



Blind background prediction using a bifurcated analysis scheme

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

The University of Chicago, Enrico Fermi Institute, 5640 South Ellis Avenue, Chicago, IL 60637, USA

ARTICLE INFO

Article history:

Received 14 January 2010

Accepted 27 January 2010

Available online 2 February 2010

Keywords:

Data analysis scheme

Background prediction

Bifurcation

ABSTRACT

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.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

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].

2. 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):

$$N_{\text{bkg}} = N_0 P(AB). \quad (1)$$

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 $P(AB) = P(A)P(B)$:

$$N_{\text{bkg}} = N_0 P(A)P(B). \quad (2)$$

This can be expanded into

$$N_{\text{bkg}} = \frac{N_0^2 P(A)P(B)P(\bar{A})P(\bar{B})}{N_0 P(\bar{A})P(\bar{B})}. \quad (3)$$

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:

$$N_{A\bar{B}} = N_0 P(A)P(\bar{B})$$

$$N_{\bar{A}B} = N_0 P(\bar{A})P(B)$$

$$N_{\bar{A}\bar{B}} = N_0 P(\bar{A})P(\bar{B}). \quad (4)$$

Here $N_{A\bar{B}}$ 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

* Corresponding author.

E-mail address: ywah@uchicago.edu (Y.W. Wah).

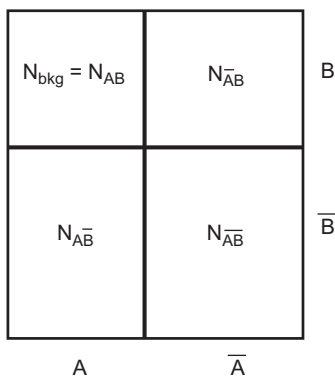


Fig. 1. Schematic of background distribution in the cut space.

applied. All of these values are outside the signal region defined in the multi-dimensional cut space allowing us to predict the background without opening the box:

$$N_{\text{bkg}} = \frac{N_{\overline{AB}} N_{\overline{A\overline{B}}}}{N_{\overline{A\overline{B}}}}. \quad (5)$$

The procedure for producing a background prediction is as follows. First, apply setup cuts to the data, the number of events in the signal box is N_0 . The setup cuts are all the cuts other than A or B , which are applied. They should remove all other background sources. There is more freedom in choosing the setup cuts than cuts A and B , since they can be correlated with each other or A and B . We then apply cut A and \overline{B} . By applying \overline{B} , we are looking at events which are outside the signal box in the multidimensional space. We count the number of events which pass these sets of cuts, $N_{\overline{AB}}$. We then do the same procedure in reverse applying the set of cuts B and the inverse of A to find $N_{\overline{A\overline{B}}}$. Finally, we apply the inverse of both cuts A and B to find $N_{\overline{A\overline{B}}}$. These values are combined to produce the background prediction. Eq. (5) predicts no background events, if $N_{\overline{AB}}$ or $N_{\overline{A\overline{B}}}$ are zero. This can be true for one of three reasons: $N_0=0$, $P(A)$ or $P(B)=0$, or $P(\overline{A})$ or $P(\overline{B})=0$. 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 $P(A)$ or $P(B)=0$, $N_{\overline{A\overline{B}}}$ 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.

3. 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 two different types of events: Backgrounds 1 and 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 a 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 1.

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 $0.25 < p < 0.75$. We

Table 1
Probability distribution functions for the variables of each event type in the Toy model.

Background	x	p
1	$f(x) = 1, x \in (0, 1]$	$f(p) = 1/x, p \in [0, x]$
	$f(x) = 0, x \notin (0, 1]$	$f(p) = 0, p \notin [0, x]$
2	$f(x) = 1, x \in [0, 1)$	$f(p) = 1/(1-x), p \in [0, 1-x]$
	$f(x) = 0, x \notin [0, 1)$	$f(p) = 0, p \notin [0, 1-x]$
	a	b
1	$f(a) = 1, a \in [0, 1]$	$f(b) = 1-b, b \in [0, 1]$
	$f(a) = 0, a \notin [0, 1]$	$f(b) = 0, b \notin [0, 1]$
2	$f(a) = 1-a, a \in [0, 1]$	$f(b) = 1, b \in [0, 1]$
	$f(a) = 0, a \notin [0, 1]$	$f(b) = 0, b \notin [0, 1]$

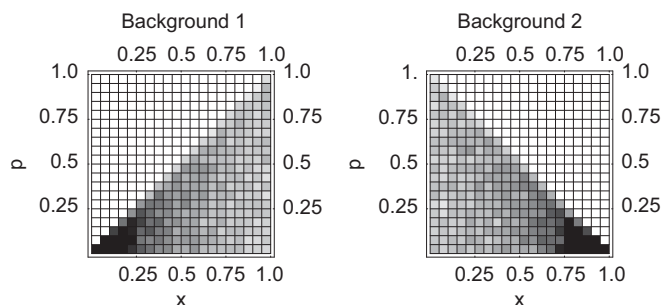


Fig. 2. Kinematic variable distributions for Backgrounds 1 and 2.

Table 2
Cut survival probabilities for each event type in the Toy model.

Event type	$P(A)$	$P(B)$
Background 1	0.5	0.25
Background 2	0.25	0.5

define our cuts on variables a and b as

$$A = (a > 0.5)$$

$$B = (b > 0.5). \quad (6)$$

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 2.

3.1. One background in the Toy model

In this section we discuss the case of a single significant background. The background prediction is given by Eq. (5). We generated 1×10^4 Background 1 events over the whole range of kinematic variables. This leaves ≈ 2200 background events in the signal region before applying cuts A and B . In Table 3, we show the observed number of events for each combination of cuts, the predicted background, and the observed background after applying both cuts A and B . The predicted background of 267.8 ± 20.6 agrees well with the 256 ± 16 observed background events.

Table 3

Single background study for the Toy model with only Background 1.

N_0	2236 ± 47
$N_{\overline{AB}}$	831 ± 29
$N_{\overline{A}\overline{B}}$	280 ± 17
$N_{\overline{A}\overline{B}}$	869 ± 29
Predicted background	267.8 ± 20.6
Observed background	256 ± 16

4. Two background case

4.1. Derivation

The previous derivation applied in the case of a single background source. If the background is made up of two different background sources, $N_0 = N_1 + N_2$, with different CSPs then there is a correlation introduced between cuts that must be accounted for. To do so, we begin by replacing Eqs. (2) and (4) with

$$N_{\text{bkg}} = N_1 P_1(A) P_1(B) + N_2 P_2(A) P_2(B)$$

$$N_{\overline{AB}} = N_1 P_1(A) P_1(\overline{B}) + N_2 P_2(A) P_2(\overline{B})$$

$$N_{\overline{A}\overline{B}} = N_1 P_1(\overline{A}) P_1(B) + N_2 P_2(\overline{A}) P_2(B)$$

$$N_{\overline{A}\overline{B}} = N_1 P_1(\overline{A}) P_1(\overline{B}) + N_2 P_2(\overline{A}) P_2(\overline{B}). \quad (7)$$

Then our previous calculation of the background has a cross term introduced. We wish to find the correction to the one background solution. We begin by substituting the above definitions into the solution for the one background case, Eq. (5):

$$\frac{N_{\overline{AB}} N_{\overline{A}\overline{B}}}{N_{\overline{AB}}} = \frac{1}{N_{\overline{AB}}} [N_1 P_1(A) P_1(\overline{B}) + N_2 P_2(A) P_2(\overline{B})] \times [N_1 P_1(B) P_1(\overline{A}) + N_2 P_2(B) P_2(\overline{A})]. \quad (8)$$

We expand the numerator:

$$N_{\overline{AB}} N_{\overline{A}\overline{B}} = N_1^2 P_1(A) P_1(\overline{A}) P_1(B) P_1(\overline{B}) + N_1 N_2 [P_1(A) P_2(\overline{A}) P_2(B) P_1(\overline{B}) + P_2(A) P_1(\overline{A}) P_1(B) P_2(\overline{B})] + N_2^2 P_2(A) P_2(\overline{A}) P_2(B) P_2(\overline{B}). \quad (9)$$

We multiply N_{bkg} by $N_{\overline{A}\overline{B}}$ to allow us to find the difference of Eq. (8) and N_{bkg} :

$$N_{\text{bkg}} N_{\overline{A}\overline{B}} = [N_1 P_1(A) P_1(B) + N_2 P_2(A) P_2(B)] \times [N_1 P_1(\overline{A}) P_1(\overline{B}) + N_2 P_2(\overline{A}) P_2(\overline{B})] = N_1^2 P_1(A) P_1(\overline{A}) P_1(B) P_1(\overline{B}) + N_1 N_2 [P_1(A) P_2(\overline{A}) P_1(B) P_1(\overline{B}) + P_2(A) P_1(\overline{A}) P_2(B) P_1(\overline{B})] + N_2^2 P_2(A) P_2(\overline{A}) P_2(B) P_2(\overline{B}) \quad (10)$$

$$\frac{N_{\overline{AB}} N_{\overline{A}\overline{B}}}{N_{\overline{AB}}} = N_{\text{bkg}} + \frac{N_1 N_2}{N_{\overline{AB}}} [P_1(A) P_2(\overline{A}) P_2(B) P_1(\overline{B}) + P_2(A) P_1(\overline{A}) P_1(B) P_2(\overline{B}) - P_2(A) P_1(\overline{A}) P_2(B) P_1(\overline{B}) - P_2(A) P_1(\overline{A}) P_2(B) P_1(\overline{B})]. \quad (11)$$

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, $P_i(\overline{A}) = 1 - P_i(A)$. Each element of the cross term has the same structure which can be expanded to

$$P_i(A) P_j(\overline{A}) P_k(B) P_l(\overline{B}) = P_i(A) P_k(B) (1 - P_j(A)) \times (1 - P_l(B)) = P_i(A) P_k(B) - P_i(A) P_j(A) P_k(B) - P_i(A) P_k(B) P_l(B) + P_i(A) P_j(A) P_k(B) P_l(B). \quad (12)$$

Summing the elements of the cross term cancels everything except the terms with two CSPs:

$$\frac{N_{\overline{AB}} N_{\overline{A}\overline{B}}}{N_{\overline{AB}}} = N_{\text{bkg}} + \frac{1}{N_{\overline{AB}}} (N_1 N_2 (P_1(A) P_2(B) + P_2(A) P_1(B)) - P_1(A) P_1(B) - P_2(A) P_2(B)) = N_{\text{bkg}} - \frac{N_1 N_2 (P_2(A) - P_1(A)) (P_2(B) - P_1(B))}{N_{\overline{AB}}}. \quad (13)$$

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

$$N_{\text{bkg}} = \frac{N_{\overline{AB}} N_{\overline{A}\overline{B}}}{N_{\overline{AB}}} + \frac{N_1 N_2}{N_{\overline{AB}}} \Delta_A \Delta_B. \quad (14)$$

The second term in Eq. (14) 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.

4.2. Properties of the two background solution

We now discuss the behavior of the Eq. (14). 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 Eq. (14) under extreme conditions. First, we show that under no conditions can the total N_{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, $N_{\overline{AB}} = 0$ and N_{bkg} is undefined. We therefore want to study N_{bkg} 's behavior as we approach this limit. We begin by setting $\Delta_B = -1$ and studying the limit as $\Delta_A \rightarrow 1$. In this case cut B removes Background 1 completely, but the Background 1 events which survive \overline{AB} still contribute to the prediction produced by Eq. (5).

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

$$P_1(B) = P_2(\overline{B}) = 0 \quad (15)$$

$$P_2(B) = P_1(\overline{B}) = 1. \quad (16)$$

Substituting these values into Eq. (7), we find

$$N_{\overline{AB}} = N_2 P_2(A) \quad (17)$$

$$N_{\overline{A}\overline{B}} = N_1 P_1(\overline{A}) = N_1 (1 - P_1(A)) \quad (18)$$

$$N_{\overline{A}\overline{B}} = N_2 P_2(\overline{A}) = N_2 (1 - P_2(A)). \quad (19)$$

We then substitute these values into Eq. (14) and sum the two terms:

$$N_{\text{bkg}} = \frac{N_1 (P_2(A) - P_2(A) P_1(A) - P_2(A) + P_1(A))}{1 - P_2(A)} = \frac{N_1 P_1(A) (1 - P_2(A))}{1 - P_2(A)} = N_1 P_1(A). \quad (20)$$

As $\Delta_A \rightarrow 1$, $P_1(A) \rightarrow 0$. Therefore the N_{bkg} goes to 0. This indicates that N_{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 Eq. (14) correctly predicts the background by simplifying Eq. (7), substituting $P(B)$ for $P_1(B)$ and $P_2(B)$:

$$N_{\text{bkg}} = (N_1 P_1(A) + N_2 P_2(A)) P(B) \quad (21)$$

$$N_{A\bar{B}} = (N_1 P_1(A) + N_2 P_2(A)) P(\bar{B}) \quad (22)$$

$$N_{\bar{A}B} = (N_1 P_1(\bar{A}) + N_2 P_2(\bar{A})) P(B) \quad (23)$$

$$N_{\bar{A}\bar{B}} = (N_1 P_1(\bar{A}) + N_2 P_2(\bar{A})) P(\bar{B}). \quad (24)$$

Substituting these values into the first term of (14) and simplifying gives

$$\frac{N_{A\bar{B}} N_{\bar{A}B}}{N_{A\bar{B}}} = (N_1 P_1(A) + N_2 P_2(A)) P(B). \quad (25)$$

This is N_{bkg} , so the correction is not necessary to properly predict the background when one of the Δ 's is zero.

4.3. 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 cuts A and B , but do have different cut survival probabilities. If $P_1(A)=0.75$ 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 do not 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 , Δ_A , and Δ_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, $N_{A\bar{B}}$, $N_{\bar{A}B}$, and $N_{\bar{A}\bar{B}}$, it is possible to effectively open the box and count N_{bkg} . Determining Δ_A and Δ_B also requires additional input. Their values can be derived from either Monte Carlo or data outside the signal region.

4.4. 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 2×10^4 . 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 (Eq. (14)). 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, Δ_A and Δ_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

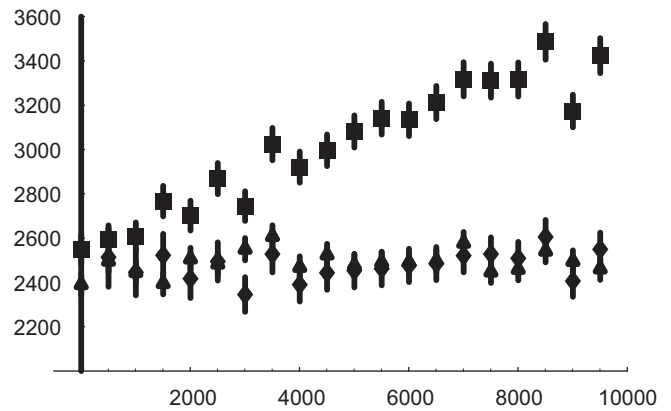


Fig. 3. Predicted and observed background for different admixtures of a Background 2. Squares are predicted background (without second background correction), triangles are the observed background in data, diamonds are the corrected prediction. The x -axis is the number of generated Background 2 events, the total number of events, $N_1 + N_2$, was held constant at 2×10^4 .

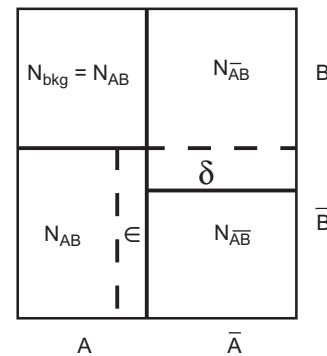


Fig. 4. Cut space with correlated cuts.

Background 2 events. Here only the predicted background without correction increases, while the observed and corrected backgrounds remain relatively flat.

5. Cut correlation

In the derivation of both the one and two background cases, we have assumed that the cuts A and B are uncorrelated. 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.

5.1. 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 (see Fig. 4).

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

$$N_{\text{bkg}} = N_0 P(A|B) P(B) = N_0 P(A) P(B|A) \quad (26)$$

$$N_{A\bar{B}} = N_0 P(A|\bar{B}) P(\bar{B}) = N_0 P(A) P(\bar{B}|A) \quad (27)$$

$$N_{\bar{A}\bar{B}} = N_0 P(\bar{A}|B)P(B) = N_0 P(\bar{A})P(B|\bar{A}) \quad (28)$$

$$N_{\bar{A}B} = N_0 P(\bar{A}|\bar{B})P(\bar{B}) = N_0 P(\bar{A})P(\bar{B}|\bar{A}). \quad (29)$$

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

$$\frac{N_{\bar{A}\bar{B}}N_{\bar{A}B}}{N_{\bar{A}\bar{B}}} = \frac{N_0^2 P(\bar{A}|\bar{B})P(\bar{B})P(\bar{A}|B)P(B)}{N_0 P(\bar{A}|B)P(\bar{B})}. \quad (30)$$

We assume the correlations are small and relate the different posterior probabilities to each other:

$$P(\bar{A}|\bar{B}) = P(\bar{A}|B) - \varepsilon \quad (31)$$

$$P(\bar{A}|B) = P(\bar{A}|\bar{B}) + \varepsilon \quad (32)$$

$$P(B|\bar{A}) = P(B|A) - \delta \quad (33)$$

$$P(\bar{B}|\bar{A}) = P(\bar{B}|A) + \delta. \quad (34)$$

The corrections ε and δ 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 ε and δ to be negligible. We have the freedom to choose to formulate the correction as a function of ε or δ . We choose to use ε . We substitute the definitions into Eq. (31):

$$\frac{N_{\bar{A}\bar{B}}N_{\bar{A}B}}{N_{\bar{A}\bar{B}}} = \frac{N_0 P(\bar{A}|B)(P(\bar{A}|B) - \varepsilon)P(B)}{P(\bar{A}|B) + \varepsilon} = \frac{N_0(P(\bar{A}|B)P(B) - \varepsilon P(B))}{1 + \frac{\varepsilon}{P(\bar{A}|B)}}. \quad (35)$$

Assuming the ε term in the denominator is small, we expand this result in ε :

$$\frac{N_{\bar{A}\bar{B}}N_{\bar{A}B}}{N_{\bar{A}\bar{B}}} \approx (N_0(P(\bar{A}|B)P(B) - \varepsilon P(B))) \times \left(1 - \frac{\varepsilon}{P(\bar{A}|B)} + \frac{\varepsilon^2}{P(\bar{A}|B)^2} + \mathcal{O}(\varepsilon^3)\right). \quad (36)$$

Multiplying this out and keeping the second order terms of ε gives

$$\frac{N_{\bar{A}\bar{B}}N_{\bar{A}B}}{N_{\bar{A}\bar{B}}} = N_0 P(\bar{A}|B)P(B) - \varepsilon N_0 \left(P(B) + \frac{P(\bar{A}|B)P(B)}{P(\bar{A}|B)}\right) + \varepsilon^2 N_0 \left(\frac{P(B)}{P(\bar{A}|B)} + \frac{P(\bar{A}|B)P(B)}{P(\bar{A}|B)^2}\right). \quad (37)$$

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

$$N_{bkg} = \frac{N_{\bar{A}\bar{B}}N_{\bar{A}B}}{N_{\bar{A}\bar{B}}} + \varepsilon N_0 P(B) \left(1 + \frac{P(\bar{A}|B)}{P(\bar{A}|B)}\right) - \varepsilon^2 N_0 \frac{P(B)}{P(\bar{A}|B)} \left(1 + \frac{P(\bar{A}|B)}{P(\bar{A}|B)}\right). \quad (38)$$

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:

$$P(B) = \frac{N_{AB} + N_{\bar{A}B}}{N_0} \approx \frac{N_{\bar{A}B}}{N_0} \quad (39)$$

$$\frac{P(\bar{A}|B)}{P(\bar{A}|B)} = \frac{N_{\bar{A}B}}{N_B} \approx \frac{N_{\text{pred}}}{N_{\bar{A}B}}. \quad (40)$$

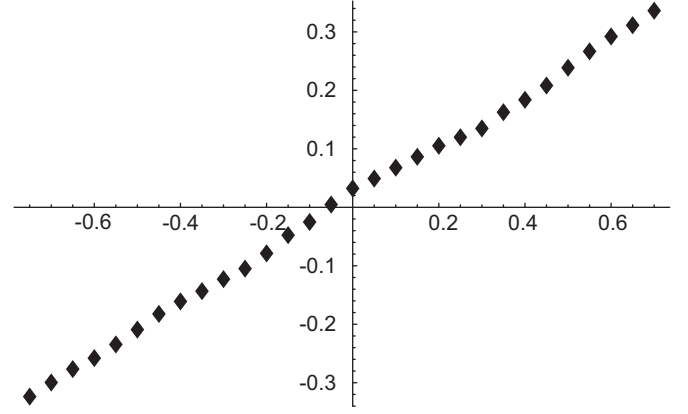


Fig. 5. The value of ε (Eq. (31)) as a function of ε' (Eq. (42)) in the Toy model.

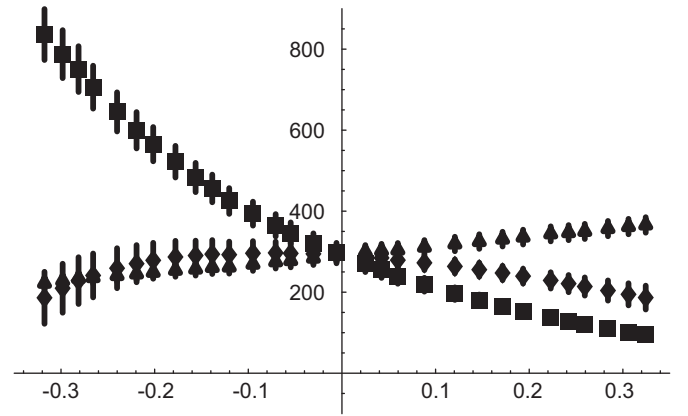


Fig. 6. Predicted and observed background for different levels of correlation between a and b in the Toy model. Triangles are data background, squares are predicted, and diamonds are predicted plus second order ε correction (Eq. (38)). The x -axis is ε .

These approximations give us the first order correction:

$$C_\varepsilon = \varepsilon N_{\bar{A}\bar{B}} \left(1 + \frac{N_{\text{pred}}}{N_{\bar{A}\bar{B}}}\right). \quad (41)$$

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 :

$$f(a) = (1 + \varepsilon'), \quad a \in [\varepsilon' b / (1 + \varepsilon'), (1 + \varepsilon') b / (1 + \varepsilon')]. \quad (42)$$

The variable ε' is the knob we use to tune the correlation. It is closely related to the variable ε that is defined in Eqs. (31) and (32) as is shown in Fig. 5.

We show the predicted and observed background in Fig. 6. As ε' increases the background in data increases, because the correlation increases the average value a , while the predicted background decreases (see Fig. 7).

In Fig. 6, we show the prediction with C_ε added. It improves the agreement for a fairly wide range of ε .

6. 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

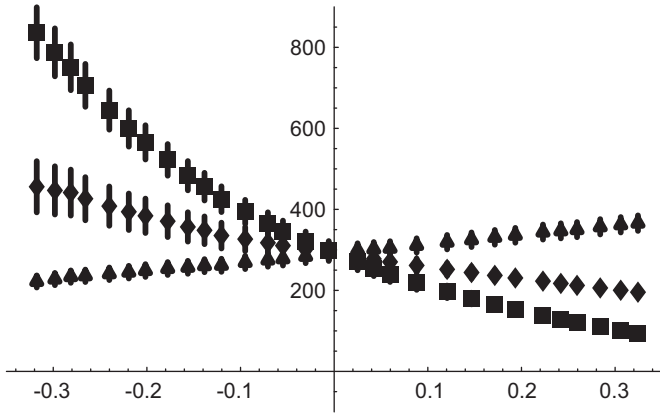


Fig. 7. Predicted and observed background for different levels of correlation between a and b in the Toy model. Triangles are data background, squares are predicted, and diamonds are predicted plus practical ε correction (Eq. (41)). The x -axis is ε .

apparatus consists of a 4π 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 $(1.4 \pm 0.4) \times 10^{-13}$ [4]. 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 π^0 's and two unobservable ν 's. Events are reconstructed by pairing the photons and calculating the π^0 decay vertices. The pairing with the smallest separation between π^0 vertices is selected. The signal region is defined in three kinematical variables: the transverse momentum of the π^0 – π^0 system, the invariant mass of the π^0 – π^0 system, $M_{\pi^0-\pi^0}$, and reconstructed decay vertex, Z .

The signal region is selected to avoid the two primary background sources: $K_L \rightarrow \pi^0 \pi^0 \pi^0$ 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 100 MeV/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 $100 \text{ MeV}/c < P_T < 200 \text{ MeV}/c$, $268 \text{ MeV}/c^2 < M_{\pi^0-\pi^0} < 450 \text{ MeV}/c^2$, $300 \text{ cm} < Z < 500 \text{ cm}$. 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 vetoes, and cut B which contained primarily cuts on photon reconstruction quality in the CsI array. Using Eq. (5), 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 $450 \text{ MeV}/c^2 < M_{\pi^0-\pi^0} < 550 \text{ MeV}/c^2$. The results of this are shown in Table 4. There is a significant discrepancy in the Low P_T region, this is due to neglecting the correlation.

We estimated the cut correlation on the $K_L \rightarrow \pi^0 \pi^0 \pi^0$ background in its peak region using both data and Monte Carlo. The value of ε was found to be (-0.049 ± 0.035) . We first apply Eq. (41) on the Low P_T region and find a correction of -6.67 ± 4.81 . This correction brings the prediction in Table 4 into much better agreement with data. We next apply Eq. (41) to the signal region and find a correction of (-0.12 ± 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

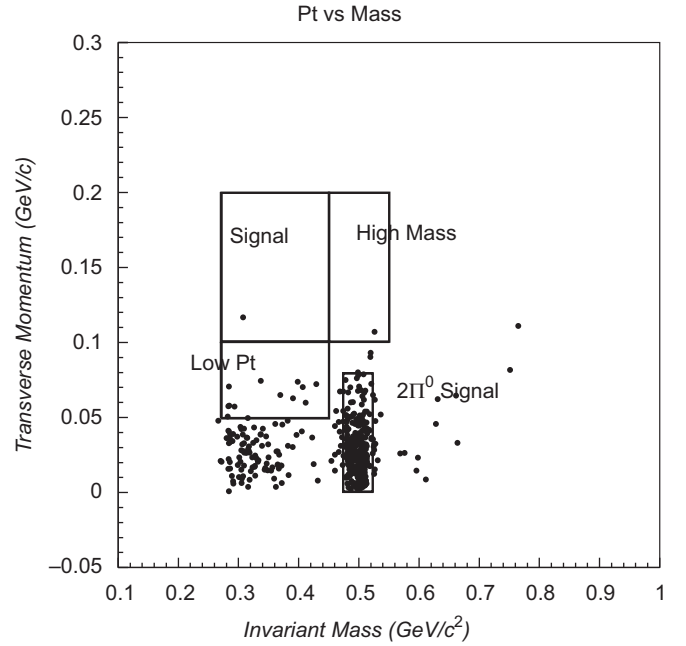


Fig. 8. P_T plotted against mass with all cuts in the E391a experiment. Z between 300 and 500 cm. The rectangular regions correspond to the regions in Table 4 [2].

Table 4

Prediction of background events in different regions without correlation or secondary background corrections.

Region	$N_{\bar{A}B}$	N_{AB}	$N_{\bar{A}B}$	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

Table 5

Values for the core neutron multiple background correction.

Δ_A	Δ_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 ± 0.05)

P_T –High Z region, $Z > 500 \text{ cm}$ and $P_T > 100 \text{ MeV}/c$. 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 $N_2 = 2.16 \pm 0.03_{\text{stat}} \pm 1.05_{\text{sys}}$. 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 5.

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 $0.43 \pm 0.32_{\text{stat.}} \pm 0.13_{\text{sys.}}$. Opening the signal box, we observed a single event, consistent with our background prediction.

7. 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 non-linear. 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 cuts A and B , so that the statistical errors on the terms of Eq. (5) are small. On the other hand, a large N_0 means the prediction is sensitive to small correlations between the cuts.

References

- [1] S. Adler, et al., Phys. Rev. Lett. 79 (1997) 2204.
- [2] J. Nix, et al., Phys. Rev. D 76 (2007) 011101(R).
- [3] Wolfram Research, Inc., Mathematica, Version 5.2, Champaign, IL, 2005.
- [4] C.W. Chiang, F.J. Gilman, Phys. Rev. D 62 (2000) 094026.