DATASCIENCE

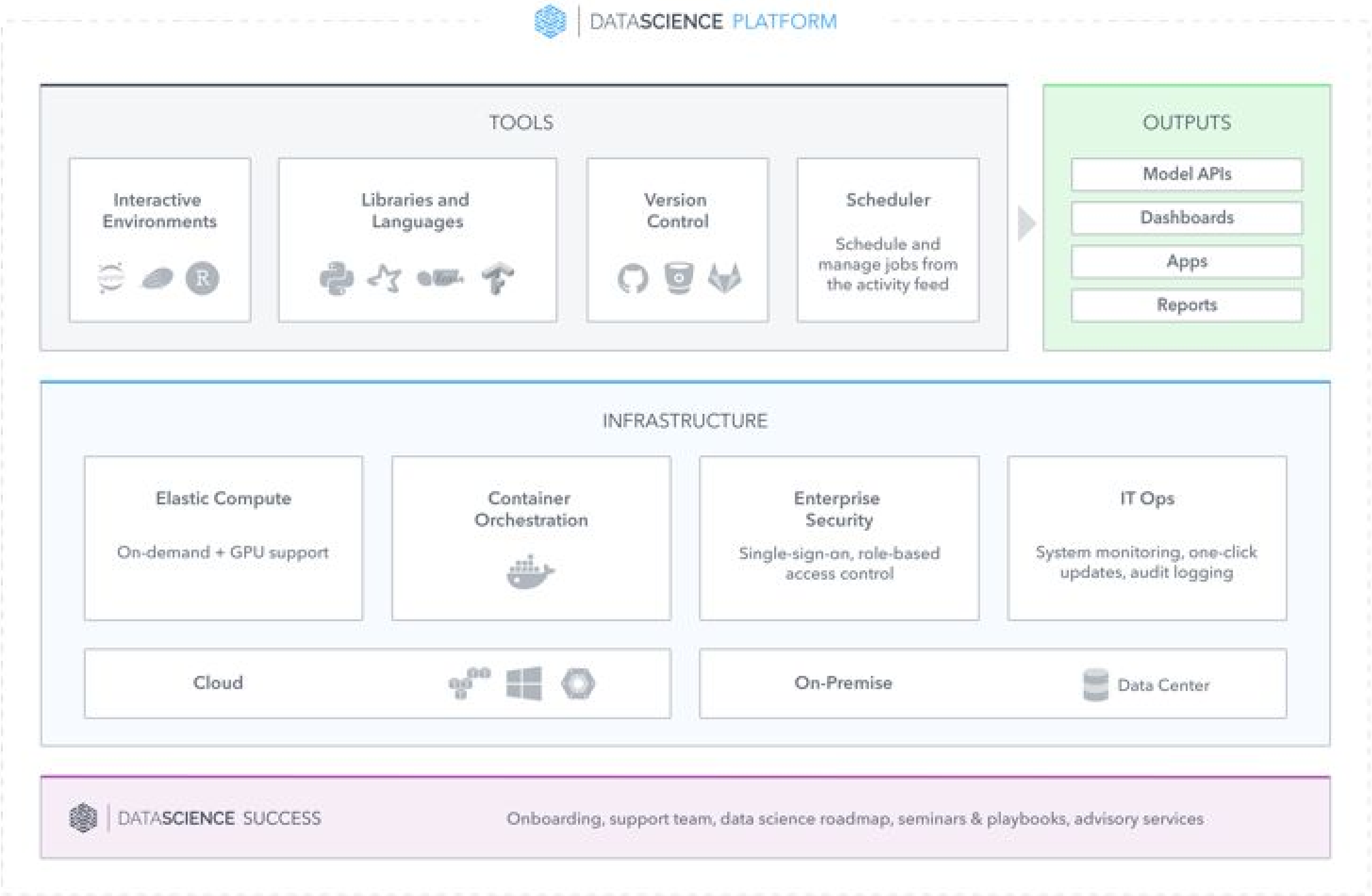
Modeling Customer Lifetime Value

Ben Van Dyke & Jean-Rene Gauthier

Data Scientists

DataScience.com

PyData Seattle, July 6 2017



The DataScience.com Platform combines the tools, libraries, and languages your team loves with the infrastructure and workflows your organization needs.

Overview

- Background
    - What is CLV and why do we care?
    - Quick review of different CLV contexts and models
- Deep dive into Pareto/NBD Probabilistic Model
- Using the model for business insights and action
- Lab: Implement Pareto/NBD Model in a jupyter notebook using pymc (https://github.com/datascienceinc/pydata-seattle-2017)

What is Customer Lifetime Value (CLV)?

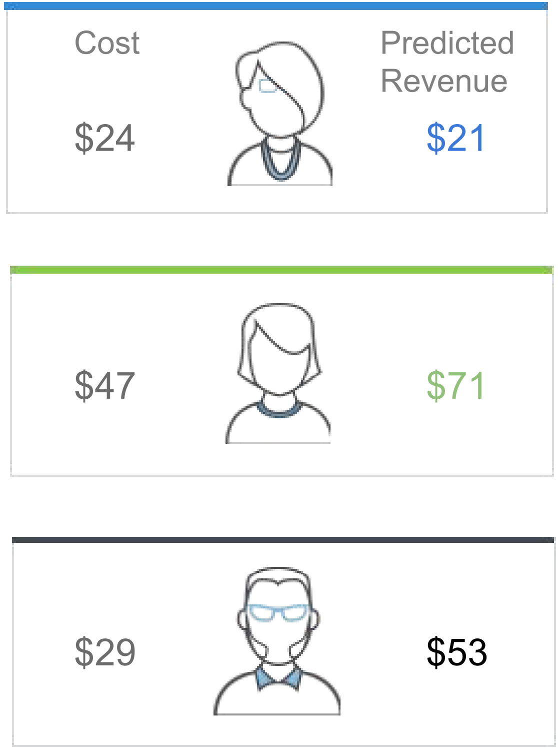
- Total profit of the entire relationship with a customer

  - Costs to attract, service, and maintain customer
  - Customer transactions (number and value)
  - Customer network effects (e.g. word-of-mouth)

Focus on long-term health of customer relationship

Why do we care about CLV?

- Customer segmentation to identify most profitable customers
- Identify traits and features of valuable customers
- Determine how to allocate resources among customers
- Enable evaluation of what a company should pay to acquire the customer relationship



Business Contexts

Contractual

- Customer ‘death’ can be observed
- Often modeled using survival-based approaches

Non Contractual

- Customer ‘death’ is unobserved
- Customer lifetime distribution often modeled via exponential models

Discrete versus continuous purchases

Discrete purchases occur at fixed periods or frequencies

Continuous purchases can happen at any time

Examples of Business Contexts

	Non-contractual Settings	Contractual Settings
Continuous Purchases	• movie rentals • medical appointments • hotel stays • grocery purchases • amazon.com	• Costco membership • credit cards
Discrete Purchases	• prescription refills • charity fund drives • event attendance	• magazine/newspaper subscriptions • fitness clubs • most insurance policies • streaming services: netflix, hulu, etc. • most cell phone plans

CLV Equation

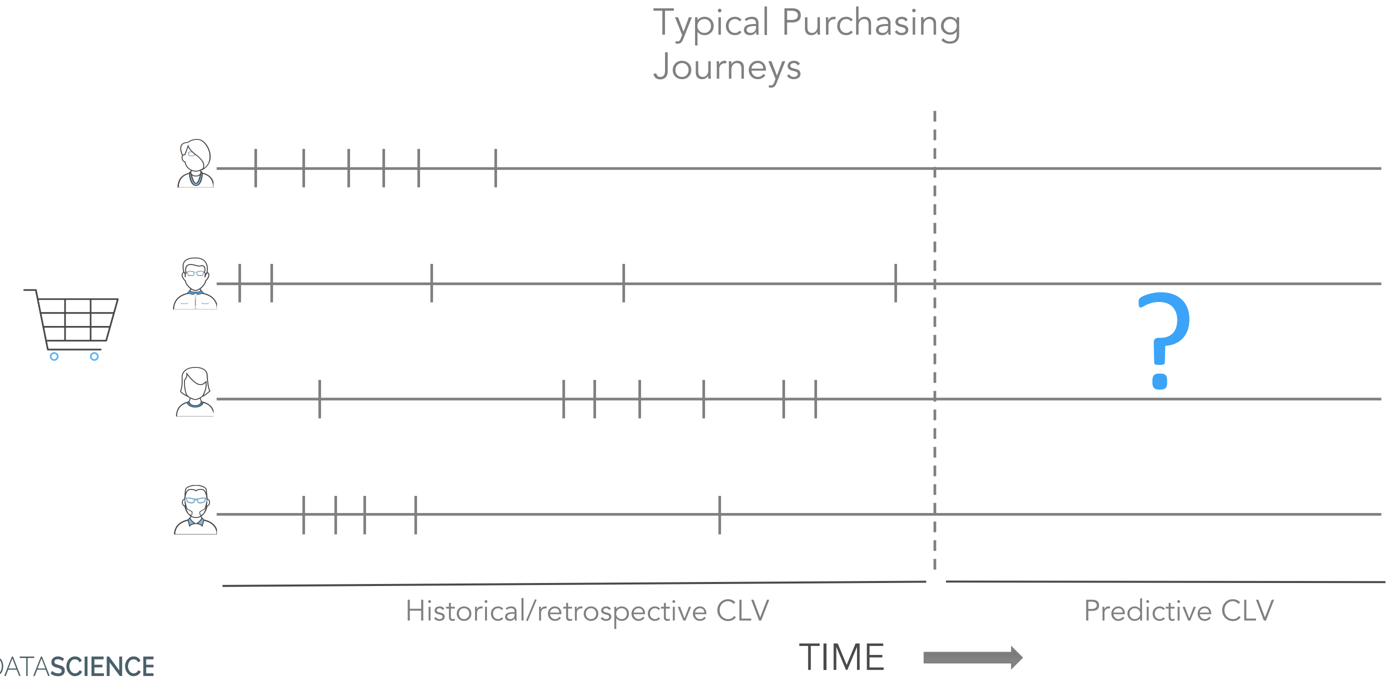
CLV = $\boxed{\text{Total number of purchases for each customer}}$ * Value of each future transaction at the customer level

Pareto/NBD CLV Model

A Hierarchical Bayesian Model

Customer Transactional Data

Transactional data from each customer is used to predict that customer's CLV



Dimensions of CLV Modeling

Purchase Count

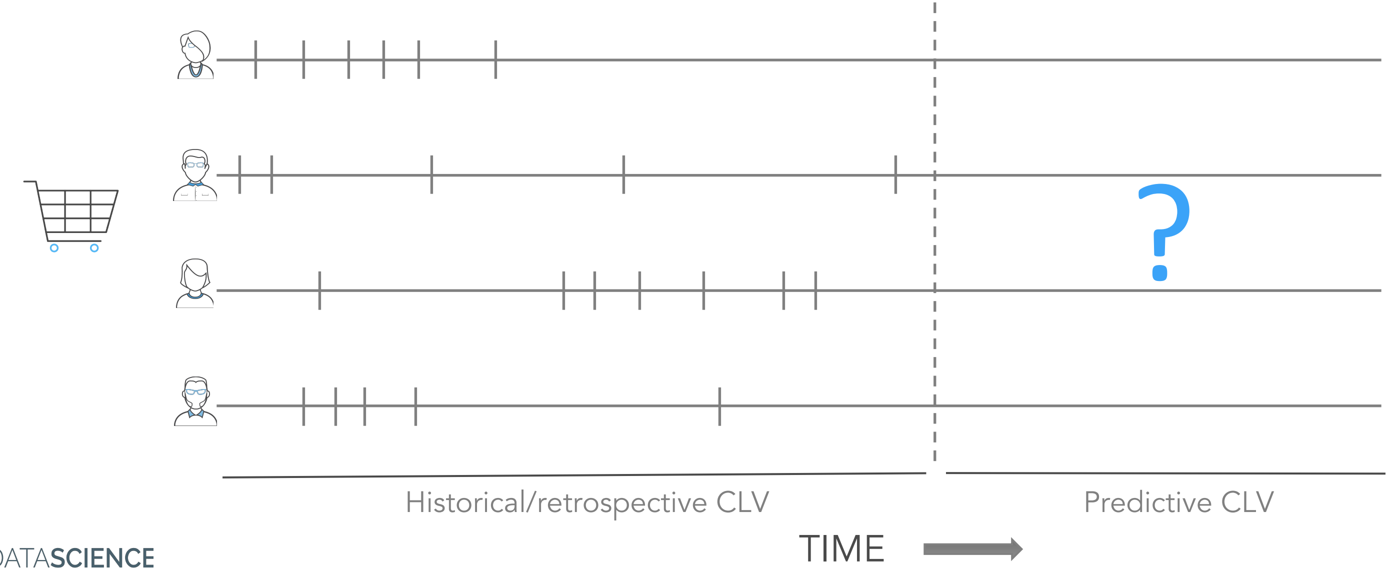
purchases in a given time window

Modeled with latent parameter **λ**

Lifetime

The customer's predicted lifetime

Modeled with latent parameter **μ**



Probabilistic Models

Purchase Count

purchases in a given time window

Modeled with latent parameter $\boldsymbol{\lambda}$

Lifetime

The customer's predicted lifetime

Modeled with latent parameter $\boldsymbol{\mu}$

Hierarchical:

Individual customer parameters, $\boldsymbol{\theta}$ (e.g. λ, μ, etc) constrained by and drawn from a population level distribution with parameters $\boldsymbol{\varphi}$

Probabilistic Models

Purchase Count

purchases in a given time window

Modeled with latent parameter **λ**

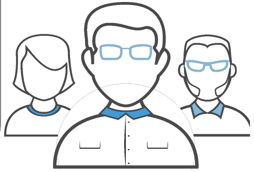
Lifetime

The customer's predicted lifetime

Modeled with latent parameter **μ**

Priors on Latent Parameters

The prior distributions represent our belief on how the latent parameters are distributed in the customer population.



Hierarchical:
Individual customer parameters, $\boldsymbol{\theta}$ (e.g. λ, μ, etc) constrained by and drawn from a population level distribution with parameters $\boldsymbol{\varphi}$

Probabilistic Models

Purchase Count

Poisson Distribution

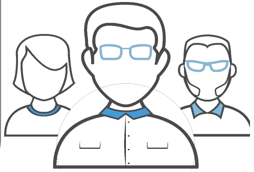
Lifetime

Exponential Distribution

Combined Models

Pareto/NBD model
- Pareto: exponential x gamma
- NBD: poisson x gamma

Gamma-Gamma: gamma x gamma



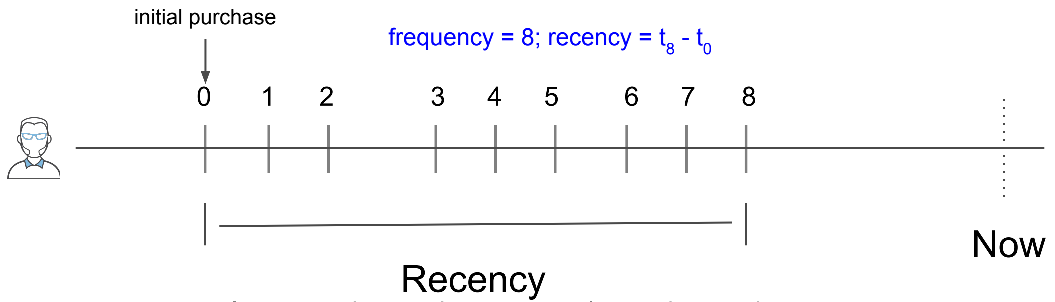
$$P(\theta|Data, \phi) \propto P(Data|\theta)P(\theta|\phi)$$

Pareto / **NBD** *(Schmittlein et al. 1987)*

Training The Pareto/NBD Model

The Recency-Frequency-Monetary Value (RFM) Data Structure

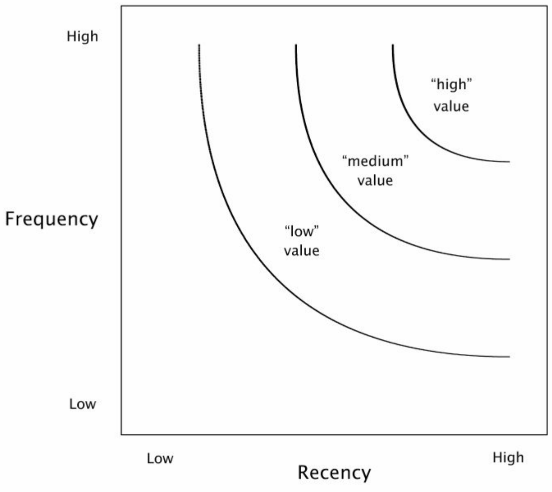
- What data structure do I need to train a Pareto/NBD model?



Recency = last purchase date - initial purchase date = t_8-t_0

(Repeat) Frequency = Number of purchases not counting first = 8

T = Last date - first purchase date = t_{now}-t_0



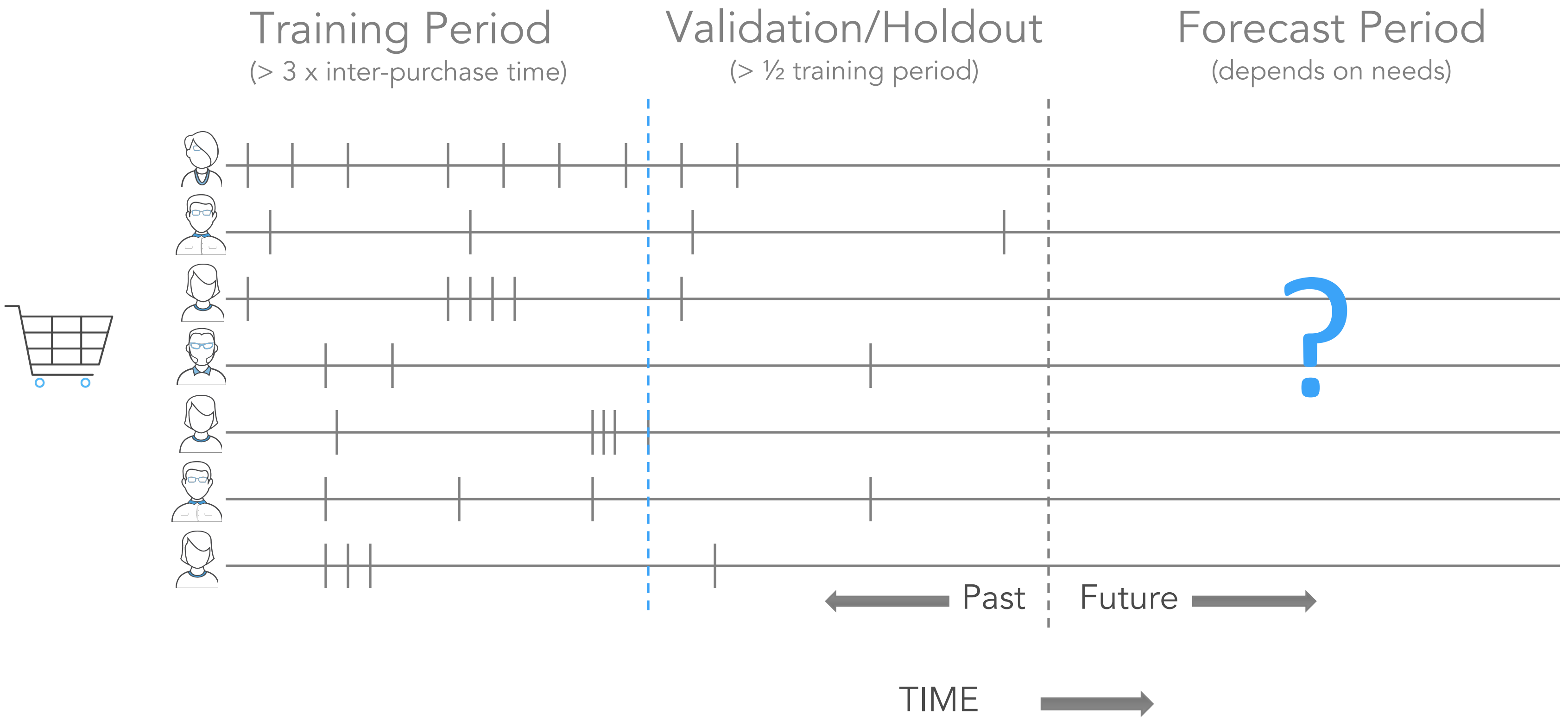
(Fader et al. 2004)

- Pareto/NBD and other models only require an RFM data structure (at the individual level) to be trained.

(Fader et al. 2004; Fader et al. 2005)

Jupyter Notebook : Generating an RFM object

Training the Pareto/NBD Model

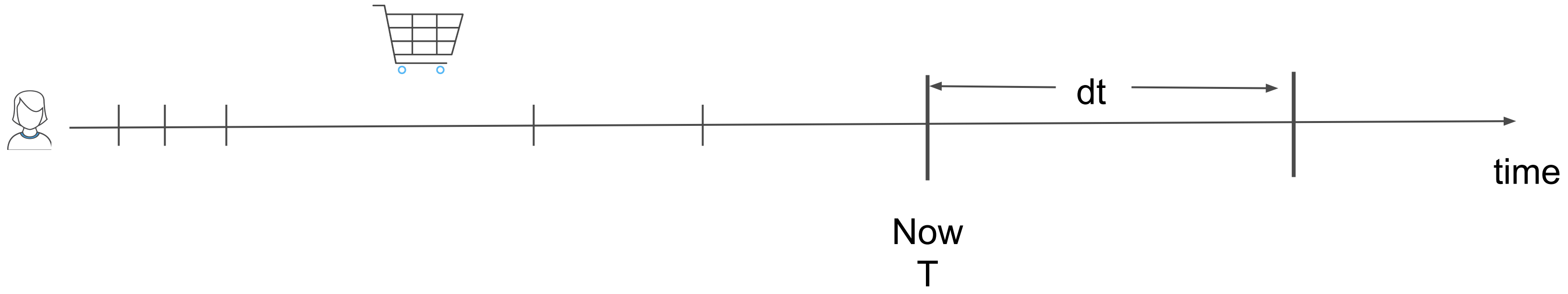


Jupyter Notebook : Training P/NBD model

Why is this Model Useful?

(Forecasting future purchases)

The Number of Future Purchases



- For each pair of (λ, μ) in the MCMC chain, one computes :

Number of future purchases = Number of future purchases given customer-level latent parameters and that customer is alive X Probability of being alive at T

$$\mathbf{E(N(dt) \mid \lambda, \mu, x, t_x, T) = E(N(dt) \mid \lambda, \mu, \text{alive at } T) \;\; x \;\; P(\tau > T \mid \lambda, \mu, x, t_x, T)}$$

The Number of Future Purchases

Number of future purchases given customer-level latent parameters and that customer is alive

$$E[N(dt) \mid \lambda, \mu, \text{alive at T}] = \frac{\lambda}{\mu}\left(1 - \mathrm{e}^{-\mu \mathrm{dt}}\right)$$

Probability of being alive at T

$$P(\tau > T \mid \lambda, \mu, x, t_x, T) = \frac{1}{1 + \mu/(\mu + \lambda)\left[e^{(\lambda+\mu)(T-t_x)} - 1\right]}$$

Jupyter Notebook : Actionable insights

In Summary

- What CLV models are and why they are useful
- No one-size-fits-all. Most CLV models are applicable in very specific business contexts.
- Intro to probabilistic CLV models in the non-contractual settings
  - Pareto/NBD Model
  - Training steps
  - Actionable quantities

Thank You!

Probabilistic Models

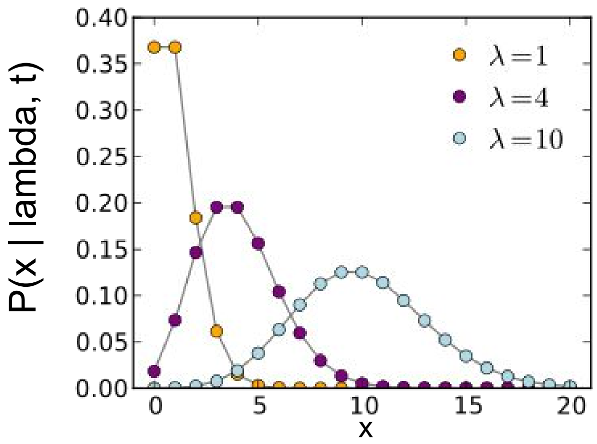
Purchase Count

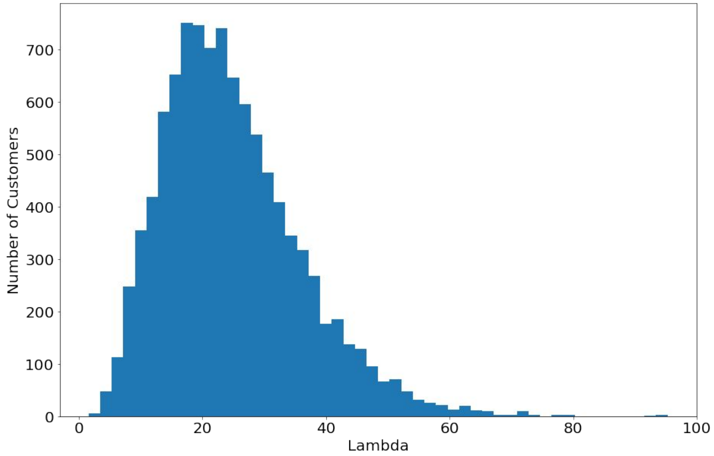
Poisson Distribution

$$P(x|\lambda, t) = \frac{(\lambda t)^x \exp -(\lambda t)}{x!}$$

x: number of purchases
t: length of time window
λ: number of purchases per unit of t, distributed as g(λ | r, α)

τ:
μ:
dis





Hierarchical:

Individual customer parameters, $\boldsymbol{\theta}$ (e.g. λ, μ, etc) constrained by and drawn from a population level distribution with parameters $\boldsymbol{\varphi}$

$$P(\theta|Data, \phi) = \frac{P(Data|\theta)P(\theta|\phi)}{P(Data|\phi)} \propto P(Data|\theta)P(\theta|\phi$$

DATASCIENCE

Historical vs. Predictive CLV: The Pitfalls of Historical LTV

Some businesses may use historical heuristics to model lifetime value; however, there are significant limitations to such an approach.

Limitations to heuristic model:

"most companies apply recency of last purchase (hiatus) analyses to distinguish between active and inactive customers […] and average past purchase behavior is employed as a simple predictor of future behavior."

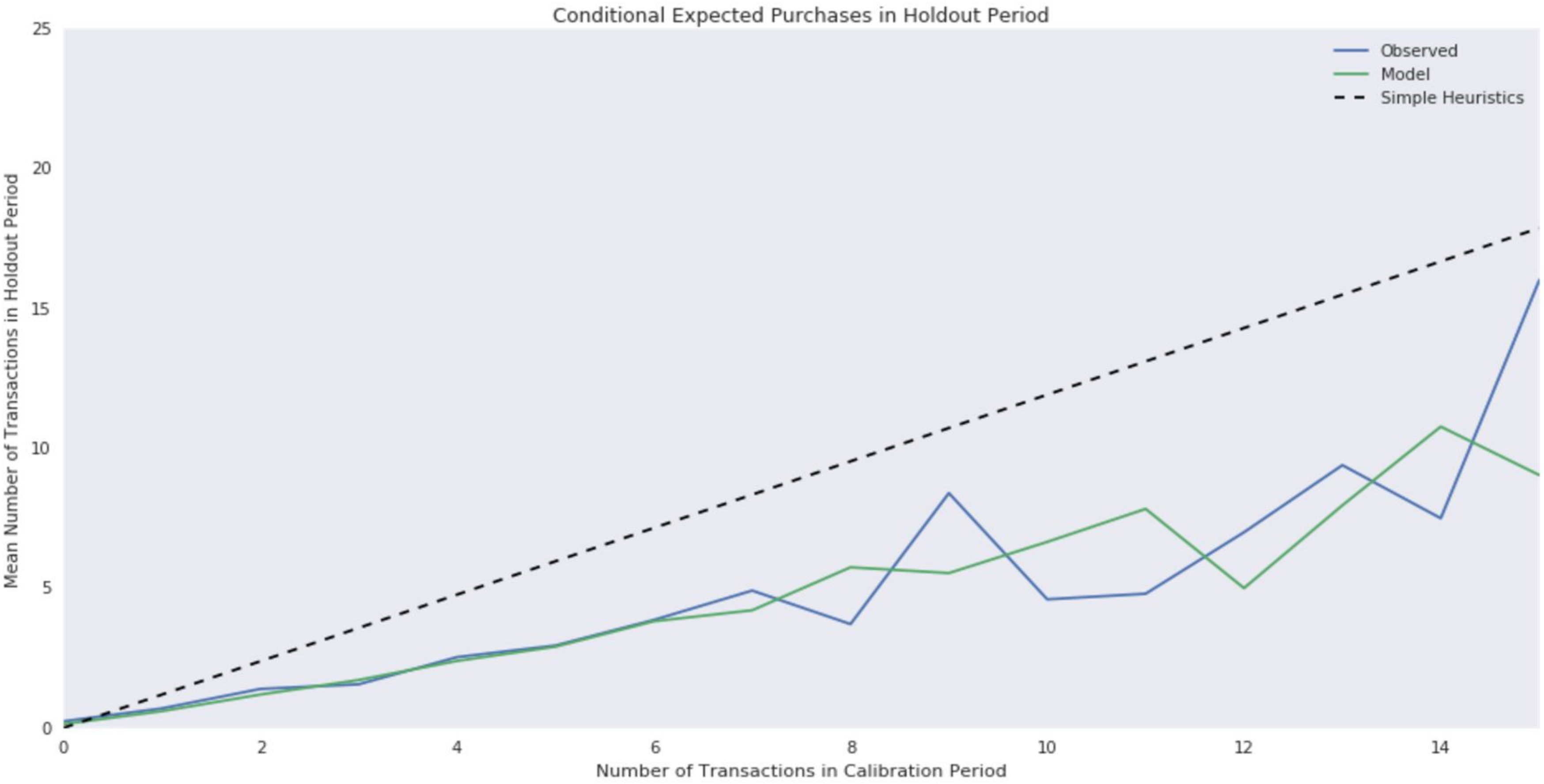
- Wubben & Wangenheim (2008)

Probabilistic approach is necessary:

Accounting for variation in the behavior of customers allows us to arrive at more accurate conclusions about customer lifetime and purchase behavior.

An Overview of the Modeling Techniques

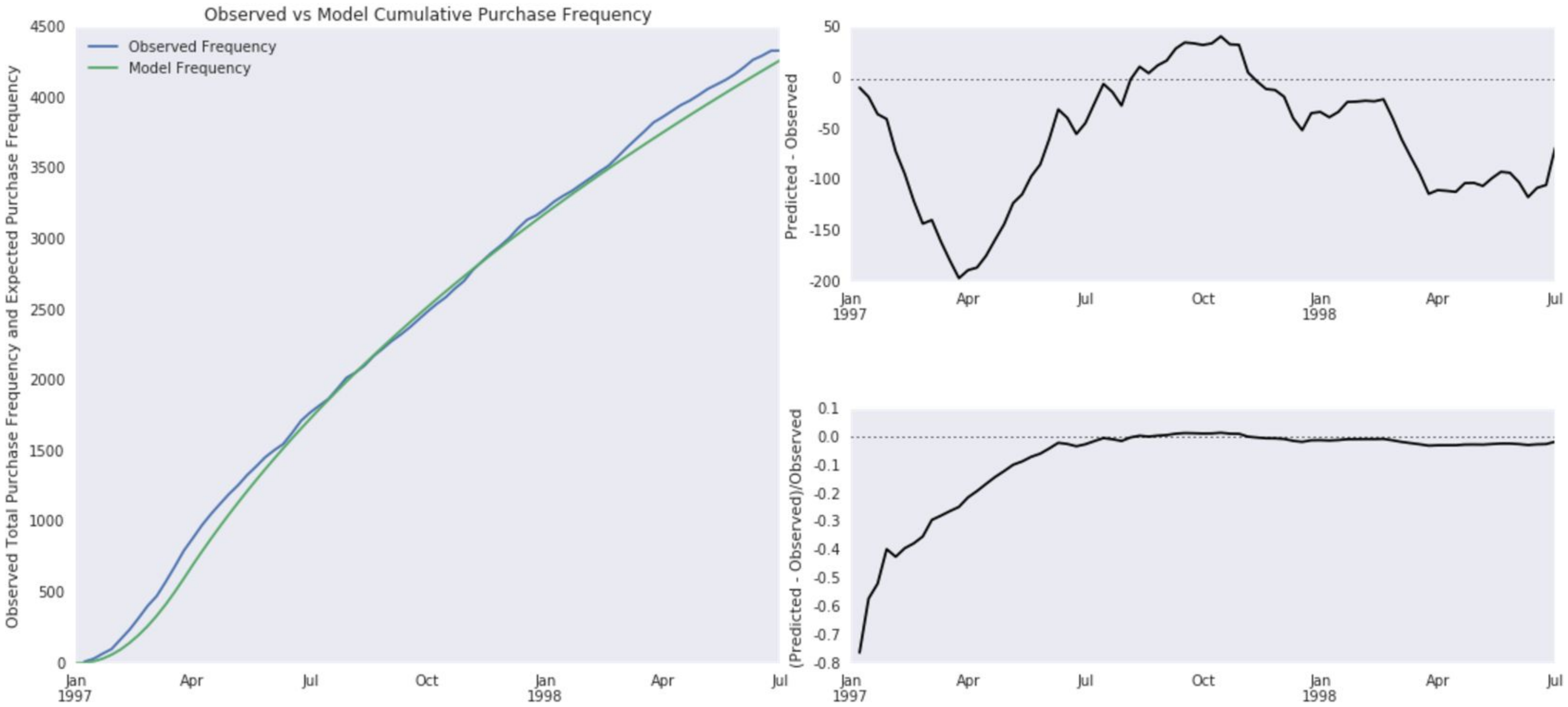
	Non-contractual Settings	Contractual Settings
Continuous Purchases	**Probabilistic Models** ● Pareto/NBD (Schmittlein et al. 1987) ● BG/NBD (Fader et al. 2005) ● Pareto/GGG (Platzer & Reutterer 2016) ● Gamma Gompertz G/G/NBD model (Bemmaror & Glady 2012) ● PDO model (Jerath et al. 2011) ● MCMC Pareto/NBD and GG/NBD (Ma & Liu 2007) ● Hierarchical Bayes (Abe 2009) **Machine Learning Models** ● CART + logit/linear regression (Jamal & Zhag 2009) ● Hybrid methods (Tsai et al. 2013) **Markov Models** ● Partially Hidden Markov models (Romero et al. 2008) ● Markov Chain model (Cheng et al. 2011)	● Exponential-gamma models (Hardie et al. 1998) ● Weibull-Gamma models (Morrisson & Schmittlein 1980) ● Hazard Models
Discrete Purchases	● BG/BB model ● Solicited Transactions Model (Colombo & Jiang, 1999) ● Basic structural model of CLV (Jain & Singh 2002)	● Shifted beta-geometric models (Kaplan 1982) ● Basic structural model of CLV (Jain & Singh 2002) ● Hierarchical Bayes models (Borle et al. 2008)


Conditional Expected Purchases in Holdout Period
Observed
Model
Simple Heuristics
Mean Number of Transactions in Holdout Period
Number of Transactions in Calibration Period
0
5
10
15
20
25
0
2
4
6
8
10
12
14

In [6]:
```
from ds_clv.plotting import plot_cumulative_build

# The plotting function will also return the cumulative build dataframe :
cumulative_build_df = plot_cumulative_build(pnbd, "1998-06-30", metric_to_compare='purchase_freque
ncy')
```

```
Relative Residuals at 1998-06-30 : -0.0157470677904
```



EVENT: LEARN DATA SCIENCE FROM EXPERTS AT GOOGLE AND NETFLIX REGISTER NOW ›

DATASCIENCE.COM

Platform Solutions Resources Education Company

REQUEST DEMO

A powerful platform for enterprise data science.

DataScience.com's platform empowers your team with open source tools, elastic infrastructure, and customizable workflows.

WATCH VIDEO

REQUEST DEMO