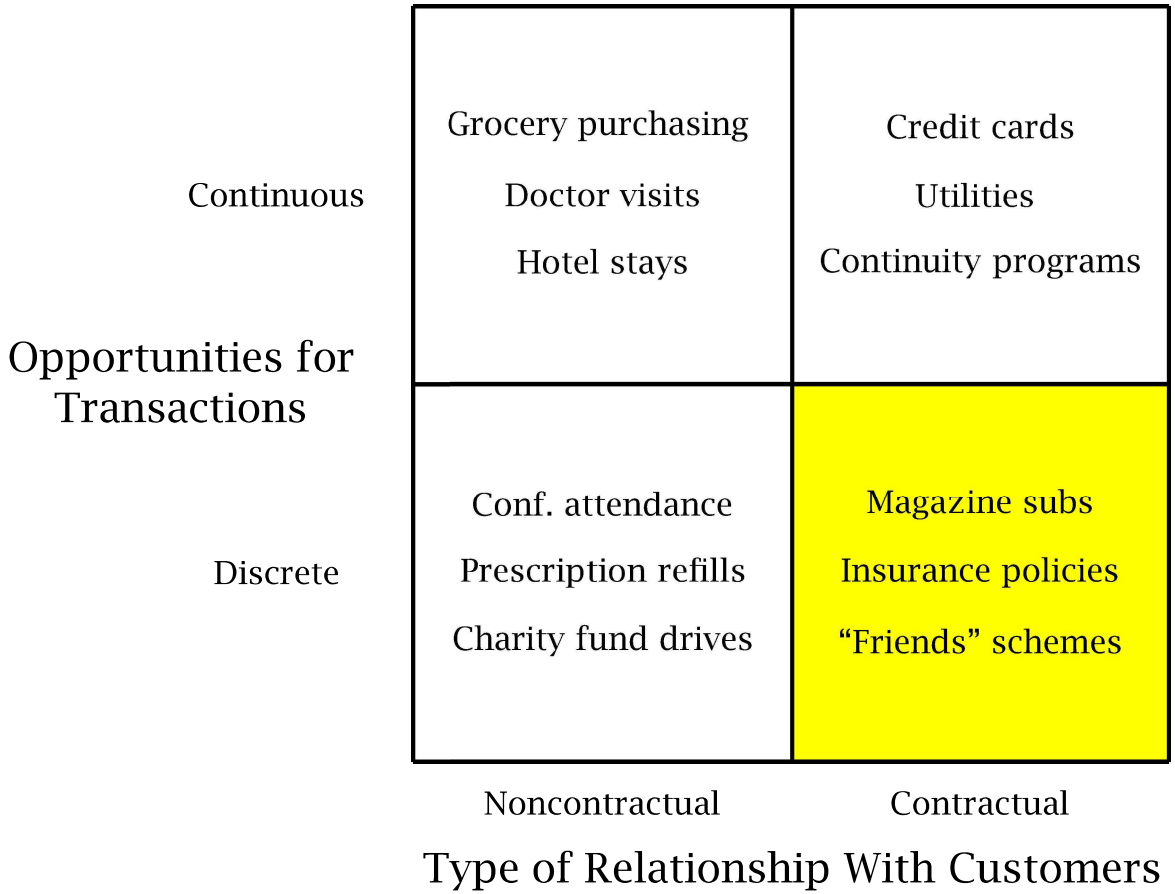
$$E(CLV) = \int_0^\infty E[v(t)]S(t)d(t)dt$$

- Standing at time T, the expected residual lifetime value of an existing customer is given by

$$E(RLV) = \int_T^\infty E[v(t)]S(t \mid t > T)d(t-T)dt$$

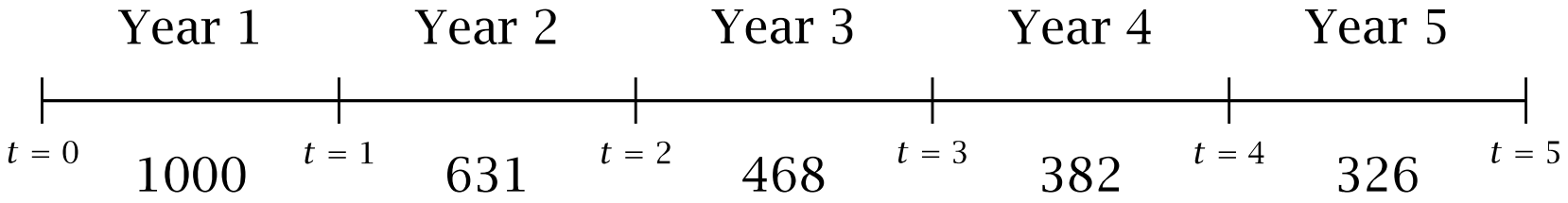
Models for Contractual Settings

Classifying Customer Bases



Motivating Problem

Consider a company with a subscription-based business model. 1000 customers are acquired at the beginning of Year 1 with the following pattern of renewals over the subsequent four years:



- 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?

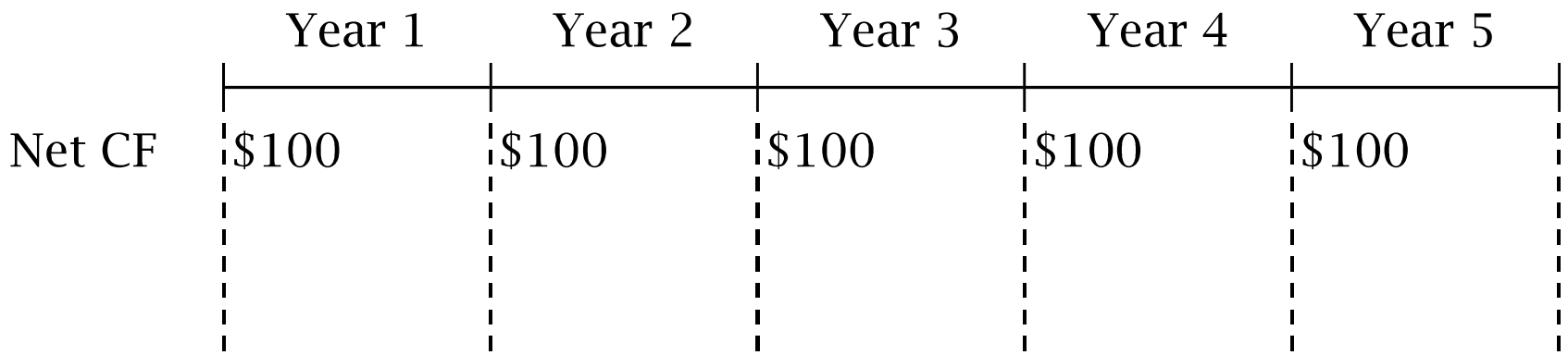
Motivating Problem

Let us assume:

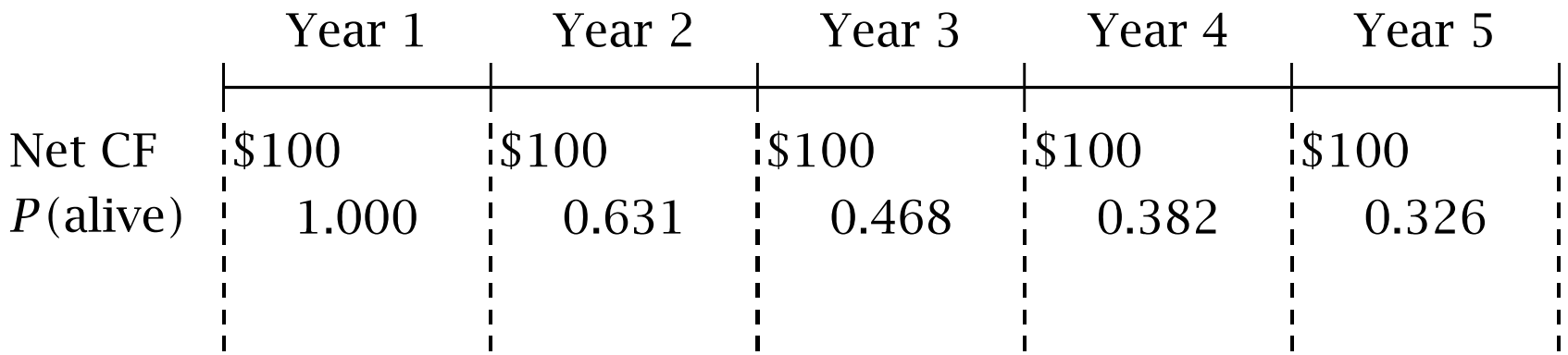
- Each contract is annual, starting on January 1 and expiring at 11:59pm on December 31.
- An average net cashflow of $100/year, which is "booked" at the beginning of the contract period.
- A 10% annual discount rate.

Spending on Customer Acquisition

Spending on Customer Acquisition



Spending on Customer Acquisition



Spending on Customer Acquisition

	Year 1	Year 2	Year 3	Year 4	Year 5
Net CF	\$100	\$100	\$100	\$100	\$100
P(alive)	1.000	0.631	0.468	0.382	0.326
discount	1	$\frac{1}{1.1}$	$\frac{1}{(1.1)^2}$	$\frac{1}{(1.1)^3}$	$\frac{1}{(1.1)^4}$

Spending on Customer Acquisition

	Year 1	Year 2	Year 3	Year 4	Year 5
Net CF	\$100	\$100	\$100	\$100	\$100
P(alive)	1.000	0.631	0.468	0.382	0.326
discount	1	$\frac{1}{1.1}$	$\frac{1}{(1.1)^2}$	$\frac{1}{(1.1)^3}$	$\frac{1}{(1.1)^4}$

For a randomly chosen customer,

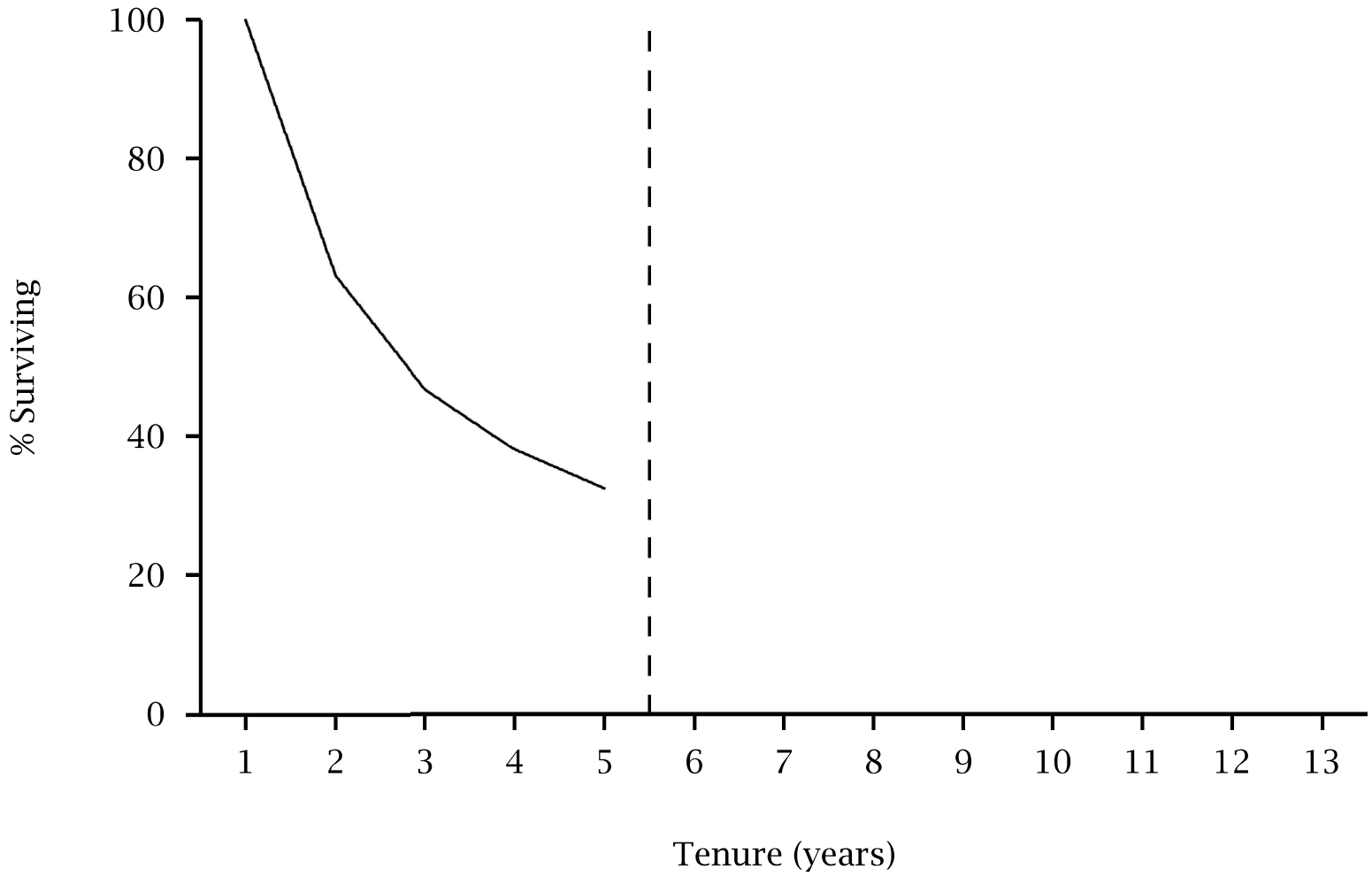
$$E(\text{CLV}) = \$100 + \$100 \times \frac{0.631}{1.1} + \$100 \times \frac{0.468}{(1.1)^2} + \$100 \times \frac{0.382}{(1.1)^3} + \$100 \times \frac{0.326}{(1.1)^4}$$

$$= \$247$$

Spending on Customer Acquisition

- This implies we can justify spending up to $247 to acquire a new customer (based on expected "profitability" over the five year period).
- But what about the value beyond year 5?

Projecting the “Survival Curve”



Natural Starting Point

Project the survival curve using functions of time:

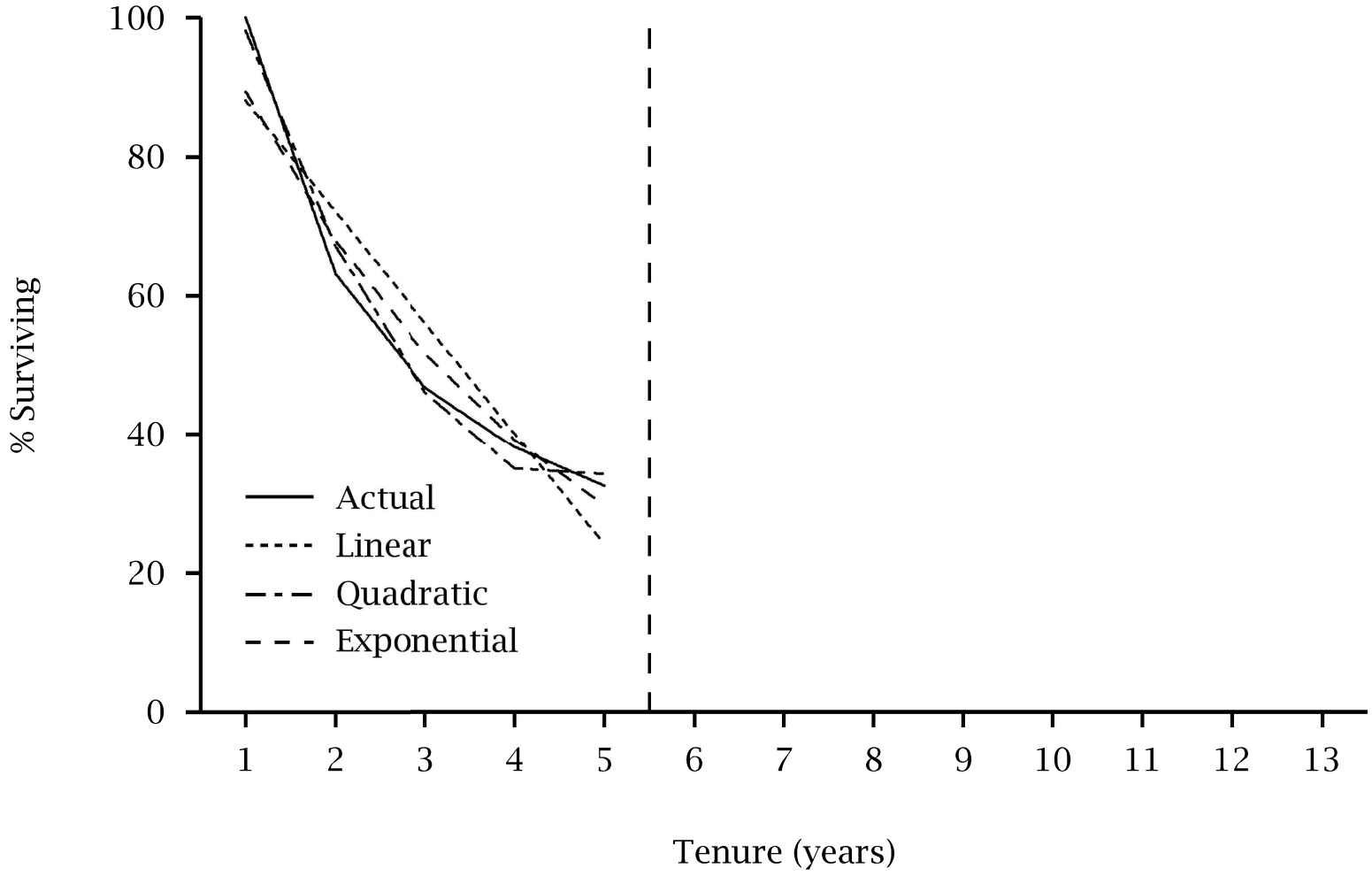
- Consider linear, quadratic, and exponential functions
- Let y = the proportion of customers surviving more than t years

$$y = 0.881 - 0.160t \qquad R^2 = 0.868$$

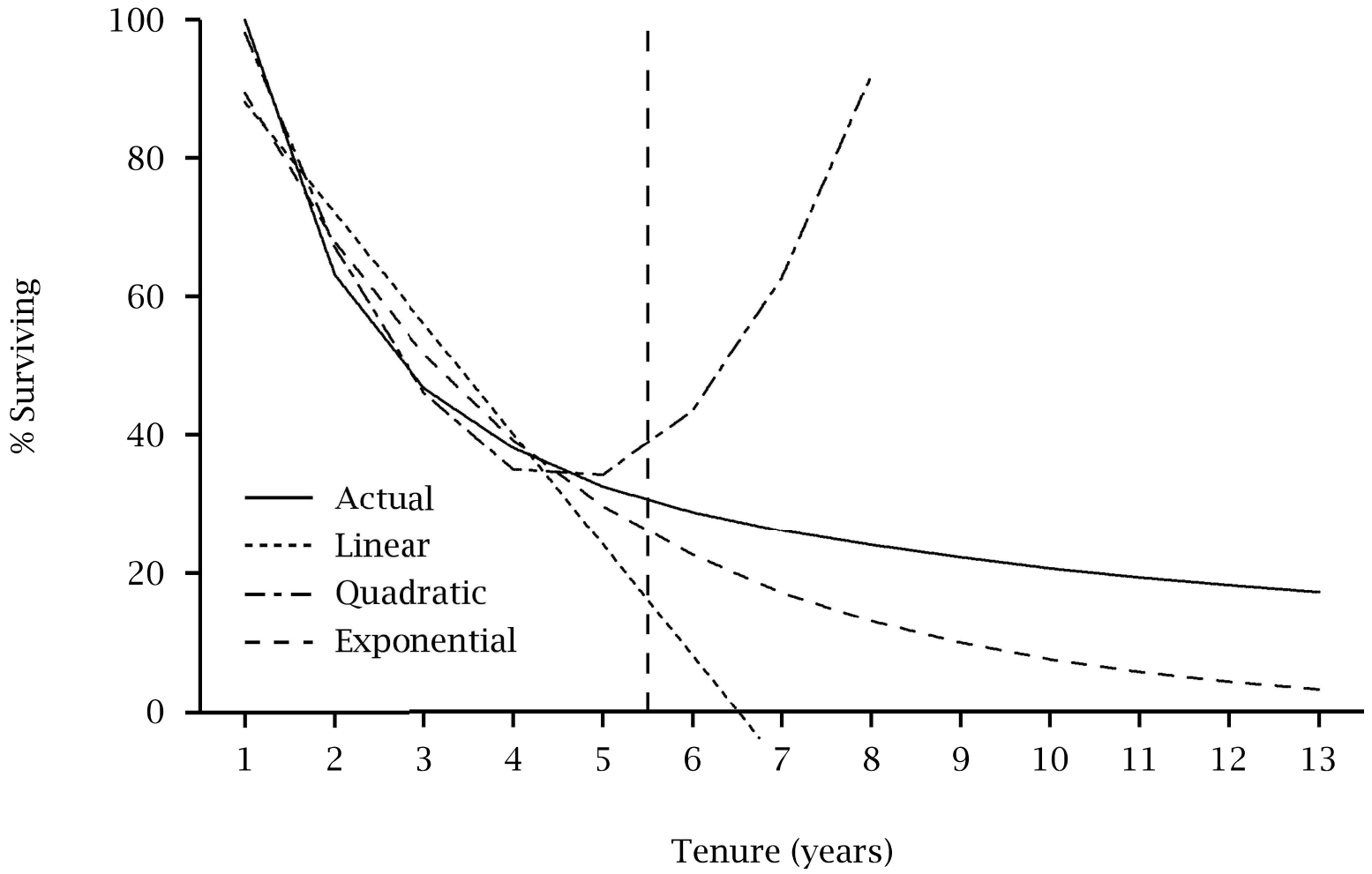
$$y = 0.981 - 0.361t + 0.050t^2 \qquad R^2 = 0.989$$

$$\ln(y) = -0.112 - 0.274t \qquad R^2 = 0.954$$

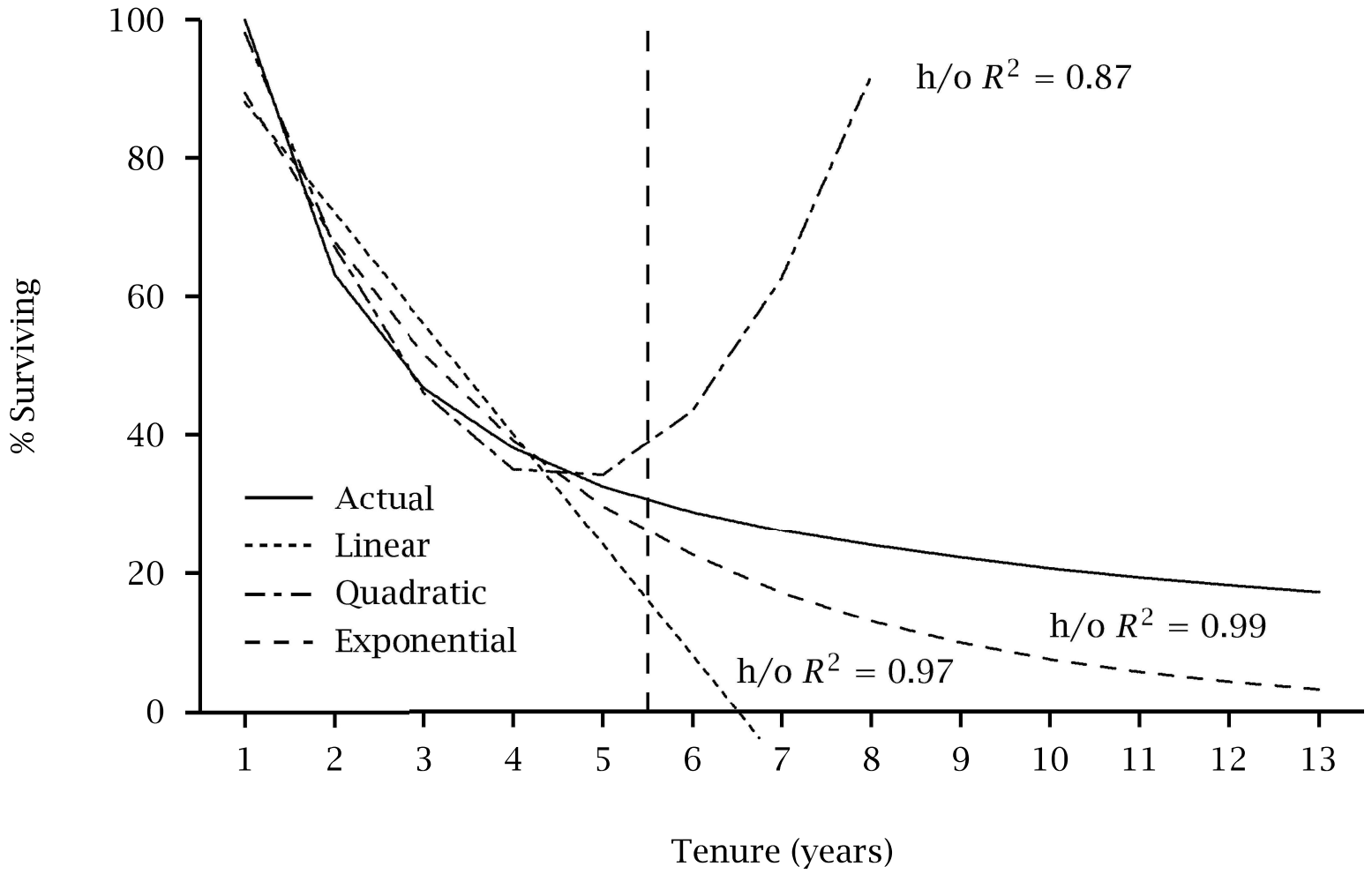
Model Fit


100
80
60
40
20
0
% Surviving
Actual
Linear
Quadratic
Exponential
1
2
3
4
5
6
7
8
9
10
11
12
13
Tenure (years)

Survival Curve Projections



Survival Curve Projections



Developing a Better Model (I)

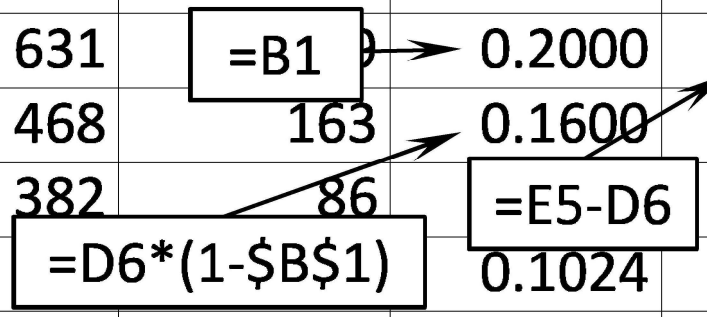
At the end of each contract period, a customer makes the renewal decision by tossing a coin: $\mathbb{H} \rightarrow$ renew, $\mathbb{T} \rightarrow$ don't renew

Length of relationship	
1 period	$\mathbb{T}$
2 periods	$\mathbb{H}$ $\mathbb{T}$
3 periods	$\mathbb{H}$ $\mathbb{H}$ $\mathbb{T}$
...	

i) $P(\mathbb{H}) = 1 - \theta$ is constant and unobserved.

ii) All customers have the same "churn probability" θ.

	A	B	C	D	E
1	theta	0.2			
2					
3					
4	t	# Cust.	# Lost	P(die)	S(t)
5	0	1000			1.0000
6	1	631		0.2000	0.8000
7	2	468	163	0.1600	0.6400
8	3	382	86		0.5120
9	4			0.1024	0.4096
10					



Developing a Better Model (I)

More formally:

- Let the random variable T denote the duration of the customer's relationship with the firm.
- We assume that the random variable T has a geometric distribution with parameter θ:

$$P(T = t \mid \theta) = \theta(1-\theta)^{t-1}, \quad t = 1, 2, 3, \ldots$$

$$S(t \mid \theta) = P(T > t \mid \theta)$$

$$= (1-\theta)^{t}, \quad t = 0, 1, 2, 3, \ldots$$

Estimating Model Parameters

Assuming

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

ii) we know $P(\mathbb{T}) = \theta$,

the probability of the observed pattern of renewals is:

$$
\begin{aligned}
&[P(T = 1 \mid \theta)]^{369}[P(T = 2 \mid \theta)]^{163}[P(T = 3 \mid \theta)]^{86} \\
&\qquad \times [P(T = 4 \mid \theta)]^{56}[S(t \mid \theta)]^{326} \\
&= [\theta]^{369}[\theta(1-\theta)]^{163}[\theta(1-\theta)^2]^{86} \\
&\qquad \times [\theta(1-\theta)^3]^{56}[(1-\theta)^4]^{326}
\end{aligned}
$$

Estimating Model Parameters

- Suppose we have two candidate coins:

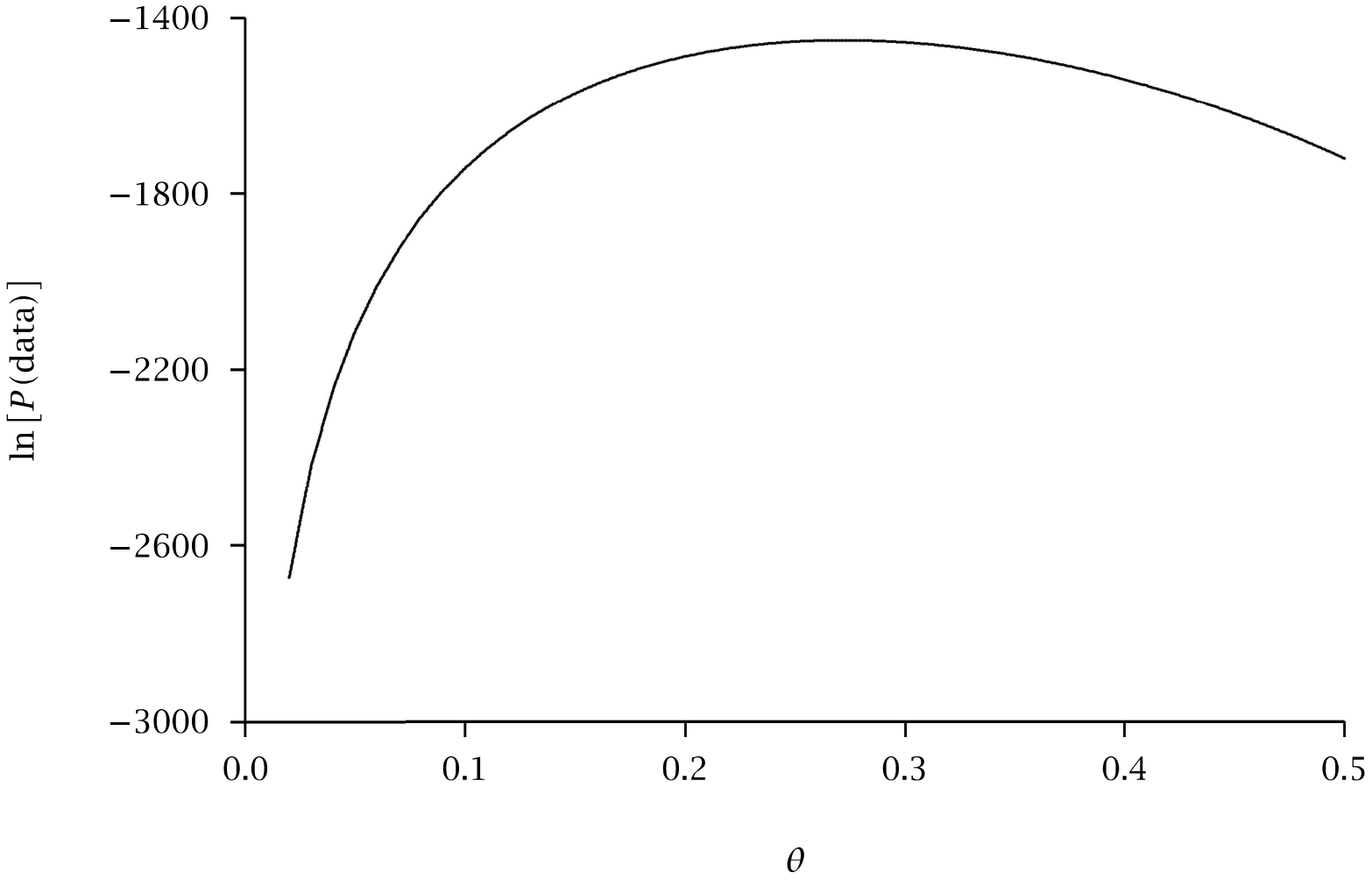
  Coin A: $\theta = 0.2$

  Coin B: $\theta = 0.5$

- Which coin is more likely to have generated the observed pattern of renewals across this set of 1000 customers?

θ	$P(\text{data} \mid \theta)$	$\ln[P(\text{data} \mid \theta)]$
0.2	6.00×10^{-647}	-1488.0
0.5	1.40×10^{-747}	-1719.7

Estimating Model Parameters



Estimating Model Parameters

We estimate the model parameters using the method of *maximum likelihood*:

- The likelihood function is defined as the probability of observing the data for a given set of the (unknown) model parameters.
- It is computed using the model and is viewed as a function of the model parameters:

$$L(\text{parameters} \mid \text{data}) = p(\text{data} \mid \text{parameters}) \,.$$

- For a given dataset, the maximum likelihood estimates of the model parameters are those values that maximize $L(\cdot)$.
- It is typically more convenient to use the natural logarithm of the likelihood function — the log-likelihood function.

Estimating Model Parameters

The log-likelihood function is given by:

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

The maximum value of the log-likelihood function is $LL = -1451.2$, which occurs at $\hat{\theta} = 0.272$.

Estimating Model Parameters

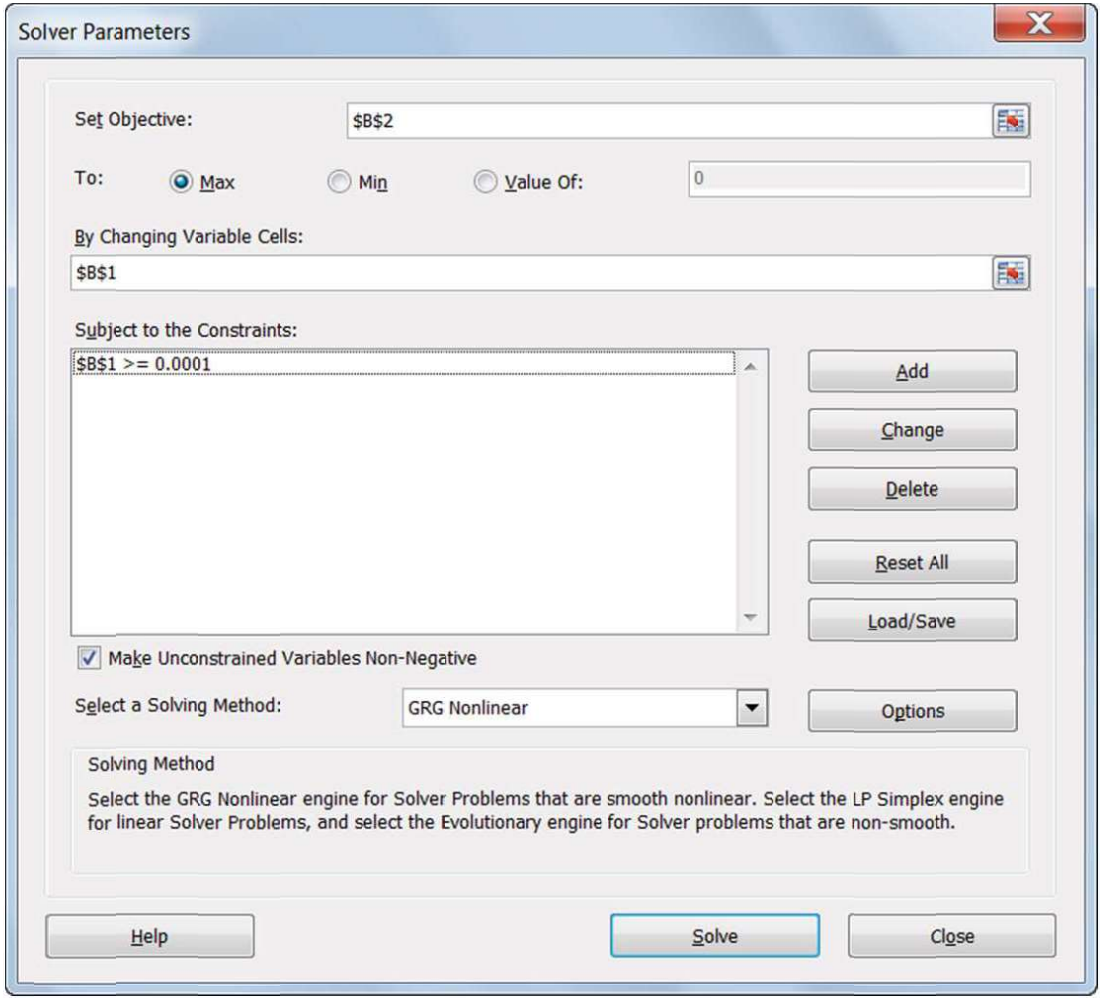
	A	B	C	D	E	F
1	theta	0.2				
2	LL	-1488.0				
3						
4	t	# Cust.	# Lost	P(die)	S(t)	
5	0	1000			1.0000	
6	1	631	369	0.2000	0.8000	-593.88
7	2	468	163	0.1600	0.6400	-298.71
8	3	382	86	0.12		-176.79
9	4	326	56	0.1024	0.4096	-127.62
10						-290.98
11						

=SUM(F6:F10)

=C6*LN(D6)

=B9*LN(E9)

Estimating Model Parameters



	A	B	C	D	E	F
1	theta	0.272				
2	LL	-1451.2				
3						
4	t	# Cust.	# Lost	P(die)	S(t)	
5	0	1000			1.0000	
6	1	631	369	0.2717	0.7283	-480.88
7	2	468	163	0.1979	0.5305	-264.09
8	3	382	86	0.1441	0.3864	-166.60
9	4	326	56	0.1050	0.2814	-126.23
10	5			0.0764	0.2050	-413.36
11	6			0.0557	0.1493	
12	7			0.0406	0.1087	
13	8			0.0295	0.0792	
14	9			0.0215	0.0577	
15	10			0.0157	0.0420	
16	11			0.0114	0.0306	
17	12			0.0083	0.0223	

Survival Curve Projection