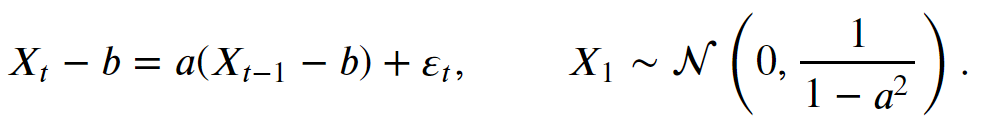
Prep Meeting 31

# AR(1) Model With off-set

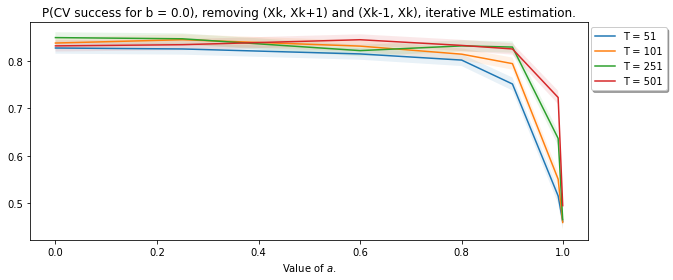
## Problem Setting

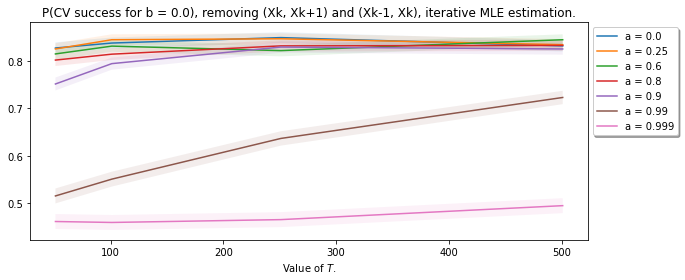
Consider the AR(1) model with offset.



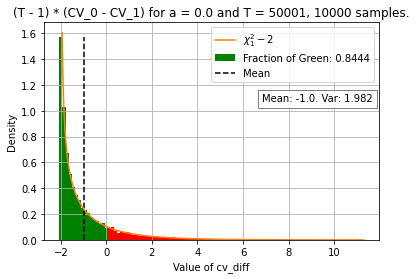
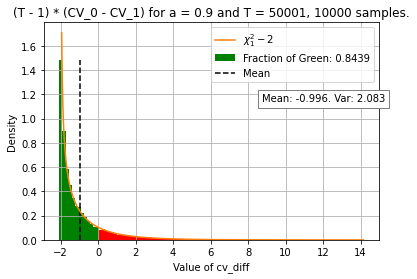
We want to use cross-validation to determine whether







# AR(1) Model success

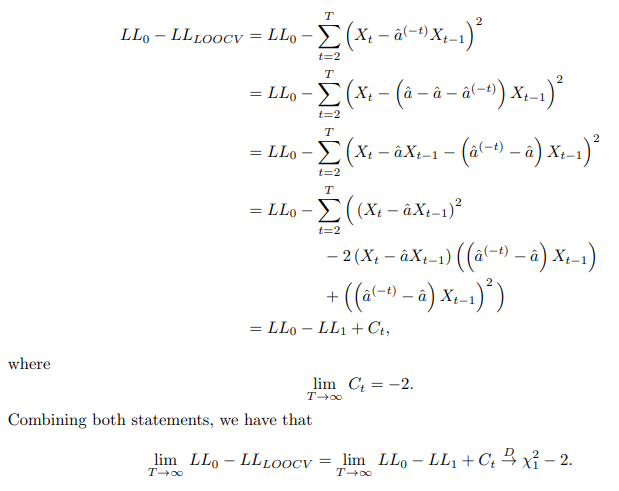
We made the problem statement very simple and concrete. Suppose we have an AR(1) model with coefficient *a\_0*. We use cross-validation to determine whether this is the case. Interestingly, the probability of success seems to range around *0.84*. We have discovered the origin of this value.

We see that the distribution follows a shifted chi squared distribution, where it is shifted two units to the left. This also holds true for any value of *a*. In turn, the probability of success is approximately P(chi – 2 < 0) = 0.843.



Now, exactly why is this true? And why is this? We know by Wilk’s Theorem that the difference between the using a\_0 and the MLE of a follows a chi squared distribution with one degree of freedom. Apparently, the loocv estimate shifts this value two units to the left.

We see that the difference can be rewritten, and we get that this difference follows a normal distribution with a mean of two and a variance that converges to zero.



# Writing about regularization approaches

1. Bootstrapping.
   1. 95% Confidence Intervals
   2. Recoverability of edges.
2. LASSO-Regularization.
   1. Given acyclic support B, use LASSO to shrink solutions. Which value for lambda?
      1. Minimizing BIC / AIC, lambda often too low.
      2. Cross-Validation / Bootstrapping?
3. Cross-Validation.
   1. LOOCV. *K*-fold CV seemed unnecessary.
4. Perhaps discuss “classical” approach, and see that that produces worse estimates.

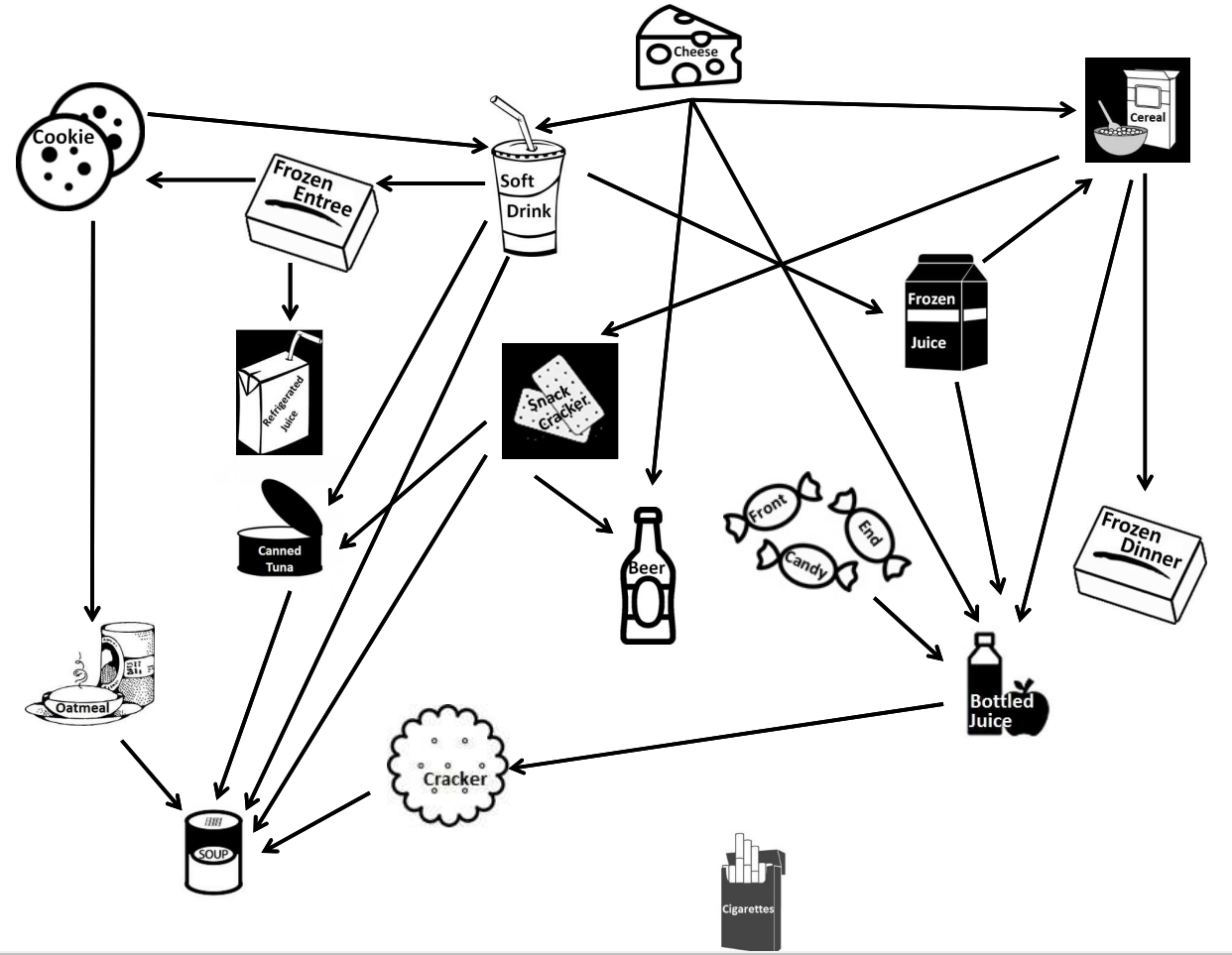
# Writing about continuous optimization, slight improvement

Finished writing up the continuous optimization part. Some slight improvements indeed give better results, and “convergence” is now also true for a larger number of dimensions. However, we still have the issue that there are optima where P is far from a permutation matrix.

Furthermore, I found that the method does not enforce DAG-ness, so this method does not seem to be very sensible. Still, an interesting try though.

# VAR(1) Real-Life data.

Interesting dataset containing lots of prices / sales / etc. Found a VAR(2) sparsity modeler that modeled the influence of **prices** of A on the **sales** of B, and I thought the figure looked quite interesting. In the process of remaking the dataset, and also in the have also requested the data.



Another interesting one could be stock exchange markets, or news channels reporting. Any method that would show directed relations.