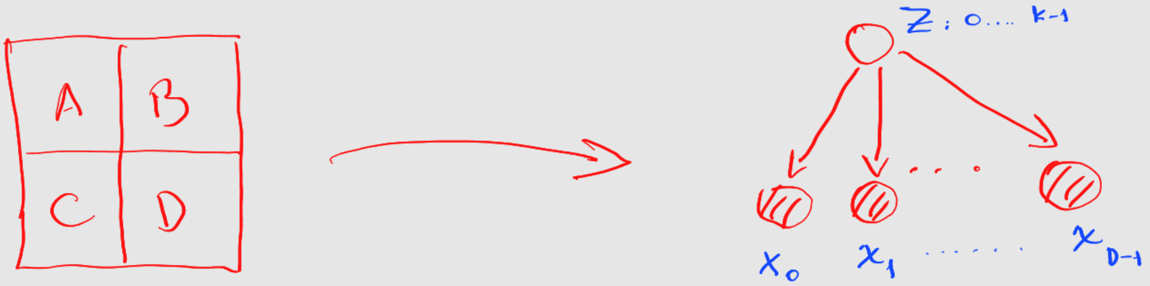
ABCD Method: improvement and generalization in a Bayesian Framework



- Two observable which must be conditionally indedendent
- Each observable has 2 outcomes
- Only 2 classes: signal and background
- signal should be only in one region of paremeter space: usually B

- Any number of observables that must be conditionally independent.
- Each observable can have any number of outcomes. Moreover, if they are continuous is usually better.
- It can have many classes. As for instance it can classify the different backgrounds and take the advantage of different prior knowledge for each one of them.
- Signal and backgrounds can be mixed in different proportions in all the phase space. One does not need a control region with no signal in it.

Application to a simplified example

PHYSICAL REVIEW LETTERS **129,** 081802 (2022)

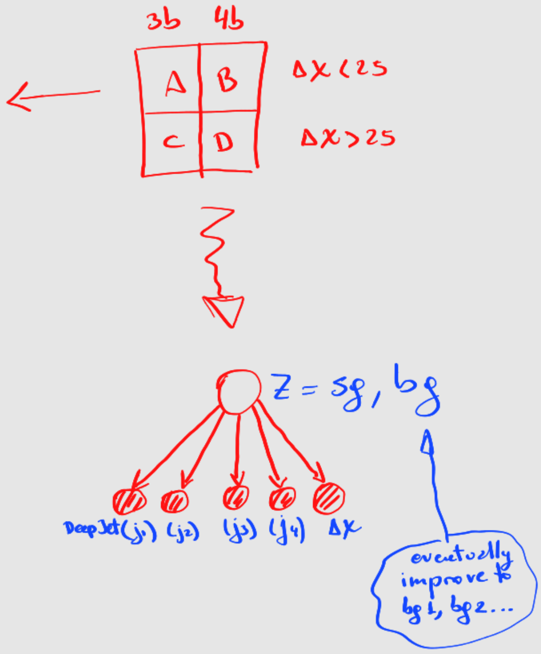
Search for Higgs Boson Pair Production in the Four b Quark Final State in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV

A. Tumasyan *et al.*[*]
(CMS Collaboration)

(Received 19 February 2022; accepted 7 July 2022; published 18 August 2022)

A search for pairs of Higgs bosons produced via gluon and vector boson fusion is presented, focusing on the four b quark final state. The data sample consists of proton-proton collisions at a center-of-mass energy of 13 TeV, collected with the CMS detector at the LHC, and corresponds to an integrated luminosity of 138 fb^{-1}. No deviation from the background-only hypothesis is observed. A 95% confidence level upper limit on the Higgs boson pair production cross section is observed at 3.9 times the standard model prediction for an expected value of 7.8. Constraints are also set on the modifiers of the Higgs field self-coupling, κ_λ, and of the coupling of two Higgs bosons to two vector bosons, κ_{2V}. The observed (expected) allowed intervals at the 95% confidence level are $-2.3 < \kappa_\lambda < 9.4$ ($-5.0 < \kappa_\lambda < 12.0$) and $-0.1 < \kappa_{2V} < 2.2$ ($-0.4 < \kappa_{2V} < 2.5$). These are the most stringent observed constraints to date on the HH production cross section and on the κ_{2V} coupling.

DOI: 10.1103/PhysRevLett.129.081802

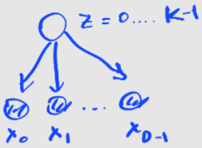


Possible road

1) Toy model where one can stess all the advantages. Compare some target using ABCD & Bayes.

2) Simplified di-Higgs example
   - MG-Pythia-Delphes
   - Shift true distribution
   - Use both methods.

ABCD Method as a particular case of a general Graphical Model

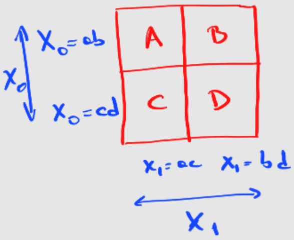


$$P(\vec{x}, z) = \prod_{k=0}^{K-1} \left(P(\vec{x} | z_k) \, P(z_k) \right)^{z_k} \rightsquigarrow P(\vec{x}) = \sum_{z_k} P(\vec{x}, z)$$

suppose $z \sim \text{Categorical}(\vec{\pi})$

$$P(\vec{x}, z) = \prod_{k=0}^{K-1} \left(\pi_k \cdot P(\vec{x} | z_k) \right)^{z_k} \rightsquigarrow \boxed{P(\vec{x}) = \sum_{k=0}^{K-1} \pi_k \cdot P(\vec{x} | z_k)}$$

Now suppose $K=2$ and $D=2$, then we have an ABCD method



$Z =$ signal, background

$$\therefore \; \underset{\substack{\uparrow \\ X_0, X_1}}{P(\vec{X})} = \pi_{bg} \Big(P(\vec{x}=A | bg) \, \delta_{\vec{x}A} + P(\vec{x}=B | bg) \cdot \delta_{\vec{x}B} + P(\vec{x}=C | bg) \cdot \delta_{\vec{x}C} + P(\vec{x}=D | bg) \, \delta_{\vec{x}D} \Big)$$

$$+ \pi_{sg} \Big(P(\vec{x}=A | sg) \, \delta_{\vec{x}A} + P(\vec{x}=B | sg) \, \delta_{\vec{x}B} + P(\vec{x}=C | sg) \, \delta_{\vec{x}C} + P(\vec{x}=D | sg) \cdot \delta_{\vec{x}D} \Big)$$

Therefore

$$P(\vec{x}=A) = \frac{N_A}{N} = \pi_{bg} \, P(\vec{x}=A | bg) + \pi_{sg} \underbrace{P(\vec{x}=A | sg)}_{}\,,$$

$$P(\vec{x}=B) = \frac{N_B}{N} = \pi_{bg} \, P(\vec{x}=B | bg) + \pi_{sg} \, P(\vec{x}=B | sg)$$

$$P(\vec{x}=C) = \frac{N_C}{N} = \pi_{bg} \, P(\vec{x}=C | bg) + \pi_{sg} \underbrace{P(\vec{x}=C | sg)}_{}\,,$$

$$P(\vec{x}=D) = \frac{N_D}{N} = \pi_{bg} \, P(\vec{x}=D | bg) + \pi_{sg} \underbrace{P(\vec{x}=D | sg)}_{}\,,$$

All of these are zero under the ABCD hypothesis. (only B has signal)

The other hypothesis is that x_0 and x_1 are independent. Therefore

$$P(\vec{x}=A | bg) = P(x_0 = ab | bg) \cdot P(x_1 = ac | bg)$$

and so on. Therefore we get:

$$\frac{N_A}{N} = \pi_{bg} P(x_0 = ab | bg) \cdot P(x_1 = ac | bg)$$

$$\frac{N_B}{N} = \pi_{bg} P(x_0 = ab | bg) P(x_1 = bd | bg) + \pi_s \underbrace{P(x_0 = ab | s) P(x_0 = bd | sg)}_{=1}$$

$$\frac{N_C}{N} = \pi_{bg} P(x_0 = cd | bg) \, P(x_1 = ac | bg)$$

$$\frac{N_D}{N} = \pi_{bg} P(x_0 = cd | bg) \, P(x_1 = bd | bg)$$

And from here it is straightforward to get:

$$\frac{N_D}{N_C} \cdot N_A = \pi_{bg} P(x_0 = ab | bg) \cdot P(x_1 = bd | bg) \cdot N \leftsquigarrow \text{expected backgroun events in region B}$$

And therefore

$$\boxed{N_S = N - \frac{N_D}{N_C} \cdot N_A} \leftsquigarrow \text{expected signal events} \checkmark$$