Before testing any model we have to establish two crucial components of the viability measure. First, we require a prediction model which we want to explain using counterfactuals. This is relevant for determining the improvement that a counterfactual yileds in contrast to the factual. Second, we need to know to what extend any given counterfacual is feasbile given the dataset at hand. Therefore, we will dedicate the first set of experiments to establishing these components.

To compute the viability of a counterfactual we need to determine its feasibility. In other words, we have to determine the possibility or impossibility of the counterfactual. We can use the data log to gauge the feasibility, by estimating the data distribution.

There are many ways to estimate the density of a data set. For our purposes, we incorporate the sequential structure of the log data and make simplifying assumptions. First, we consider every activity as a state in the case. Second, each state is only dependent on its immediate predecessor and neither on future nor on any any states prior to its immediate predecessor. Third, the collection of attributes within an event depend on the activity which emits it. The second assumption is commonly known as *Markov Assumption*. With these assumptions in place, we can model the distribution by knowing the state transition probability and the density to emit a collection of event attributes given the activity. The probability distributions are shown in **??**.

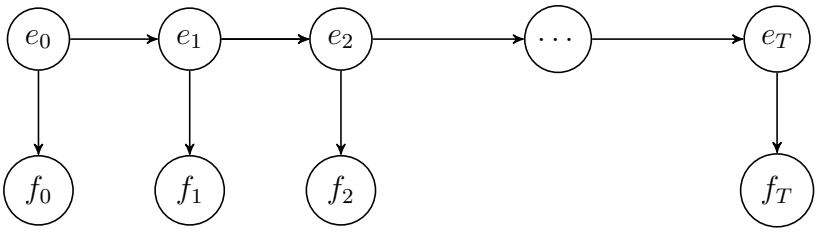

Figure 1: The feaibility model in graphical form.

Here, e_t represents the transition from one event state to another. Likewise, f represent the emission of the feature attributes. Hence, the probability of a particular sequence is the product of the transition probability multiplied with state emission probability for each step. Note, that this is the same as the feasibility measure as in **??**. **[Make formula in viability section consistent with this! Formula needs to use e instead of a and change starting index from 1 to 0.]**

$$p\left(e_{0:T}, f_{0:T}\right) = p\left(e_0\right) p\left(f_0 \mid e_0\right) \prod_1^T p\left(e_t \mid e_{t-1}\right) p\left(f_t \mid e_t\right) \tag{1}$$

Practical Matters

The general computation of these products is trivial. However, we need to probabilities for $p\left(e_0\right)$, $p\left(f_t \mid e_t\right)$ and $p\left(e_t \mid e_{t-1}\right)$ as the true distributions are unobservable.

Starting with the transition dynamics part of the equation $p\left(e_t \mid e_{t-1}\right)$, we can estimate the model parameters by counting the transitions from one event state **(acitvity)** to another **(acitvity)**.**[Define the difference between event und activity in a better way and earlier.]** $p\left(e_0\right)$ is a special case, as it does not have a preceding event.

The emission probabilities are more complicated for three reasons: First, the event distribution does not necessarily belong to the same family as the feature distribution. Hence, we cannot use any simple method to estimate these conditionals. We need to estimate the probability for each event seperately. The second issue directly follows from the first. If we estimate each event distribution by partioning the data by events, we naturally have less data to estimate each model's parameters. Although, event partitioning, are not an issue for common events states **(activities)**, they can make emission probabilities of less frequent event states exceptionally hard to estimate. One can turn to Bayesian Methods, which hand these situations better by specifying a prior.

However, the third issue exacerbates the main issue of using bayesian methods. Namely, because features do not necessarily have to be from the same distributional family, we have to model each conditional distribution with a mixture of distributions. Hence, simple bayesian updates are not possible either **and require more time expensive methods such as Markov-Chain-Monte-Carlo methods or similar**.**[Check if this is true. Maybe we *can* use MCSC]**.

From these issues, we can conclude there are multiple viable ways to model these conditional distributions and we have to choose an fitting method[1].

[1]Note, that we did not mention modelling $p\left(f_0 \mid e_0\right)$ as it is practically the same distribution as $p\left(f_t \mid e_t\right)$.