Attribution Modeling:

Monte Carlo Markov Chain (MCMC),

Survival Analysis and Shapley Value

QuantFarm

What is a Markov Chain?

A Markov Chain is a statistical formula that makes predictions for the future of a process based solely on its present state and a probability distribution derived from the sequence of events preceding it.

$P(X_t = s_j | X_{t-i}$

$0 \ 5 \ w_0$

$\sum_{j=1}^{N} w_{ij} = 1 \ \forall i$

Simplified into 3 main parts:

- The Transition Probability (wij) = The Probability of the Previous State (Sequence A, Xt- Given the Current State (Sequence B, Xt)
- The Transition Probability (wij) is No Less Than 0 and No Greater Than 1
- The Sum of the Transition Probabilities Equals 1 (Everyone Must Go Somewhere)

Using Markovian Chains for Attribution Modeling

Proof of Concept

1. Assume three customer journeys

Channel1 -> Channel2 -> Channel3 -> purchase!

Channel1 -> unsuccessful conversion

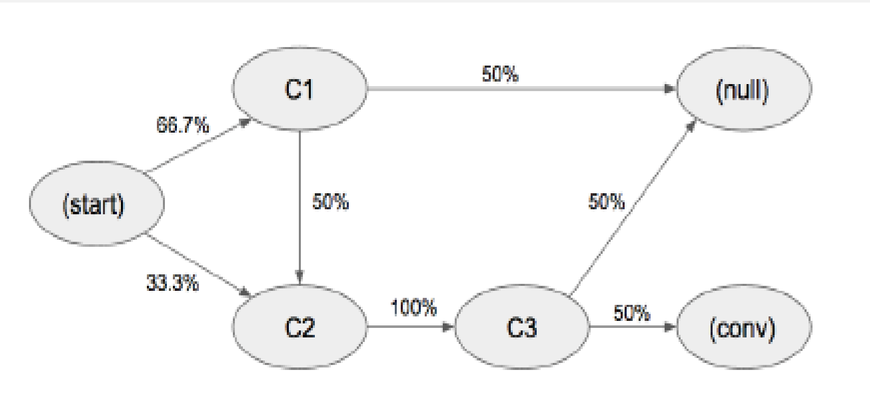
Channel2 -> Channel3 -> unsuccessful conversion

2. Add states, split pairs, transform, calculate probabilities...

FROM	TO	PROBABILITY	TOTAL PROBABILITY
(start)	C1	1/3	66.70%
(start)	C1	1/3	
(start)	C2	1/3	33.30%
total from (start)		3/3	
C1	C2	1/2	50%
C1	(null)	1/2	50%
total from C1		2/2	
C2	C3	1/2	100%
C2	C3	1/2	
total from C2		2/2	
C3	(conversion)	1/2	50%
C3	(null)	1/2	50%
total from C3		2/2	

Based on model, we distribute conversion revenue across channels

3. Generate model



MONTE CARLO MARKOV CHAIN

History of MONTE CARLO SIMULATION

- Applied in 1940 by the scientist while working on Atomic Bomb.
- They used it to calculate probabilities of fission uranium reaction with one another.
- With uranium in short supply, there was little room for experimental trial and error.
- Scientist discovered that enough simulated data could compute reliable probabilities & reduce the amount of uranium needed for testing.

MONTE CARLO MARKOV CHAIN

What is MONTE CARLO SIMULATION?

- Technique that uses simple random sampling to simulate data to use with statistical models.
- Helps in designing better process by determining relationship between input & output variability.
- Sample randomly and aggregate the results into an estimate of what's going to happen.

Working of MONTE CARLO SIMULATIONS

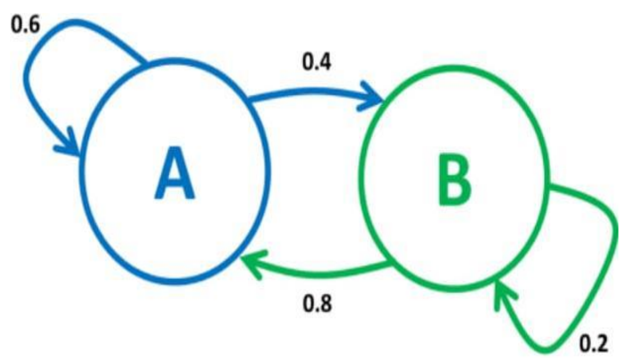
- Performs risk analysis by building models of possible results by substituting a range of values
- Calculates results using a different set of random values from the probability functions
- Perform thousands of recalculations before it is complete

MONTE CARLO MARKOV CHAIN

What is MONTE CARLO MARKOV CHAIN?

Markov Chain

- Mathematical systems that hop from one "state" to another.
- Provides probability of transitioning from one state to another.
- Based on Principle of "memorylessness“
- The next state of the process only depends on the previous state.



Monte Carlo Markov Chain

- A general purpose technique for generating fair samples for Markov Chain based on Probability distribution..

MONTE CARLO MARKOV CHAIN (MCMC) DATA

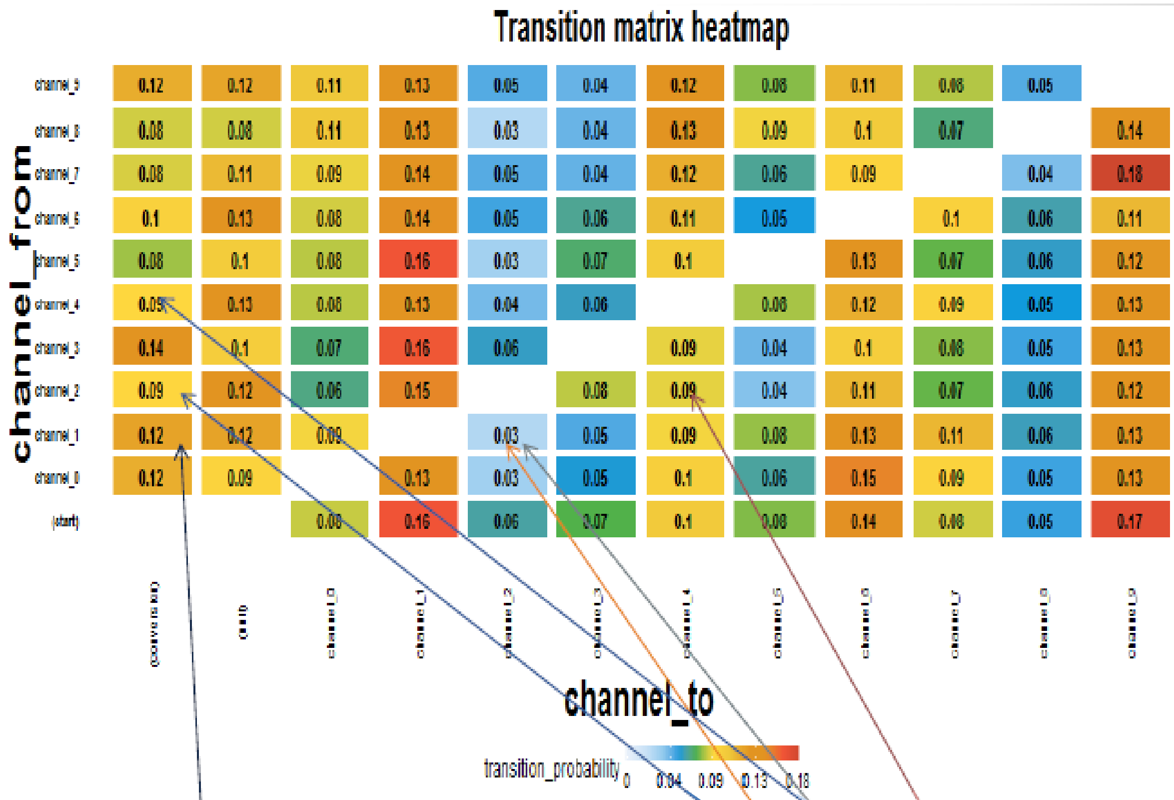
Path	Cony	Cony_null	Time	Last Touch
Channel 3	0	1	12	channel_3
channel 4 > channel 9 > channel 9 > channel 6 > channel 2 > channel_1 > channel _O	1	0	13	channel()
channel 7 > channel 9 > channel 9 > channel 8 > channel 8 > channel 9 > channel 5 > channel _O > channel	1	0	7	channel_O
channel_1 > channel 9 > channel _O > channel 6	0	1	11	channel_6
channel 4 > channel_6 > channel 4 > channel 9 > channel 2 > channel 7	0	1	9	channel_7
channel_1 > channel_1 > channel 4 > channel 7 > channel 6	0	1	2	channel_6
channel 6 > channel 5 > channel 6 > channel 6 > channel 7 > channel 7 > channel 6	1	0	9	channel_6
channel 4 > channel 5 > channel 5 > channel 3 > channel 5 > channel 6 > channel 8 > channel 4 > channel	0	1	2	channel_1
channel 4 > channel 8	1	0	1	channel_8
channel _O > channel 4 > channel_1 > channel_1	0	1	12	channel_1
channel 6 > channel _O > channel 6	0	1	5	channel_6
channel 9 > channel_1 > channel 4 > channel_9 > channel 6	0	1	2	channel_6
channel 6 > channel 6 > channel 6	0	1	13	channel_6
channel _O > channel 6 > channel_1 > channel 6	0	1	2	channel_6
channel _O > channel 6 > channel 7 > channel 4 > channel 7 > channel 2 > channel 7	1	0	12	channel_7

"Path" containing customer paths

"Cony" containing whether conversion has taken place or not
"Cony null" containing paths that do not lead to conversion
"Time" is time to conversion
"Last Touch" is last touch marketing Channel

MONTE CARLO MARKOV CHAIN

TRANSITION PROBABILITY MATRIX FOR CHANNELS



channel name total conversions

1	channel_3	36.30745
2	channel_4	51.93238
3	channel_9	65.48897
4	channel_6	58.35964
5	channel_2	26.89489
6	channel_1	66.62196
7	channel_0	47.65405
8	channel_7	45.94396
9	channel_8	31.46333
10	channel_5	38.33338

- ❖ Total estimated no of conversions for Channel_3 alone is 36
- ❖ From the Transition Probability Matrix, Probability of conversion for Channel is 12%
- ❖ Probability of conversion for Path Channel_1>Channel_2>conversion is 0.27% (0.03.0.09)
- ❖ Probability of conversion for Path Channel_1>Channel_2>Channel_4>conversion is 0.024%(0.03*0.09*0.09)

SURVIVAL ANALYSIS

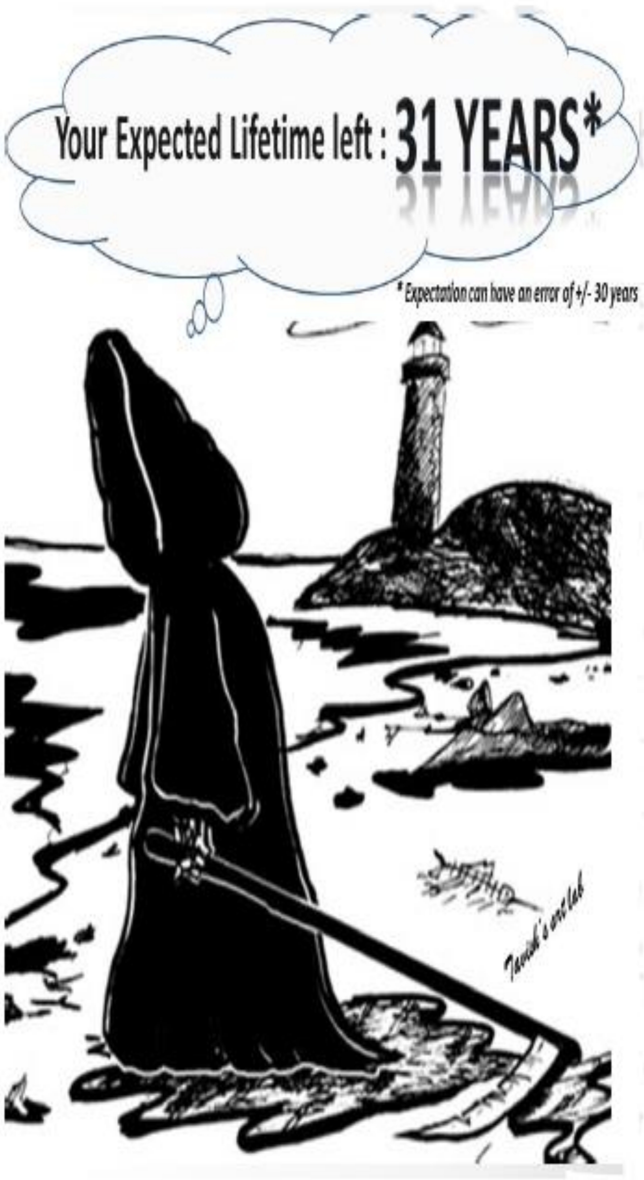
SURVIVAL ANALYSIS

WHAT IS SURVIVAL ANALYSIS?

- A set of methods for analysing data where the outcome variable is the time until the occurrence of an event of interest.
- Subjects are usually followed over a specified time period and the focus is on the time at which the event of interest occurs.
- Can handle censored data

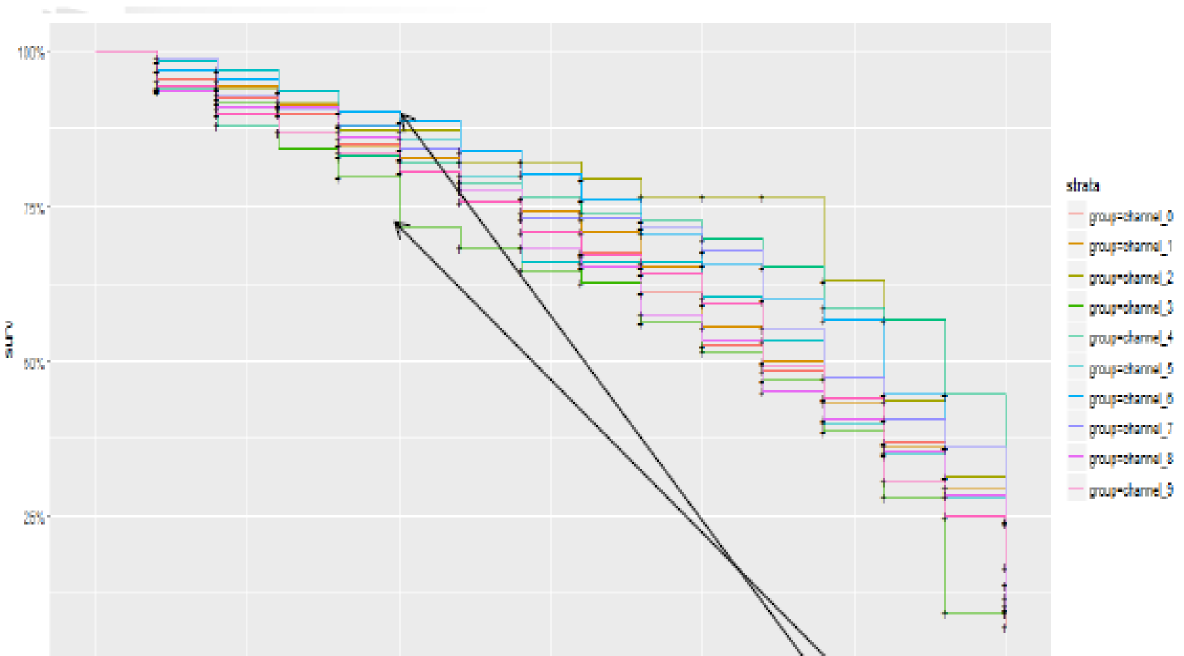
- Dependent Variable consist of following two parts:

1. Time to Event
2. Event Status



STATISTICAL MODEL: SURVIVAL

Non Parametric Survival Model For Different Channels



group=channel_0

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	92	4	0.9565	0.0213	0.9157	0.999
2	86	3	0.9232	0.0279	0.8700	0.980
3	80	2	0.9001	0.0316	0.8402	0.964
4	75	4	0.8521	0.0380	0.7808	0.930
5	70	2	0.8277	0.0406	0.7519	0.911
6	62	3	0.7877	0.0447	0.7047	0.880
7	57	4	0.7324	0.0494	0.6417	0.836
8	50	4	0.6738	0.0534	0.5768	0.787
9	43	4	0.6111	0.0569	0.5092	0.734
10	35	5	0.5238	0.0607	0.4174	0.657
11	27	2	0.4850	0.0621	0.3774	0.623
12	22	2	0.4409	0.0638	0.3320	0.586
13	18	3	0.3674	0.0658	0.2587	0.522
14	13	3	0.2826	0.0664	0.1784	0.448
15	8	6	0.0707	0.0463	0.0195	0.256

- For the Channel _3 probability of conversion on 5th day is 71%.
- For the Channel _6 probability of conversion on 5th day is 88%
- For the Channel _0 probability of conversion on 2nd day is 92%.
- For the Channel _0 probability of conversion on 4th day is 85%

STATISTICAL MODEL: SURVIVAL

Semi-Parametric Survival Analysis

```
Call:
coxph(formula = surdal ~ ., data = sub, ties = "breslow")

  n= 742, number of events= 345

                     coef exp(coef) se(coef)      z Pr(>|z|)
datasurchannel_1 -0.06486   0.93726  0.21949 -0.295    0.768
datasurchannel_2 -0.43550   0.64694  0.33055 -1.317    0.188
datasurchannel_3  0.23622   1.26646  0.26116  0.905    0.366
datasurchannel_4 -0.28994   0.74831  0.24081 -1.204    0.229
datasurchannel_5 -0.16991   0.84374  0.28279 -0.601    0.548
datasurchannel_6 -0.33460   0.71563  0.23699 -1.412    0.158
datasurchannel_7 -0.15997   0.85217  0.26095 -0.613    0.540
datasurchannel_8  0.16201   1.17587  0.29170  0.555    0.579
datasurchannel_9  0.03011   1.03057  0.22062  0.136    0.891
```

- In comparison to last touch Marketing Channel _0, "Channel_1" decreases the chance of conversion by 0.93 times
- In comparison to last touch Marketing Channel _0, Channel_3 increases the chance of conversion by 1.26 times

```
                 exp(coef) exp(-coef) lower .95 upper .95
datasurchannel_1    0.9372     1.0670    0.6095     1.441
datasurchannel_2    0.6469     1.5457    0.3385     1.237
datasurchannel_3    1.2665     0.7896    0.7591     2.113
datasurchannel_4    0.7483     1.3363    0.4668     1.200
datasurchannel_5    0.8437     1.1852    0.4847     1.469
datasurchannel_6    0.7156     1.3974    0.4497     1.139
datasurchannel_7    0.8522     1.1735    0.5110     1.421
datasurchannel_8    1.1759     0.8504    0.6638     2.083
datasurchannel_9    1.0306     0.9703    0.6688     1.588

Concordance= 0.554  (se = 0.019 )
Rsquare= 0.015   (max possible= 0.995 )
Likelihood ratio test= 11.19  on 9 df,   p=0.2627
Wald test            = 11.21  on 9 df,   p=0.2617
```

STATISTICAL MODEL: SURVIVAL

Parametric Survival Analysis

```
[1] "best parametric model is Gaussian"

Call:
survreg(formula = surdal ~ ., data = sub, dist = "gaussian")
                  Value Std. Error      z        p
(Intercept)      10.912     0.7927 13.765 4.12e-43
datasurchannel_1  0.164     0.9605  0.171 8.64e-01
datasurchannel_2  1.734     1.3789  1.257 2.09e-01
datasurchannel_3 -0.835     1.1626 -0.718 4.73e-01
datasurchannel_4  0.450     1.0208  0.441 6.59e-01
datasurchannel_5  0.554     1.1878  0.466 6.41e-01
datasurchannel_6  1.330     1.0160  1.309 1.90e-01
datasurchannel_7  0.790     1.1103  0.712 4.77e-01
datasurchannel_8 -1.317     1.2750 -1.033 3.02e-01
datasurchannel_9 -0.553     0.9642 -0.573 5.67e-01
Log(scale)        1.658     0.0381 43.553 0.00e+00

Scale= 5.25

Gaussian distribution
Loglik(model)= -1242.9   Loglik(intercept only)= -1248.7
        Chisq= 11.42 on 9 degrees of freedom, p= 0.25
Number of Newton-Raphson Iterations: 3
n= 742
```

- In comparison to last touch Marketing Channel _0, "Channel_1" increases the time to conversion by 0.164 units
- In comparison to last touch Marketing Channel _0, Channel_3 decreases the time to conversion by 0.835 units

SHAPLEY VALUE

Shapley Value: An Introduction

A Game theory based axiomatic value attribution to assign credits to each player who are cooperating

Structure data to reflect conversion performance of all unique combination of channels

Identify baseline 'importance' value for each campaign that represents the expected Conversion performance

Run series of regressions comparing the importance of each campaign with each of the others as a pair, triplet, or higher order combination

Advantages

- Fast emerging as a standard for attribution modeling in marketing
- Provides stable, statistically reliable & consistent attribution weights
- Addresses the problem of multi-collinearity between independent variables in the model by providing an accurate decomposition of the total variance explained (R2).
- Handles changes in sample without leading to unstable and fluctuating estimates

An example: Shapley Value based attribution

Unique Path	Transactions
SEM	100
Display	125
SEO	50
SEM → Display	270
SEO → SEM	375
SEO → Display	350
SEM → Display → SEO	500

- Evaluating the contribution of each channel in the path that contributed to 500 transaction

Calculating the Shapley Value

Shapley Value calculation for all possible combinations

SEM 100	Display 270-100=170	SEO 500-270=230
SEM 100	SEO 375-100=275	Display 500-375=125
Display 125	SEM 270-125=145	SEO 500-270=230
Display 125	SEO 350-125=225	SEM 500-350=150
SEO 50	Display 350-50=300	SEM 500-350=150
SEO 50	SEM 375-50=325	Display 500-375=125

Marginal Contribution

SEM

=1/6 x (100 + 100 + 145 + 150 + 150 + 325)
=161.66

Display

=1/6 x (170 + 125 + 125 + 125 + 300 + 125)
=161.66

SEO

=1/6 x (230 + 275 + 230 + 225 + 50 + 50)
=176.66

Attribution %		
SEM	Display	SEO
161.66/(161.66+161.66+176.66)	161.66/(161.66+161.66+176.66)	176.66/(161.66+161.66+176.66)
32%	32%	36%

THANK YOU