

# Ground truth inference of binary classifier accuracies - the independent classifiers case

The accuracy of binary classifiers can be estimated exactly without knowing the true labels for the classifier decisions. Three independent classifiers are sufficient. This notebook demonstrates the above claims by construction.

The method proposed here is algebraic as opposed to the maximum-likelihood methods commonly seen in ground truth inference algorithms. This approach has been patented by Data Engines Corporation.

---

## The Polynomial System - Algebra of Error for Independent Binary Classifiers

The following set of quartic polynomials define algebraic system that must be solved to estimate the accuracy of three classifiers as well as the true prevalence of the labels. There are  $8 = 2^3$  polynomials, one for each of the possible voting pattern of three classifiers doing binary label classification.

In[457]: **groundTruthPolynomials** =  $\{-f_{\alpha,\alpha,\alpha} + P_{\alpha} P_{1,\alpha,\alpha} P_{2,\alpha,\alpha} P_{3,\alpha,\alpha} + P_{\beta} P_{1,\alpha,\beta} P_{2,\alpha,\beta} P_{3,\alpha,\beta},$

$-f_{\alpha,\alpha,\beta} + P_{\alpha} P_{1,\alpha,\alpha} P_{2,\alpha,\alpha} P_{3,\beta,\alpha} + P_{\beta} P_{1,\alpha,\beta} P_{2,\alpha,\beta} P_{3,\beta,\beta},$

$-f_{\alpha,\beta,\alpha} + P_{\alpha} P_{1,\alpha,\alpha} P_{2,\beta,\alpha} P_{3,\alpha,\alpha} + P_{\beta} P_{1,\alpha,\beta} P_{2,\beta,\beta} P_{3,\alpha,\beta},$

$-f_{\alpha,\beta,\beta} + P_{\alpha} P_{1,\alpha,\alpha} P_{2,\beta,\alpha} P_{3,\beta,\alpha} + P_{\beta} P_{1,\alpha,\beta} P_{2,\beta,\beta} P_{3,\beta,\beta},$

$-f_{\beta,\alpha,\alpha} + P_{\alpha} P_{1,\beta,\alpha} P_{2,\alpha,\alpha} P_{3,\alpha,\alpha} + P_{\beta} P_{1,\beta,\beta} P_{2,\alpha,\beta} P_{3,\alpha,\beta},$

$-f_{\beta,\alpha,\beta} + P_{\alpha} P_{1,\beta,\alpha} P_{2,\alpha,\alpha} P_{3,\beta,\alpha} + P_{\beta} P_{1,\beta,\beta} P_{2,\alpha,\beta} P_{3,\beta,\beta},$

$-f_{\beta,\beta,\alpha} + P_{\alpha} P_{1,\beta,\alpha} P_{2,\beta,\alpha} P_{3,\alpha,\alpha} + P_{\beta} P_{1,\beta,\beta} P_{2,\beta,\beta} P_{3,\alpha,\beta},$

$-f_{\beta,\beta,\beta} + P_{\alpha} P_{1,\beta,\alpha} P_{2,\beta,\alpha} P_{3,\beta,\alpha} + P_{\beta} P_{1,\beta,\beta} P_{2,\beta,\beta} P_{3,\beta,\beta}\}$

Out[457]:  $\{-f_{\alpha,\alpha,\alpha} + P_{\alpha} P_{1,\alpha,\alpha} P_{2,\alpha,\alpha} P_{3,\alpha,\alpha} + P_{\beta} P_{1,\alpha,\beta} P_{2,\alpha,\beta} P_{3,\alpha,\beta},$

$-f_{\alpha,\alpha,\beta} + P_{\alpha} P_{1,\alpha,\alpha} P_{2,\alpha,\alpha} P_{3,\beta,\alpha} + P_{\beta} P_{1,\alpha,\beta} P_{2,\alpha,\beta} P_{3,\beta,\beta},$

$-f_{\alpha,\beta,\alpha} + P_{\alpha} P_{1,\alpha,\alpha} P_{2,\beta,\alpha} P_{3,\alpha,\alpha} + P_{\beta} P_{1,\alpha,\beta} P_{2,\beta,\beta} P_{3,\alpha,\beta},$

$-f_{\alpha,\beta,\beta} + P_{\alpha} P_{1,\alpha,\alpha} P_{2,\beta,\alpha} P_{3,\beta,\alpha} + P_{\beta} P_{1,\alpha,\beta} P_{2,\beta,\beta} P_{3,\beta,\beta},$

$-f_{\beta,\alpha,\alpha} + P_{\alpha} P_{1,\beta,\alpha} P_{2,\alpha,\alpha} P_{3,\alpha,\alpha} + P_{\beta} P_{1,\beta,\beta} P_{2,\alpha,\beta} P_{3,\alpha,\beta},$

$-f_{\beta,\alpha,\beta} + P_{\alpha} P_{1,\beta,\alpha} P_{2,\alpha,\alpha} P_{3,\beta,\alpha} + P_{\beta} P_{1,\beta,\beta} P_{2,\alpha,\beta} P_{3,\beta,\beta},$

$-f_{\beta,\beta,\alpha} + P_{\alpha} P_{1,\beta,\alpha} P_{2,\beta,\alpha} P_{3,\alpha,\alpha} + P_{\beta} P_{1,\beta,\beta} P_{2,\beta,\beta} P_{3,\alpha,\beta},$

$-f_{\beta,\beta,\beta} + P_{\alpha} P_{1,\beta,\alpha} P_{2,\beta,\alpha} P_{3,\beta,\alpha} + P_{\beta} P_{1,\beta,\beta} P_{2,\beta,\beta} P_{3,\beta,\beta}\}$

## The Ground Truth and the Inference Problem

In classification problems, the ground truth is the correct label for each data point in a dataset. The statistics of ground truth solved by the above polynomial system are:

- the prevalence of the labels.
- the accuracies of the classifiers for each label.

### Synthetic/Fake/Manufactured Ground Truth

The purpose of this notebook is to demonstrate the mathematics of ground truth inference with polynomials. So to keep it simple, we will be manufacturing our own ground truth. The following commands create the ground truth for this problem - a list of labels for the true label of each data point in the dataset.

```
In[403]:= desiredALabelPrevalence = 23 / 100
datasetSize = 10 000
labelGroundTruth =
  Table[If[RandomReal[] < desiredALabelPrevalence,  $\alpha$ ,  $\beta$ ], {datasetSize}];
```

```
Out[403]=  $\frac{23}{100}$ 
```

```
Out[404]= 10 000
```

Let's check that the ground truth makes sense

```
In[407]:= RandomSample[labelGroundTruth, 10]
```

```
Out[407]= { $\alpha$ ,  $\beta$ ,  $\alpha$ ,  $\beta$ ,  $\beta$ ,  $\beta$ ,  $\beta$ ,  $\alpha$ ,  $\beta$ ,  $\beta$ }
```

```
In[408]:= labelGroundTruth // Tally
```

```
Out[408]= {{ $\beta$ , 7675}, { $\alpha$ , 2325}}
```

Now we construct the synthetic data for the classifier label decisions. Again, we are using synthetic data to create a simple example that focuses on the math of the algorithm. You can substitute numbers from real independent classifiers if you have them.

1. The first thing we need to specify is the accuracy of each of the classifiers. It is this accuracy that we seek to estimate with the polynomial system that is constructed from the classifier decisions alone. I'll arbitrarily pick them uniformly from the same range. I need two accuracies, one for each label since we are doing binary classification.

```
In[409]:= classifierAccuracies = Table[
  {(* The  $\alpha$  accuracy *)  $\alpha \rightarrow$  RandomReal[{0.7, 0.9}],
    (* The  $\beta$  accuracy *)  $\beta \rightarrow$  RandomReal[{0.7, 0.9}]} // Association, {3}]
```

```
Out[409]= {<|  $\alpha \rightarrow 0.724074$ ,  $\beta \rightarrow 0.762341$  |>,
  <|  $\alpha \rightarrow 0.873416$ ,  $\beta \rightarrow 0.777599$  |>, <|  $\alpha \rightarrow 0.758298$ ,  $\beta \rightarrow 0.757377$  |>}
```

2. Now we need a function that takes a classifier's accuracies, a list true labels, and produces a sample of its classification decisions.

- OtherLabel is a convenience function to create incorrect classification decisions.

- SyntheticClassification is the actual function that produces a sample of what a classifier with the specified accuracies would produce on a dataset.

```
In[410]:= OtherLabel[α] = β
          OtherLabel[β] = α
```

```
Out[410]= β
```

```
Out[411]= α
```

```
In[412]:= SyntheticClassification[classifierAccuracies_Association, dataset_List] :=
          dataset //
          Map[
            If[
              (* Throw a die, if below accuracy,
               pick the true label, if not, pick the other label *)
              RandomReal[] < classifierAccuracies[#, #, OtherLabel[#]] &,
              #] &
```

3. The classifier decisions are next. Three lists, one for each classifier, as long as the dataset.

```
In[413]:= classifierDecisions =
          Map[SyntheticClassification[#, labelGroundTruth] &, classifierAccuracies];
```

Let's make sure things are okay

```
In[416]:= classifierDecisions[[1]] // {RandomSample[#, 10], # // Length} &
          classifierDecisions[[2]] // {RandomSample[#, 10], # // Length} &
          classifierDecisions[[3]] // {RandomSample[#, 10], # // Length} &
```

```
Out[416]= {{α, β, β, β, β, α, α, β, α, β}, 10 000}
```

```
Out[417]= {{β, β, β, β, β, β, α, β, α, α}, 10 000}
```

```
Out[418]= {{β, α, β, β, β, α, β, β, β, β}, 10 000}
```

Let's also make sure that we are getting noisy classification of the true labels

```
In[419]:= Map[Tally, classifierDecisions]
Out[419]= {{{β, 6385}, {α, 3615}}, {{β, 6287}, {α, 3713}}, {{α, 3653}, {β, 6347}}}
```

4. Now we can construct the left side input of the ground truth polynomials for independent binary classifiers - the frequency of the voting patterns

```
In[423]:= votingPatternCounts = classifierDecisions // Transpose // Tally
Out[423]= {{{β, β, α}, 1181}, {{β, β, β}, 3433}, {{α, α, β}, 695}, {{α, β, α}, 526},
          {{β, α, β}, 1072}, {{α, β, β}, 1147}, {{α, α, α}, 1247}, {{β, α, α}, 699}}
```

Let's sort them and pretty print them to see the pattern better

```
In[426]:= votingPatternCounts // Sort // Column
          {{α, α, α}, 1247}
          {{α, α, β}, 695}
          {{α, β, α}, 526}
          {{α, β, β}, 1147}
Out[426]= {{β, α, α}, 699}
          {{β, α, β}, 1072}
          {{β, β, α}, 1181}
          {{β, β, β}, 3433}
```

Everything looks okay with our synthetic data. Note that `votingPatternCounts` no longer contains any information about the true label for any of the data points in `labelGroundTruth`. Hence, any algorithm that only uses `votingPatternCounts` - the observed frequency of classifier voting patterns, would be carrying out ground truth inference. The goal of the next section is to estimate `classifierAccuracies` and the prevalence of the labels using only `votingPatternCounts`.

---

## Exact Solution for Three Independent Binary Classifiers

The above polynomial can be solved exactly for the unknown ground truth statistics given the tally of voting patterns by the three classifiers.

The solution for the prevalence is a quadratic equation for the prevalence of, say, label A and complicated coefficients based on the 8 voting pattern frequencies of the three classifiers over the dataset. This means that there are two solutions for prevalence -  $x\%$  or  $(1-x)\%$ . Likewise for the accuracies of the classifiers, although there is a symmetry between the accuracies for both labels.

This illustrates, in the case of classification tasks, the difference between accuracy and precision.

Curating a very small subset of the target dataset immediately establishes the correct solution. Accuracy is cheap, precision is expensive.

In general, these polynomial systems are solved using ideas from algebraic geometry such as a Groebner basis and Buchberger's algorithm.

### Choosing an independent variables set

There are two labels so we have two prevalences that are unknown. But, by definition, they must sum up to 100% in the dataset. So one variable is enough. It is an arbitrary choice, I picked  $P_\alpha$ .

Similar thing happens with the accuracies of the classifiers. There are two labels and they can give two answers for each label so each classifier has 4 variables. And similar to the case of the prevalence, there are two equations the 4 variables have to obey. I chose representing everything in terms of the probability of being correct given label.

If you put together the two paragraphs above, you can figure out that  $N$  independent classifiers would require  $2*N + 1$  variables to be solved for. So three classifiers should yield a polynomial system that can be reduced to 7 unknown variables ( $2*3 + 1$ ).

I've encapsulated the arbitrary choice of what those 7 variables need to be in the following rules

```
In[456]:= reduceToIndependentVariablesRules = {
  (* The prevalence *)
   $P_\beta \rightarrow (1 - P_\alpha)$ ,
  (* The alpha label rules for each of the three classifiers *)
  (*Sequence@@Flatten@Table[ $P_{i,\beta,\alpha} \rightarrow (1 - P_{i,\alpha,\alpha})$ , {i,3}]*)
   $P_{i,\beta,\alpha} \rightarrow (1 - P_{i,\alpha,\alpha})$ ,
  (* The beta label rules for each of the three classifiers *)
   $P_{i,\alpha,\beta} \rightarrow (1 - P_{i,\beta,\beta})$ 
}
```

```
Out[456]:= { $P_\beta \rightarrow 1 - P_\alpha$ ,  $P_{i,\beta,\alpha} \rightarrow 1 - P_{i,\alpha,\alpha}$ ,  $P_{i,\alpha,\beta} \rightarrow 1 - P_{i,\beta,\beta}$ }
```

And let's transform our polynomials into using an independent set of variables

```
In[458]:= groundTruthPolynomials /. reduceToIndependentVariablesRules
Out[458]:= {-f $_{\alpha,\alpha,\alpha}$  +  $P_\alpha P_{1,\alpha,\alpha} P_{2,\alpha,\alpha} P_{3,\alpha,\alpha}$  + (1 -  $P_\alpha$ ) (1 -  $P_{1,\beta,\beta}$ ) (1 -  $P_{2,\beta,\beta}$ ) (1 -  $P_{3,\beta,\beta}$ ),
  -f $_{\alpha,\alpha,\beta}$  +  $P_\alpha P_{1,\alpha,\alpha} P_{2,\alpha,\alpha}$  (1 -  $P_{3,\alpha,\alpha}$ ) + (1 -  $P_\alpha$ ) (1 -  $P_{1,\beta,\beta}$ ) (1 -  $P_{2,\beta,\beta}$ )  $P_{3,\beta,\beta}$ ,
  -f $_{\alpha,\beta,\alpha}$  +  $P_\alpha P_{1,\alpha,\alpha}$  (1 -  $P_{2,\alpha,\alpha}$ )  $P_{3,\alpha,\alpha}$  + (1 -  $P_\alpha$ ) (1 -  $P_{1,\beta,\beta}$ )  $P_{2,\beta,\beta}$  (1 -  $P_{3,\beta,\beta}$ ),
  -f $_{\alpha,\beta,\beta}$  +  $P_\alpha P_{1,\alpha,\alpha}$  (1 -  $P_{2,\alpha,\alpha}$ ) (1 -  $P_{3,\alpha,\alpha}$ ) + (1 -  $P_\alpha$ ) (1 -  $P_{1,\beta,\beta}$ )  $P_{2,\beta,\beta} P_{3,\beta,\beta}$ ,
  -f $_{\beta,\alpha,\alpha}$  +  $P_\alpha$  (1 -  $P_{1,\alpha,\alpha}$ )  $P_{2,\alpha,\alpha} P_{3,\alpha,\alpha}$  + (1 -  $P_\alpha$ )  $P_{1,\beta,\beta}$  (1 -  $P_{2,\beta,\beta}$ ) (1 -  $P_{3,\beta,\beta}$ ),
  -f $_{\beta,\alpha,\beta}$  +  $P_\alpha$  (1 -  $P_{1,\alpha,\alpha}$ )  $P_{2,\alpha,\alpha}$  (1 -  $P_{3,\alpha,\alpha}$ ) + (1 -  $P_\alpha$ )  $P_{1,\beta,\beta}$  (1 -  $P_{2,\beta,\beta}$ )  $P_{3,\beta,\beta}$ ,
  -f $_{\beta,\beta,\alpha}$  +  $P_\alpha$  (1 -  $P_{1,\alpha,\alpha}$ ) (1 -  $P_{2,\alpha,\alpha}$ )  $P_{3,\alpha,\alpha}$  + (1 -  $P_\alpha$ )  $P_{1,\beta,\beta} P_{2,\beta,\beta}$  (1 -  $P_{3,\beta,\beta}$ ),
  -f $_{\beta,\beta,\beta}$  +  $P_\alpha$  (1 -  $P_{1,\alpha,\alpha}$ ) (1 -  $P_{2,\alpha,\alpha}$ ) (1 -  $P_{3,\alpha,\alpha}$ ) + (1 -  $P_\alpha$ )  $P_{1,\beta,\beta} P_{2,\beta,\beta} P_{3,\beta,\beta}$ }
```

And let's check that we only have seven non-frequency variables

```
In[461]:= groundTruthPolynomials /. reduceToIndependentVariablesRules // Variables /@# & //
  Flatten // DeleteDuplicates // SortBy[#, First] &
Out[461]:= {f $_{\alpha,\alpha,\alpha}$ , f $_{\alpha,\alpha,\beta}$ , f $_{\alpha,\beta,\alpha}$ , f $_{\alpha,\beta,\beta}$ , f $_{\beta,\alpha,\alpha}$ , f $_{\beta,\alpha,\beta}$ ,
  f $_{\beta,\beta,\alpha}$ , f $_{\beta,\beta,\beta}$ ,  $P_\alpha$ ,  $P_{1,\alpha,\alpha}$ ,  $P_{1,\beta,\beta}$ ,  $P_{2,\alpha,\alpha}$ ,  $P_{2,\beta,\beta}$ ,  $P_{3,\alpha,\alpha}$ ,  $P_{3,\beta,\beta}$ }
```

## Preparing the observed voting frequencies rules

Recall that we have the voting pattern counts already

```
In[462]:= votingPatternCounts // Sort // Column
Out[462]:= {{ $\alpha$ ,  $\alpha$ ,  $\alpha$ }, 1247}
  {{ $\alpha$ ,  $\alpha$ ,  $\beta$ }, 695}
  {{ $\alpha$ ,  $\beta$ ,  $\alpha$ }, 526}
  {{ $\alpha$ ,  $\beta$ ,  $\beta$ }, 1147}
  {{ $\beta$ ,  $\alpha$ ,  $\alpha$ }, 699}
  {{ $\beta$ ,  $\alpha$ ,  $\beta$ }, 1072}
  {{ $\beta$ ,  $\beta$ ,  $\alpha$ }, 1181}
  {{ $\beta$ ,  $\beta$ ,  $\beta$ }, 3433}
```

Let's turn them into rules so we can substitute their value into the ground truth polynomials. We need to normalize them so we need the total count first. We now it is 10K, but let's get it by computation so we can check everything is okay

```
In[465]:= totalVotes = votingPatternCounts // Last /@# & // Total
totalVotes == datasetSize
```

```
Out[465]= 10 000
```

```
Out[466]= True
```

```
In[467]:= frequencyRules = votingPatternCounts //
Map[(Subscript[f, Sequence@@#[[1]]] → #[[2]] / datasetSize) &, #] &
```

```
Out[467]= {fβ,β,α →  $\frac{1181}{10\,000}$ , fβ,β,β →  $\frac{3433}{10\,000}$ , fα,α,β →  $\frac{139}{2000}$ , fα,β,α →  $\frac{263}{5000}$ ,
fβ,α,β →  $\frac{67}{625}$ , fα,β,β →  $\frac{1147}{10\,000}$ , fα,α,α →  $\frac{1247}{10\,000}$ , fβ,α,α →  $\frac{699}{10\,000}$ }
```

## The final polynomial set

```
In[468]:= finalPolynomialSet =
(groundTruthPolynomials /. reduceToIndependentVariablesRules) /. frequencyRules
```

```
Out[468]= {-  $\frac{1247}{10\,000}$  + Pα P1,α,α P2,α,α P3,α,α + (1 - Pα) (1 - P1,β,β) (1 - P2,β,β) (1 - P3,β,β),
-  $\frac{139}{2000}$  + Pα P1,α,α P2,α,α (1 - P3,α,α) + (1 - Pα) (1 - P1,β,β) (1 - P2,β,β) P3,β,β,
-  $\frac{263}{5000}$  + Pα P1,α,α (1 - P2,α,α) P3,α,α + (1 - Pα) (1 - P1,β,β) P2,β,β (1 - P3,β,β),
-  $\frac{1147}{10\,000}$  + Pα P1,α,α (1 - P2,α,α) (1 - P3,α,α) + (1 - Pα) (1 - P1,β,β) P2,β,β P3,β,β,
-  $\frac{699}{10\,000}$  + Pα (1 - P1,α,α) P2,α,α P3,α,α + (1 - Pα) P1,β,β (1 - P2,β,β) (1 - P3,β,β),
-  $\frac{67}{625}$  + Pα (1 - P1,α,α) P2,α,α (1 - P3,α,α) + (1 - Pα) P1,β,β (1 - P2,β,β) P3,β,β,
-  $\frac{1181}{10\,000}$  + Pα (1 - P1,α,α) (1 - P2,α,α) P3,α,α + (1 - Pα) P1,β,β P2,β,β (1 - P3,β,β),
-  $\frac{3433}{10\,000}$  + Pα (1 - P1,α,α) (1 - P2,α,α) (1 - P3,α,α) + (1 - Pα) P1,β,β P2,β,β P3,β,β}
```

Let's check that only the seven unknown statistics of the ground truth remain in the final polynomial system.

```
In[470]:= statisticsOfGroundTruth =
finalPolynomialSet // Variables /@# & // Flatten // DeleteDuplicates
```

```
Out[470]= {Pα, P1,α,α, P2,α,α, P3,α,α, P1,β,β, P2,β,β, P3,β,β}
```

## The “magical” solution

```
In[472]:= Solve[Map[# == 0 &, finalPolynomialSet], statisticsOfGroundTruth] // N
Out[472]= { { Pα → 0.237026, P1,α,α → 0.746384, P2,α,α → 0.872898,
             P3,α,α → 0.743696, P1,β,β → 0.758068, P2,β,β → 0.784526, P3,β,β → 0.752253 },
            { Pα → 0.762974, P1,α,α → 0.241932, P2,α,α → 0.215474, P3,α,α → 0.247747,
              P1,β,β → 0.253616, P2,β,β → 0.127102, P3,β,β → 0.256304 } }
```

Recall what the “real” ground truth is in this manufactured case

```
In[473]:= classifierAccuracies
Out[473]= { < | α → 0.724074, β → 0.762341 | >,
            < | α → 0.873416, β → 0.777599 | >, < | α → 0.758298, β → 0.757377 | > }
```

But note that classifierAccuracies are the dice we three to produce the dataset. So we should really be comparing our answer to the sample statistic - ground truth inference is always about the sample statistics as befits a non-parametric method!

Next time I’ll calculate the actual recognition of the data by classifier 1 so we can see how the solution from the ground truth polynomials, 0.746, compares with the sample and not with 0.724.