Blind background prediction using a bifurcated analysis scheme

J. Nix, J. Ma, G.N. Perdue, Y.W. Wah

Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637, USA
(Dated: February 1, 2008)

A technique for background prediction using data, but maintaining a closed signal box is described. The result is extended to two background sources. Conditions on the applicability under correlated cuts are described. This technique is applied to both a toy model and an analysis of data from a rare neutral kaon decay experiment.

PACS numbers: 06.20.Dk,14.40.Aq

In this paper we describe a bifurcation analysis procedure for data driven background prediction in a blind analysis of a closed signal region. In this type of analysis we define a signal region, in some set of variables, where we keep the events hidden during the analysis. We wish to predict the number of background events under the full set of cuts in this signal region. We use the term cut to refer to a specific selection criteria. An event that passes a cut is selected and an event which fails a cut is thrown out. The procedure uses the application of inverse cuts to properly measure the veto power of the different sets of cuts while keeping the signal region closed. This technique was first developed for a single background source in the stopped $K^+$ experiments E787 and E949 at Brookhaven [1]. The work in this paper was inspired by the use of the bifurcation technique in the E391a experiment [2].

We begin with a derivation of the bifurcation analysis in the case of one background source and uncorrelated cuts. We then extend this to two background sources and a simple model of correlation between cuts. Throughout this paper the method will be applied to a toy model to predict background and then applied to an example from the E391a experiment. We utilize the Mathematica software package to implement the toy model [3].

ONE BACKGROUND CASE

We begin discussing this method in the case of a single background source. Here a collection of setup cuts have been applied which eliminate all other sources of background. Our goal is to predict the amount of background in the signal region when we apply the cuts $A$ and $B$, which we refer to as the bifurcation cuts. The number of events we observe will be determined by the number of events before applying the cuts $A$ and $B$ (after applying the setup cuts) and the cut survival probability (CSP):

$$N_{\text{bkg}} = N_0P(AB).$$

We consider events to lie in a multi-dimensional space with a dimension corresponding to every variable on which we can cut. Our set of cuts defines a multidimensional signal region which we wish to keep blind. If two cuts show no correlation in the events that they cut, this implies that these two cuts are orthogonal in this space.

A diagram of this situation is shown in Fig. 1. The CSP can then be decomposed into $P(AB) = P(A)P(B)$:

$$N_{\text{bkg}} = N_0P(A)P(B).$$

This can be expanded into

$$N_{\text{bkg}} = \frac{N_0^2P(A)P(B)P(\bar{A})P(\bar{B})}{N_0P(A)P(B)}.$$

Here $\bar{A}$ and $\bar{B}$ are the inverses of cuts $A$ and $B$, events which pass cut $A$ fail cut $\bar{A}$. Then we can calculate this from data based on the number of observed events in the signal region under the different cut conditions:

$$N_{\bar{A}B} = N_0P(A)P(\bar{B}),$$

$$N_{AB} = N_0P(B)P(\bar{A}),$$

$$N_{\bar{A}\bar{B}} = N_0P(\bar{A})P(\bar{B}).$$

Here $N_{AB}$ is the number of background events observed with the application of cut $A$ and the inverse of cut $B$. $N_{\bar{A}B}$ is the observed background events with the inverse of cut $A$ and cut $B$ applied. $N_{\bar{A}\bar{B}}$ is the count when the inverse of both $A$ and $B$ are applied. All of these values are outside the signal region defined in the multidimensional cut space allowing us to predict the background without opening the box:

$$N_{\text{bkg}} = \frac{N_{AB}N_{\bar{A}B}}{N_{\bar{A}\bar{B}}}.$$
events which pass these sets of cuts, $N_{AB}$. We then do the same procedure in reverse applying the set of cuts $B$ and the inverse of $A$ to find $N_{AB}$. Finally, we apply the inverse of both cuts $A$ and $B$ to find $N_{AB}$. These values are combined to produce the background prediction.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|}
\hline
$N_{bkg}$ & $N_{AB}$ \\
\hline
$N_{AB}$ & $N_{AB}$ \\
\hline
\end{tabular}
\caption{Schematic of background distribution in the cut space.}
\end{table}

Eqn (7) predicts no background events, if $N_{AB}$ or $N_{AB}$ are zero. This can be true for one of three reasons: $N_0 = 0$, $P(A)$ or $P(B)=0$, or $P(A)$ or $P(B) = 0$. The first two possibilities are what we expect and reflect cases where there should be no background events. The third possibility is more problematic. When $P(\bar{A})$ or $P(\bar{B}) = 0$, $N_{AB}$ should also be zero, but statistical fluctuations may prevent that from being true. This condition results from a poor choice of cuts where one cut eliminates almost no background events. If possible a different choice of cuts for $A$ and $B$ should be made.

**The Toy Model**

Each event is described by four variables. Two kinematic variables, $p$ and $x$, which are used to describe the signal region and two cut variables, and $a$ and $b$, which will be used to define the cuts. The cut variables, $a$ and $b$, are independent of the kinematic variables, $p$ and $x$. All of these variables range from 0 to 1.

We define 2 different types of events: Background 1 and Background 2. They both have $x$ variables with uniform distributions between 0 and 1. Their $p$’s have uniform distributions between 0 and $x$ for Background 1 and between 0 and $1-x$ for Background 2. Background 1 has an uniform distribution of $a$ between 0 and 1 and variable $b$ has a linearly decreasing density with values between 0 and 1. Background 2’s $a$ distribution is a linearly decreasing density with values between 0 and 1 and an uniform distribution between 0 and 1 for $b$. The distribution of each variable for the two background types is shown in Table I. The distributions of the kinematic variables $p$ and $x$ are shown in Fig. [2]. We define a signal region by specifying the allowed kinematic variables: $0.25 < x < 0.75$ and $0.25 < p < 0.75$.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|}
\hline
Background & $x$ & $p$ \\
\hline
1 & $f(x) = 1$, $x \in [0, 1]$ & $f(p) = 1/x$, $p \in [0, x]$ \\
& $f(x) = 0$, $x \notin [0, 1]$ & $f(p) = 0$, $p \notin [0, x]$ \\
\hline
2 & $f(x) = 1$, $x \in [0, 1]$ & $f(p) = 1/(1-x)$, $p \in [0, 1-x]$ \\
& $f(x) = 0$, $x \notin [0, 1]$ & $f(p) = 0$, $p \notin [0, 1-x]$ \\
\hline
Background & $a$ & $b$ \\
\hline
1 & $f(a) = 1$, $a \in [0, 1]$ & $f(b) = 1 - b$, $b \in [0, 1]$ \\
& $f(a) = 0$, $a \notin [0, 1]$ & $f(b) = 0$, $b \notin [0, 1]$ \\
\hline
2 & $f(a) = 1 - a$, $a \in [0, 1]$ & $f(b) = 1$, $b \in [0, 1]$ \\
& $f(a) = 0$, $a \notin [0, 1]$ & $f(b) = 0$, $b \notin [0, 1]$ \\
\hline
\end{tabular}
\caption{Probability distribution functions for the variables of each event type in the Toy Model.}
\end{table}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig1.png}
\caption{Kinematic variable distributions for Backgrounds 1 and 2.}
\end{figure}

We define our cuts on variables $a$ and $b$ as:

$$A = (a > 0.5),$$
$$B = (b > 0.5).$$

A and $B$ are true or false statements. If they are false the event is cut. With the cut points defined, we can then calculate the CSP for each event type: $P(A)$ or $P(B)$. In this toy model the CSPs can be calculated analytically because we know the underlying distributions. These values are shown in Table II.

**One Background in the Toy Model**

In this section we discuss the case of a single significant background. The background prediction is given by Eqn (7). We generated $1 \times 10^4$ Background 1 events over the whole range of kinematic variables. This leaves
We expand the numerator:
\[
N_{AB}N_{\bar{A}B} = N_{1}^{2}P_{1}(A)P_{1}(\bar{A})P_{1}(B)P_{1}(\bar{B}) \\
+ N_{1}N_{2}[P_{1}(A)P_{2}(\bar{A})P_{2}(B)P_{1}(\bar{B})] \\
+ P_{2}(A)P_{1}(\bar{A})P_{1}(B)P_{2}(\bar{B}) \\
+ N_{2}^{2}P_{2}(A)P_{2}(\bar{A})P_{2}(B)P_{2}(\bar{B}).
\]  

(15)

We multiply \(N_{bkg}\) by \(N_{\bar{A}B}\) to allow us to find the difference of Eqn (14) and \(N_{bkg}\):
\[
N_{bkg}N_{\bar{A}B} = [N_{1}P_{1}(A)P_{1}(B) + N_{2}P_{2}(A)P_{2}(B)] \\
\times [N_{1}P_{1}(\bar{A})P_{1}(\bar{B}) + N_{2}P_{2}(\bar{A})P_{2}(\bar{B})] \\
= N_{1}^{2}P_{1}(A)P_{1}(\bar{A})P_{1}(B)P_{1}(\bar{B}) \\
+ N_{1}N_{2}[P_{1}(A)P_{2}(\bar{A})P_{1}(B)P_{1}(\bar{B})] \\
+ P_{2}(A)P_{1}(\bar{A})P_{2}(B)P_{1}(\bar{B}) \\
+ N_{2}^{2}P_{2}(A)P_{2}(\bar{A})P_{2}(B)P_{2}(\bar{B}).
\]  

(16)

The cross term vanishes if \(P_{1}(A) = P_{2}(A)\) and \(P_{1}(B) = P_{2}(B)\), where for the purposes of the cuts the two backgrounds are the same.

We can simplify the cross term by rewriting the CSPs of the inverse cuts in terms of the CSPs of the cuts, \(P_{1}(\bar{A}) = 1 - P_{1}(A)\). Each element of the cross term has the same structure which can be expanded to:
\[
P_{1}(A)P_{1}(\bar{A})P_{k}(B)P_{l}(\bar{B}) = P_{1}(A)P_{k}(B)(1 - P_{l}(A)) \times (1 - P_{l}(B)) \\
= P_{1}(A)P_{k}(B) \\
- P_{1}(A)P_{k}(A)P_{k}(B) \\
- P_{1}(A)P_{k}(B)P_{k}(B) \\
+ P_{1}(A)P_{k}(A)P_{k}(B)P_{k}(B),
\]  

(17)

(18)

Summing the elements of the cross term cancels everything except the terms with two CSPs:
\[
\frac{N_{AB}N_{\bar{A}B}}{N_{AB}} = N_{bkg} \\
\frac{1}{N_{AB}}[(N_{1}N_{2}(P_{1}(A)P_{2}(B) + P_{2}(A)P_{1}(B) \\
- P_{1}(A)P_{1}(B) - P_{2}(A)P_{2}(B))] \\
= N_{bkg} \frac{N_{1}N_{2}(P_{2}(A) - P_{1}(A))(P_{2}(B) - P_{1}(B))}{N_{AB}}.
\]  

(19)
We can further simplify the cross term by defining \( \Delta_A = P_2(A) - P_1(A) \) and \( \Delta_B = P_2(B) - P_1(B) \).

\[
N_{\text{bkg}} = \frac{N_{AB}N_{\bar{AB}}}{N_{\bar{AB}}} + \frac{N_1N_2}{N_{AB}}\Delta_A\Delta_B \quad (20)
\]

The second term in Eqn (20) is not the contribution of a particular source to the background prediction. It is a correction to the prediction of the total number of background events from both sources.

**Properties of the Two Background Solution**

We now discuss the behavior of Eqn (20). This solution has the reasonable property that it is symmetric with respect to the definitions of the cuts \( A \) and \( B \) and the backgrounds 1 and 2.

We consider the behavior of Eqn (20) under extreme conditions. First, we show that under no conditions can the total \( N_{\text{bkg}} \) be negative. The correction term will have its maximum negative value when \( \Delta_A = 1 \) and \( \Delta_B = -1 \) or \( \Delta_A = -1 \) and \( \Delta_B = 1 \). Under these conditions, \( N_{\bar{AB}} = 0 \) and \( N_{\text{bkg}} \) is undefined. We therefore want to study \( N_{\text{bkg}} \)'s behavior as we approach this limit. We begin by setting \( \Delta_B = -1 \) and studying the limit as \( \Delta_A \to 1 \). In this case cut \( B \) removes Background 1 completely, but the Background 1 events which survive \( \bar{A}B \) still contribute to the prediction produced by Eqn (7).

The condition that \( \Delta_B = -1 \) sets what values the CSPs of the B can take:

\[
P_1(B) = P_2(\bar{B}) = 0,
\]

\[
P_2(B) = P_1(\bar{B}) = 1.
\]

Substituting these values into Eqns (11) (13) we find:

\[
N_{\bar{AB}} = N_2P_2(A),
\]

\[
N_{\bar{A}B} = N_1P_1(\bar{A}) = N_1(1 - P_1(A)),
\]

\[
N_{\bar{AB}} = N_2P_2(\bar{A}) = N_2(1 - P_2(A)).
\]

We then substitute these values into Eqn (20) and sum the two terms:

\[
N_{\text{bkg}} = \frac{N_1(P_2(A) - P_2(A)P_1(A) - P_2(A) + P_1(A))}{1 - P_2(A)}
= \frac{N_1P_1(1 - P_2(A))}{1 - P_2(A)} = N_1P_1(A).
\]

As \( \Delta_A \to 1 \), \( P_1(A) \to 0 \). Therefore the \( N_{\text{bkg}} \) goes to 0. This indicates that \( N_{\text{bkg}} \) never has a non-physical negative value.

We now consider the case of \( \Delta_B = 0 \). Here the correction term is zero, but the contribution from the second background is not. We can see that the first term in Eqn (20) correctly predicts the background by simplifying Eqns (10) (13) substituting \( P(B) \) for \( P_1(B) \) and \( P_2(B) \):

\[
N_{\text{bkg}} = (N_1P_1(A) + N_2P_2(A))P(B),
\]

\[
N_{\bar{AB}} = (N_1P_1(A) + N_2P_2(A))P(\bar{B}),
\]

\[
N_{\bar{A}B} = (N_1P_1(\bar{A}) + N_2P_2(\bar{A}))P(B),
\]

\[
N_{\bar{AB}} = (N_1P_1(\bar{A}) + N_2P_2(\bar{A}))P(\bar{B}).
\]

Substituting these values into the first term of (20) and simplifying gives:

\[
\frac{N_{AB}N_{\bar{AB}}}{N_{\bar{AB}}} = (N_1P_1(A) + N_2P_2(A))P(B).
\]

This is \( N_{\text{bkg}} \), so the correction is not necessary to properly predict the background when one of the \( \Delta \)'s is zero.

**Interpretation of the Two Background Solution**

It may seem counterintuitive that two backgrounds cannot be combined simply. This can best be understood as the second background introducing an implicit correlation.

As an example, consider two backgrounds which individually have no correlation between cut \( A \) and cut \( B \), but do have different cut survival probabilities. If \( P_1(\bar{A}) = 0.75 \) and \( P_1(B) = 0.75 \), while \( P_2(A) = 0.25 \) and \( P_2(B) = 0.25 \) then the resulting combination of the two backgrounds will have a correlation. Events which survive cut \( A \) are have a greater chance to survive cut \( B \), because events which survive cut \( A \) are more likely to be part of Background 1. Events that don’t survive cut \( A \) are less likely to survive cut \( B \), because they are more likely to be part of background 2. Therefore there is a correlation between cuts \( A \) and \( B \), even though for the individual backgrounds they are uncorrelated.

The values of \( N_1 \), \( N_2 \), \( \Delta_A \), and \( \Delta_B \) are not directly accessible in data without opening the signal box. There are two options: either derive these values from Monte Carlo or from other regions in signal space. Determining \( N_1 \) and \( N_2 \) generally will require both an alternative way of predicting one of the backgrounds and the value of \( N_0 \), the total number of background after setup cuts. This raises the question whether determining \( N_0 \) biases the analysis. From \( N_0 \) and the other observed background numbers, \( N_{AB}, N_{\bar{AB}}, \) and \( N_{\bar{AB}} \) it is possible to effectively open the box and count \( N_{\text{bkg}} \). Determining \( \Delta_A \) and \( \Delta_B \) also requires additional input. Their values can be derived from either Monte Carlo or data outside the signal region.
Two Background Toy Model

In our toy model, we can study the effect of multiple background sources by varying the relative strength of a second background. We begin by calculating with the false assumption that there is a single background mode. We vary the relative admixture of Background 2. The total number of events, \( N_1 + N_2 \), was held constant at \( 2 \times 10^4 \). The discrepancy between the prediction and the observed background increases as the number of background events from the second source increases, as shown in Fig. 3.

We now apply the correction term to the background prediction (Eqn 20). In the case of this toy model, we know the values of \( N_1 \) and \( N_2 \) because we have set them. In a real analysis, it would be necessary to determine these values through either Monte Carlo studies or studies of different signal regions which are then extrapolated into the signal box. The differences in the cut probabilities, \( \Delta_A \) and \( \Delta_B \), also need to be determined from outside sources. In this model \( \Delta_A = -25\% \) and \( \Delta_B = 25\% \). Since the probability differences are of opposite signs the correction is negative and reduces the predicted background.

In Fig. 3, we show the results of keeping the total number of background events the same while increasing the fraction of Background 2 events. Here only the predicted background without correction increases, while the observed and corrected backgrounds remain relatively flat.

Impact of Cut Correlation

To derive a correction to the background prediction, we use a simple model of cut correlation. We describe a case where the cuts have a weak linear correlation. In this model, the posterior probability for each cut is different than the prior probability.

![Cut space with correlated cuts.](image)

FIG. 4: Cut space with correlated cuts.

We begin by specifying the background values in terms of the CSPs, which are now posterior probabilities, specifying the dependance on both cut conditions:

\[
N_{\text{bkg}} = N_0 P(A|B)P(B) = N_0 P(A)P(B|A), \tag{32}
\]

\[
N_{\bar{A}B} = N_0 P(\bar{A}|\bar{B})P(\bar{B}) = N_0 P(\bar{A})P(\bar{B}|\bar{A}), \tag{33}
\]

\[
N_{A\bar{B}} = N_0 P(A|\bar{B})P(\bar{B}) = N_0 P(A)P(\bar{B}|\bar{A}), \tag{34}
\]

\[
N_{\bar{A}\bar{B}} = N_0 P(\bar{A}|\bar{B})P(\bar{B}) = N_0 P(\bar{A})P(\bar{B}|\bar{A}). \tag{35}
\]

We proceed in the same fashion as for the two background case and substitute these definitions into the solution (Eqn 7) for the single background, uncorrelated case:

\[
\frac{N_{A\bar{B}}N_{\bar{A}B}}{N_{A\bar{B}}} = \frac{N_0^2 P(A|B)P(\bar{B})P(\bar{A}|B)P(B)}{N_0 P(A|B)P(B)}. \tag{36}
\]

We assume the correlations are small and relate the

CUT CORRELATION

In the derivation of both the one and two background cases, we have assumed that the cuts \( A \) and \( B \) are un-correlated. Of course, in real applications, it is unlikely to find two cuts which are perfectly uncorrelated. We would therefore like to find some general figure of merit to determine how correlation introduces errors into the background prediction. The following discussion will assume only one background source.
different posterior probabilities to each other:

\[
P(A|B) = P(A|B) - \epsilon, \quad (37)
\]

\[
P(\bar{A}|\bar{B}) = P(\bar{A}|B) + \epsilon, \quad (38)
\]

\[
P(B|\bar{A}) = P(B|A) - \delta, \quad (39)
\]

\[
P(\bar{B}|\bar{A}) = P(\bar{B}|A) + \delta. \quad (40)
\]

The corrections \(\epsilon\) and \(\delta\) should be small. What we mean by small will be defined at the end of the derivation by what values are necessary for the corrections which are first order in \(\epsilon\) and \(\delta\) to be negligible. We have the freedom to choose to formulate the correction as a function of \(\epsilon\) or \(\delta\). We choose to use \(\epsilon\). We substitute the definitions into Eqn 31:

\[
\frac{N_{AB}N_{\bar{A}B}}{N_{AB}} = \frac{N_0 P(\bar{A}|B)(P(A|B) - \epsilon P(B))}{P(A|B) + \epsilon}
\]

\[
= \frac{N_0(\epsilon P(A|B) - \epsilon P(B))}{1 + \frac{P(A|B)}{P(\bar{A}|B)}}. \quad (41)
\]

Assuming the \(\epsilon\) term in the denominator is small, we expand this result in \(\epsilon\):

\[
\frac{N_{AB}N_{\bar{A}B}}{N_{AB}} \approx \left(N_0(\epsilon P(A|B) - \epsilon P(B))\right)
\]

\[
\times \left(1 - \frac{\epsilon}{P(A|B)} + \frac{\epsilon^2}{P(A|B)^2} + \mathcal{O}(\epsilon^3)\right). \quad (42)
\]

Multiplying this out and keeping the second order terms of \(\epsilon\) gives:

\[
\frac{N_{AB}N_{\bar{A}B}}{N_{AB}} = N_0 P(A|B)P(B) - \epsilon N_0 (P(B) + \frac{P(A|B)P(B)}{P(A|B) + \epsilon})
\]

\[
+ \epsilon^2 N_0 \left(\frac{P(B)}{P(A|B)} + \frac{P(A|B)P(B)}{P(A|B)^2}\right). \quad (43)
\]

The first term with no \(\epsilon\) factors is \(N_{bkg}\). The condition for the correlations to have a negligible impact on our background prediction is that the \(\epsilon\) terms be much smaller than \(\frac{N_{AB}N_{\bar{A}B}}{N_{AB}}\):

\[
N_{\text{bkg}} = \frac{N_{AB}N_{\bar{A}B}}{N_{AB}} + \epsilon N_0 P(B)(1 + \frac{P(A|B)}{P(A|B) + \epsilon})
\]

\[
- \epsilon^2 N_0 \left(\frac{P(B)}{P(A|B)} + \frac{P(A|B)P(B)}{P(A|B)^2}\right). \quad (45)
\]

These terms require opening the signal box to know the correct values of the CSPs. We can however approximate these values with less knowledge, under the assumption that the number of events in the signal box is small:

\[
P(B) = \frac{N_{AB} + N_{\bar{A}B}}{N_0} \approx \frac{N_{AB}}{N_0}. \quad (46)
\]

\[
P(A|B) = \frac{N_{AB}}{N_0} \approx \frac{N_{\text{pred}}}{N_{AB}}. \quad (47)
\]

These approximations give us the first order correction:

\[
C_\epsilon = \epsilon N_{AB}(1 + \frac{N_{\text{pred}}}{N_{\bar{A}B}}). \quad (48)
\]

Returning to our toy model, we introduce a correlation between the \(a\) and \(b\) variables in Background 1. We add a term linearly dependent on \(b\) to \(a\), and then scale \(a\) to keep it between 0 and 1 and to reduce the change in background due to the change in the average value of \(a\):

\[
f(a) = (1 + \epsilon'), a \in [\epsilon'/1, (1 + \epsilon')(1 + \epsilon')/(1 + \epsilon)]. \quad (49)
\]

The variable \(\epsilon'\) is the knob we use to tune the correlation. It is closely related to the variable \(\epsilon\) that is defined in Eqs (38) and (39) as is shown in Fig. 5.

We show the predicted and observed background in Fig. 5. As \(\epsilon'\) increases the background in data increases, because the correlation increases the average value \(a\), while the predicted background decreases.

In Fig. 6 we show the prediction with \(C_\epsilon\) added. It improves the agreement for a fairly wide range of \(\epsilon\).

**EXAMPLE FROM E391A**

For a more realistic example, we discuss the use of this technique in the first search for the rare kaon decay \(K_L \rightarrow \pi^0\pi^0\nu\bar{\nu}\) at E391a. E391a is dedicated experiment for the search for the rare kaon decay \(K_L \rightarrow \pi^0\nu\bar{\nu}\) located at KEK. The experimental apparatus consists of a 4\(\pi\) hermetic photon veto system and an array of CsI, an inorganic crystal scintillator, for signal detection. The \(K_L \rightarrow \pi^0\pi^0\nu\bar{\nu}\) decay has a Standard Model branching ratio of \((1.4 \pm 0.4) \times 10^{-13}\). The bifurcation method is to predict the background to this mode from data.

The final state of \(K_L \rightarrow \pi^0\pi^0\nu\bar{\nu}\) is four photons from the \(\pi^0\)'s and two unobservable \(\nu\)'s. Events are reconstructed by pairing the photons and calculating the \(\pi^0\)
decay vertices. The pairing with the smallest separation between $\pi^0$ vertices is selected. The signal region is defined in three kinematical variables: the transverse momentum of the $\pi^0-\pi^0$ system, the invariant mass of the $\pi^0-\pi^0$ system, $M_{\pi^0-\pi^0}$, and reconstructed decay vertex, $Z$.

The signal region is selected to avoid the two primary background sources: $K_L \rightarrow \pi^0\pi^0\pi^0\nu\bar{\nu}$ decays and neutron interactions with a membrane located downstream of the fiducial region. The $K_L \rightarrow \pi^0\pi^0\pi^0$ events have a mass peak below the $K_L$ mass and with relatively low $P_T$'s of under 100MeV/c. Their decay vertices cover the full range of the fiducial decay region. The neutron related events are well separated in the decay vertex having a peak at the membrane position, 50 cm downstream of the fiducial region. The $K_L \rightarrow \pi^0\pi^0\nu\bar{\nu}$ signal box is defined as $100\text{MeV/c} < P_T < 200\text{MeV/c}$, $268\text{MeV/c}^2 < M_{\pi^0-\pi^0} < 400\text{MeV/c}^2$, $300\text{cm} < Z < 500\text{cm}$. The distribution of events in the $P_T$-mass plane is shown in Fig. 8.

We divided our cuts into three groupings: the setup cuts, cut $A$ which contained most of the photon vertices, and cut $B$ which contained primarily cuts on photon reconstruction quality in the CsI array. Using Eqn 45 we first tested the method on regions surrounding the signal box. We define a Low $P_T$ region with the same bounds on $M_{\pi^0-\pi^0}$ and $Z$ as the signal region and $P_T < 100\text{MeV/c}$. The High Mass Region is defined with the same bounds on $P_T$ and $Z$ as the signal region and $400\text{MeV/c}^2 < M_{\pi^0-\pi^0} < 500\text{MeV/c}^2$. The results of this are shown in Table IV. There is a significant discrepancy in the Low $P_T$ region, this is due to neglecting the correlation.

![FIG. 6: Predicted and observed background for different levels of correlation between $a$ and $b$ in the Toy Model. Triangles are data background, squares are predicted, and diamonds are predicted plus second order $\epsilon$ correction (Eqn 45). The x-axis is $\epsilon$.](image)

![FIG. 7: Predicted and observed background for different levels of correlation between $a$ and $b$ in the Toy Model. Triangles are data background, squares are predicted, and diamonds are predicted plus practical $\epsilon$ correction (Eqn 15). The x-axis is $\epsilon$.](image)

![FIG. 8: $P_T$ plotted against mass with all cuts in the E391a experiment. $Z$ between 300 cm and 500 cm. The rectangular regions correspond to the regions in Table IV [2].](image)

### TABLE IV: Prediction of background events in different regions without correlation or secondary background corrections.

| Region      | $N_{A\beta}$ | $N_{A\beta}$ | $N_{A\beta}$ | Prediction | Data |
|-------------|--------------|--------------|--------------|------------|------|
| Low $P_T$   | 393          | 72           | 115          | 21.1 ± 3.3 | 13   |
| High Mass   | 46           | 9            | 4            | 0.78 ± 0.48 | 1    |
| Signal      | 84           | 18           | 2            | 0.43 ± 0.32 | 1    |

We estimated the cut correlation on the $K_L \rightarrow \pi^0\pi^0\nu\bar{\nu}$ background in its peak region using both data and Monte Carlo. The value of $\epsilon$ was found to be $(-0.049 \pm 0.035)$. We first apply Eqn 15 on the Low $P_T$ region and find...
a correction of $-6.67 \pm 4.81$. This correction brings the prediction in Table [V] into much better agreement with data. We next apply Eqn [18] to the signal region and find a correction of $(-0.12 \pm 0.12)$ events.

Low energy neutron interactions are difficult to simulate in general and we had the additional difficulty of not knowing the precise shape of the membrane in the beamline. Therefore, we used data to estimate the level of neutron background. The value of $N_2$ was estimated by fitting the distribution events in the High $P_T$ -High $Z$ region, $Z > 500$cm and $P_T > 100$MeV/c. Events in this region are predominately core neutron events. We applied a set of loose cuts to ensure high statistics. We subtracted off the contribution of the $K_L \rightarrow \pi^0\pi^0\pi^0$ background using a Monte Carlo prediction of their contribution under the loose set of cuts. The fitted core neutron distribution was extended into the signal region and the total number core neutron events in the signal region was predicted from the distribution. Then the predicted number of events in the signal region were scaled down using the factor by which the population of events in the High $P_T$ -High $Z$ region were reduced. Using this procedure we estimate $N_0 = 2.16 \pm 0.03_{\text{stat}} \pm 0.15_{\text{syst}}$. The large systematic error comes from the subtraction of the $K_L \rightarrow \pi^0\pi^0\pi^0$ contribution in the High $P_T$ -High $Z$ region. The differences between the cut survival probabilities for core neutron events and $K_L \rightarrow \pi^0\pi^0\pi^0$ events were found by comparing the difference in cut effectiveness between the $K_L \rightarrow \pi^0\pi^0\pi^0$ peak in the Low Mass-Low $P_T$ region and the core neutron peak in the High $Z$ region. The values which go into the correction and the computed correction are shown in Table [V].

| $\Delta A$ | $\Delta B$ | $N_1$ | $N_2$ | Correction |
|------------|------------|-------|-------|------------|
| 26.4% | 8.6% | (101.9$\pm$10.2 $\pm 1.05$) | (2.1$\pm$0.03 $\pm 1.05$) | (0.06 $\pm 0.05$) |

TABLE V: Values for the core neutron multiple background correction.

The $K_L \rightarrow \pi^0\pi^0\pi^0$ decay has three ways in which it can cause backgrounds depending on the number of photons lost through veto inefficiency or fusion of photon clusters in the CsI. These three channels are not separated in the signal space, so our estimate of cut correlation includes contribution from these multiple background types.

The statistical uncertainties on both the correlation and the core neutron correction are large relative to their size. Therefore we decided to use these values as estimates of the systematic error made by neglecting the correlation and multiple background corrections. This results in a background prediction of $0.43 \pm 0.32_{\text{stat}} \pm 0.13_{\text{syst}}$. Opening the signal box, we observed a single event, consistent with our background prediction.

**DISCUSSION**

The bifurcation analysis technique allows us to produce data driven background predictions while still maintaining a blind analysis. In this paper we have shown how to extend the bifurcation analysis to the case of two background sources and correlated cuts.

The correction for a second background source is effective for any level of secondary background. It does require information beyond which is available directly from a blind analysis. It requires a combination of Monte Carlo information about the relative strengths of the two backgrounds and how the cut survival probabilities vary between the two backgrounds. It also requires knowledge of $N_0$, the number of events in the signal box after the setup cuts.

Correlations between the two cuts cannot be handled as easily. Even when the correlation between cuts is linear, the effects on the background prediction are nonlinear. Therefore, special care must be taken when selecting the cuts for the bifurcation analysis to avoid correlation.

One particular aspect of the impact of the prediction on the cut correlation is its dependence on the value of $N_0$. This leads to two competing forces in optimizing the division of cuts into the setup cuts and the bifurcation cuts. From the perspective of minimizing statistical error in the bifurcation prediction, one would like a large value for $N_0$ with powerful cuts for cut $A$ and $B$, so that the statistical errors on the terms of Eqn [4] are small. On the other hand, a large $N_0$ means the prediction is sensitive to small correlations between the cuts.

[1] S. Adler et al. “Evidence for the Decay $K^+ \rightarrow \pi^+\nu\bar{\nu}$,” *Phys. Rev. Lett.* 79, 2204 (1997)
[2] J. Nix et al. “First Search for $K_L \rightarrow \pi^0\pi^0\nu\bar{\nu}$,” *Phys. Rev. D* 76, 011101(R) (2007)
[3] Wolfram Research, Inc., Mathematica, Version 5.2, Champaign, IL (2005).
[4] C.W. Chiang and F. J. Gilman. “$K_{L, S} \rightarrow \pi\pi\nu\bar{\nu}$ decays within and beyond the standard model,” *Phys. Rev. D* 62, 094026 (2000).