Subscription Survival Analysis in R

Jim Porzak

Senior Data Scientist, Viadeo.com
Contact author: jporzak@gmail.com

Keywords: marketing, survival, subscription, customer.

Survival analysis started in reliability engineering and medical research [1]. More recently it has been used by marketers to better understand customers in subscription based businesses [2, 3]. While the basic math doesn't change, there are practical differences when applied to marketing. First, N is huge. Secondly, there are generally many cohorts driven by marketing variations around product, offer, price, acquisition source, and various subscriber properties. Thirdly, the assumption of hazard stability over time certainly doesn't hold.

In order to demonstrate subscription survival techniques with non-proprietary data, we first needed a tool to generate pseudo-random subscriber data. subsurvGen generates data that closely mimics real-world situations based on the choice of generation parameters.

Since N is huge, we use non-parametric methods to calculate hazard and survival. Emphasis is placed on marketer-friendly display of results – both graphically and in tabular form. If subscription average sales price is known, the n-year long term revenue (LTR) is also calculated.

Tactical marketers need survival and LTR at the most granular cohort. Strategic marketers, on the other hand, are interested in LTR at global or regional level. subsurvRollup supports flexible roll-ups across multidimensional cohorts to any level.

Hazard space is much more aligned with the marketer's world. Their actions to improve retention are measured by changes in hazard. “What-if” modeling, like “if we reduced churn by 10%,” is supported by subsurvModel which gives the projected impact on n-year average tenure and LTR.

There is a challenging dilemma between the strategic marketer's need for a multi-year LTR – which is one limit on allowable acquisition cost – and the tactical need to measure short term changes in hazard. subsurvModel also extrapolates short term hazards out to n-years based on the historical decay in hazard space.

The presentation concludes with examples of how subscription survival analysis answered real-world marketing questions.

References

