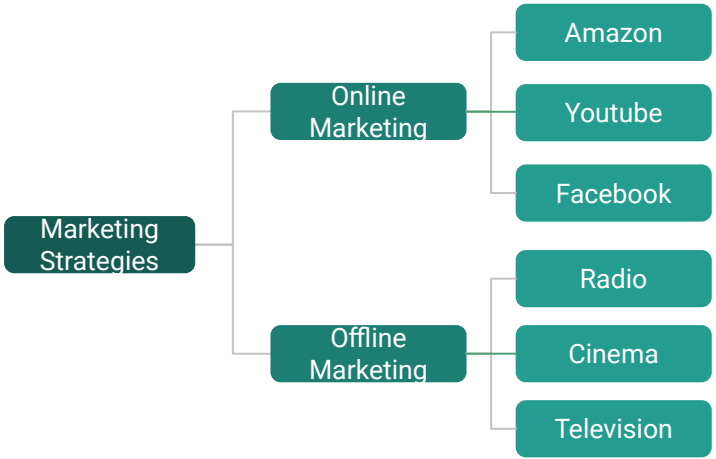
CS7290 Applied Bayesian Machine Learning Project

In Cooperation with Artefact

Mian Zhang

Background: A Marketing Problem



Discover the impact of various marketing channels on sales for Product A

Goal: Help to solve problems and improve their work

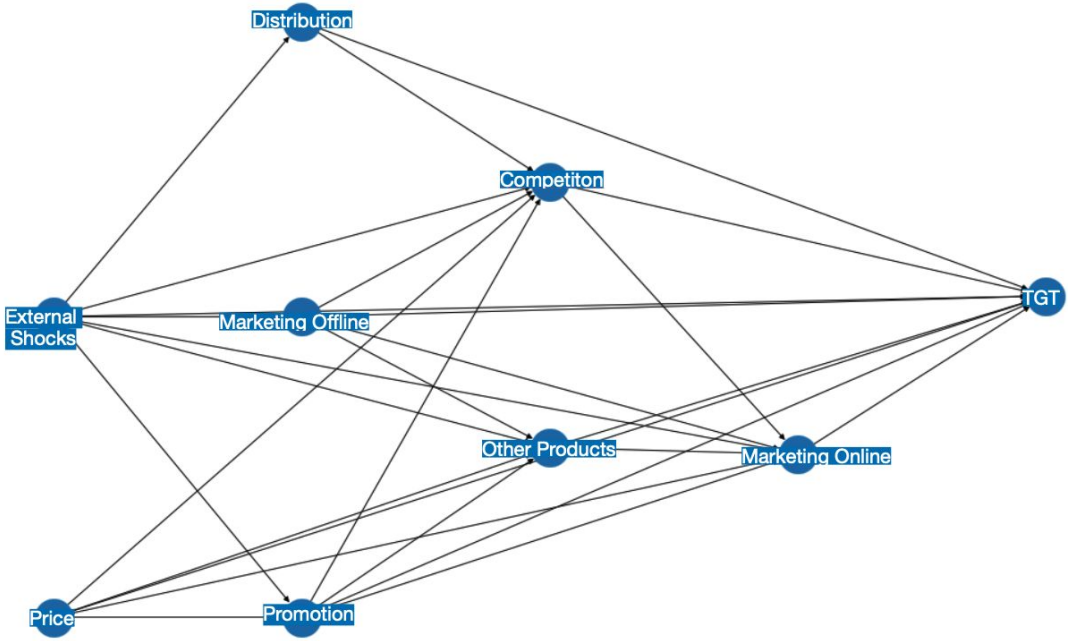
ARTEFACT
VALUE BY DATA

Problems:

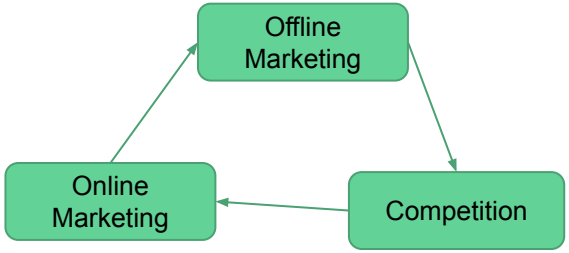
- How to deal with cyclicity in directed graph?
- What's the right metric to use for model quality control?
- How to choose the right priors for variables and for coefficients of the model?

There are two cycles in the directed graph

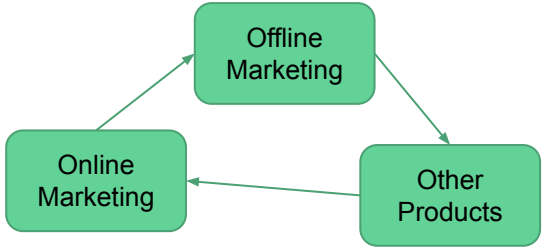
Business assumptions + statistical test = Directed Graph



Cycle 1



Cycle 2

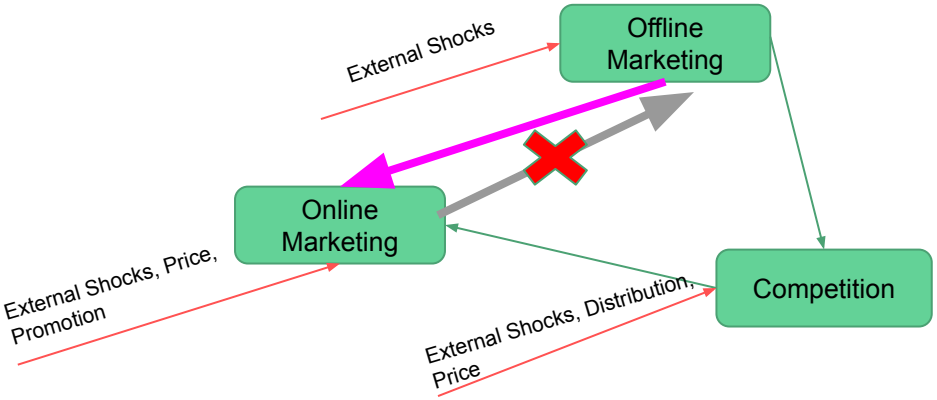


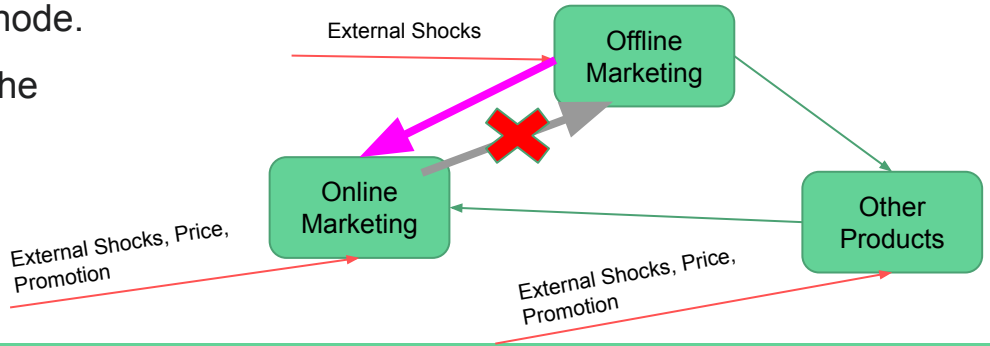
Topological Sorting can reduce cyclicity in DAG

Topological sorting is a graph traversal in which each node *v* is visited only after all its dependencies are visited.

Solution:

- Remove all lines among these nodes.
- Compare the degree of freedom of each node.
- Check if there are original lines violating the topological sorting rules.





Model Building

1. Define Mean(from dataset) and Sigma(from dataset) of variables.
2. Define Mu(normally distributed) of variables.
3. Define interaction(standard normally distributed) between variables based on DAG.
4. Assume variables are normally distributed with sigma and specific mu(its own mu + interaction) based on DAG.
    - promotion_mean = 1.1, promotion_sigma = 0.9, price_mean = 1.1, price_sigma = 0.9
    - promotion_mu = Normal(promotion_mean, promotion_sigma)
    - price_mu = Normal(pricce_mean, price_sigma)
    - price_to_promotion = Normal(0,1)
    - price = Normal(mu = price_mu, sigma = price_sigma)
    - promotion = Normal(**mu** = promotion_mu + price_to_promotion*price, **sigma** = promotion_sigma)

What's the right metric to evaluate models?

	Description	Pros	Cons
RMSE (Root Mean Square Error)	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$	Reveal the difference between predicted and observed data	Cannot detect overfitting issues
MAPE (Mean Absolute Percentage Error)	$\frac{100}{n}\sum_{t=1}^{n}\left\|\frac{A_t - F_t}{A_t}\right\|$		
R-Hat	Values close to one indicate convergence to the underlying distribution	Reveal the degree of convergence of a random Markov Chain	
WAIC (Widely Applicable Information Criteria)	Using the computed log pointwise posterior predictive density (LPPD) and correcting for the effective number of parameters to adjust for overfitting.	Reveal the loss of information by providing a trade-off between goodness of fit and model complexity in a Bayesian way	Absolute value has no meaning, need to be used for models comparison
LOO (Leave-one-out cross-validation)	A cross validation method which is applied once for each instance, using all other instances as a training set and using the selected instance as a single-item test set.		

How to choose the right priors? (Original Model)

- Original model defined **variables** as **Normal distribution** with **self-defined** mu and sigma, which is based on **domain knowledge**.
- Original model defined **coefficients** as **Standard Normal distribution** and **Half Normal Distribution**, which are weakly informative priors.

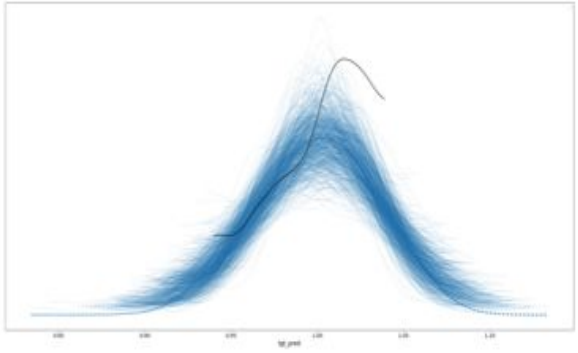
Real data

count	104.000
mean	1.000
std	0.026
min	0.940
25%	0.981
50%	1.006
75%	1.021
max	1.039

Simulated from prior distribution

count	10400.000
mean	0.360
std	7.012
min	-30.083
25%	-3.495
50%	-0.247
75%	3.661
max	62.950

Posterior Predictive Check



Evaluation Metrics

R-hat	WAIC	LOO
1.00~1.02	-804.72	-772.19

Weakly Informative ---> More Informative

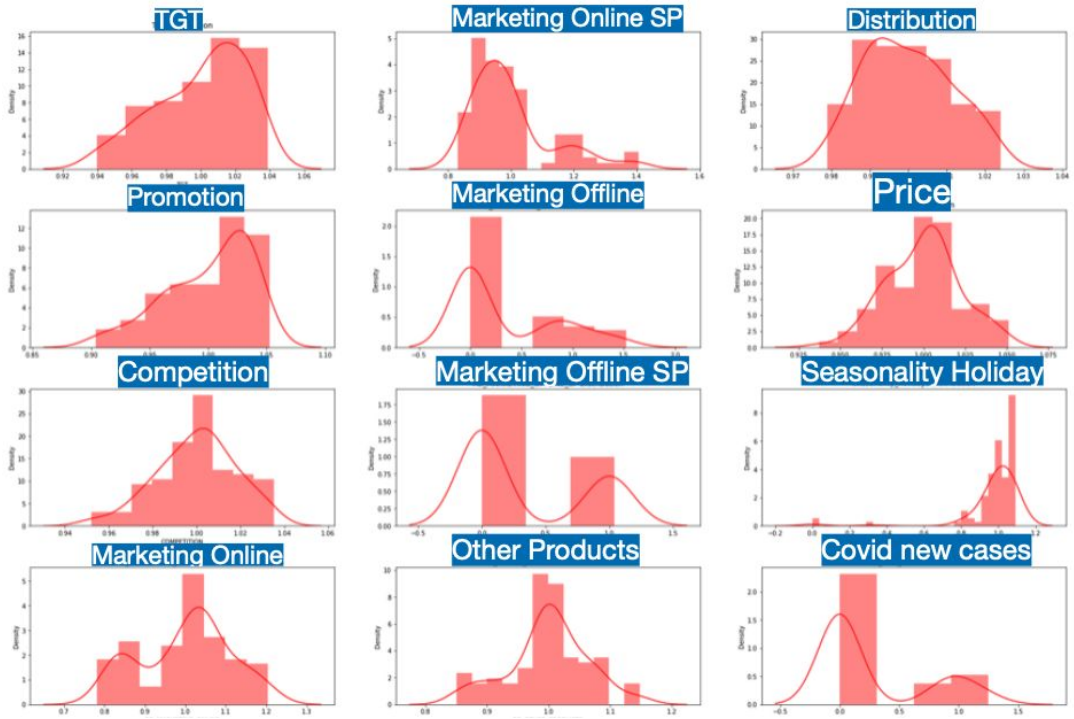
- The original model was only **weakly identified**, since the given priors were consistent with all sorts of parameter values that did not make scientific sense.
- So I decided to take the time to specify **more informative priors**, and thus the original model was played as a default version that could be improved.

Results:

	R-hat(1.00 is perfect)	WAIC(Lower is better)	LOO(Lower is better)	Rank
Original Model	Mostly between 1.00~1.10	-804.72	-772.19	2
Model 1		-771.60	-736.86	4
Model 2		-782.59	-748.42	3
Model 3		**-843.12**	**-808.14**	**1**

What did I change?(First Change)

Obviously, some variables are not normally distributed!



Distribution Modification

Variables	Before	After
Seasonality holidays	Normal	Left-Skewed Normal
Covid new cases	Normal	Bernoulli
Promotion	Normal	Left-Skewed Normal
Distribution	Normal	Right-Skewed Normal
Competition	Normal	Left-Skewed Normal
Sales	Normal	Left-Skewed Normal

What did I change?(Second Change)

After sampling from a model without seeing any data, I modified mu and sigma of some interaction coefficients according to the summary statistics.

	mean	sd
beta_seasonality_holidays_to_promo	-0.267	0.488
beta_covid_new_cases_to_promo	-0.104	0.415
beta_seasonality_holidays_to_distribution	0.284	0.467
beta_covid_new_cases_to_distribution	-0.268	0.492
beta_price_to_competition	-0.075	0.790
beta_seasonality_holidays_to_competition	0.305	0.856
beta_covid_new_cases_to_competition	-0.105	0.833
beta_offline_to_competition	-0.808	0.802
beta_price_to_otherprod	-0.427	0.723
beta_seasonality_holidays_to_otherprod	-0.003	0.672
beta_covid_new_cases_to_otherprod	-0.782	0.802
beta_promo_to_otherprod	-0.384	0.812
beta_online_to_otherprod	0.017	0.712
beta_offline_to_otherprod	0.210	0.823
beta_seasonality_holidays_to_offline	-0.163	0.684
beta_covid_new_cases_to_offline	0.015	0.972
beta_promo_to_offline	-0.185	0.913
beta_price_to_offline	-0.182	0.576
seasonality_holidays_alpha	-0.009	1.142
covid_new_cases_alpha	0.136	1.059

Results

Real data

count	104.000
mean	1.000
std	0.026
min	0.940
25%	0.981
50%	1.006
75%	1.021
max	1.039

Simulated data from original model

count	10400.000
mean	0.360
std	7.012
min	-30.083
25%	-3.495
50%	-0.247
75%	3.661
max	62.950

Simulated data from Modified model

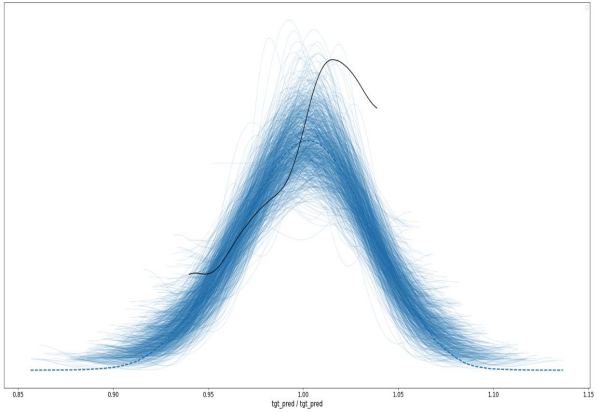
count	10400.000
mean	-0.087
std	3.832
min	-22.806
25%	-1.820
50%	0.296
75%	2.189
max	13.672

Gap

Decreased by 7.277

Decreased by 49.278

Posterior Predictive Check



Evaluation Metrics

R-hat	WAIC	LOO
1.00~1.10	-843.12	-808.14

Better Predictive Ability!

Insights

- Due to the limitation of the tasks assigned to me, it’s hard to provide many really useful information about the business, but some insights can be obtained from interaction.
- For interaction priors, if we pick those with | mean | > 0.3 and std < 0.6, then we will get:

Insights

	Sales	Offline	Other Products
Price			
Holidays			
Promotion			
Competition			

- Launch more promotion activities, especially on holidays.
- Don’t make the price too high.

Future

- Continue looking for better priors.
- Use continuous data to build DAGs.

Thank you!