Electron Fake Rates

Claudio 7 Feb 2010
Outline

• Options for denominator (FO)
• FR from QCD (Pt30)
• Checks on Wjets and top
• QCD sample dependence
Numerator & Denominator (FO)

• The numerator is set

```cpp
bool electronSelection_cand01(const unsigned int index)
{
    if (!electronId_noMuon(index)) return false;
    if (!electronId_cand01(index)) return false;
    if (!electronImpact_cand01(index)) return false;
    if (electronIsolation_relsusy_cand1(index, true) > 0.10) return false;
    if (isFromConversionPartnerTrack(index)) return false;
    return true;
}
```

Also $P_T > 10$ and $|\eta| < 2.4$

• For denominator (FO) explored 3 choices
  ① No $d_0$ cut, no ID cut, relax isolation to 0.4 (V1)
    • Extrapolate in $d_0$, ID, and isolation
  ② No $d_0$ cut, no ID cut (V2)
    • Extrapolate in $d_0$ and ID
  ③ No $d_0$ cut, relax isolation to 0.4 (V3)
    • Extrapolate in $d_0$ and isolation
Denominator (FO) considerations (1)

• Isolation extrapolations can introduce jet $P_T$ dependence
  – Think of 30 GeV jets. If the electron is 20 GeV, there is only $\sim 10$ GeV (maximum) of extra stuff ($S$), so $\text{iso} = S/P_T$ goes between 0 and $\sim 0.5$. OTOH, for 60 GeV jets, it would go btw 0 and $\sim 2$. So definitely do not want to extrapolate from too far away in iso. Here we picked 0.4 since this is something we used in the past.

• ID efficiency can also have some jet $P_T$ dependence
  – The higher the jet $P_T$, the more “stuff” is around, the more likely it is that something lands on top of the “electron” and spoil ID variables like $h/e$.
  – Nothing much that can be done here, except to pick “reasonable” jet $P_T$’s, ie, jet $P_T$’s similar to the ones in your analysis.
Denominator (FO) considerations (2)

• Extrapolations in ID can depend on the Heavy Flavor content of the sample.
  – If I have many real electrons from $b \rightarrow e$, the prob. that a FO passes ID cuts is larger than in a sample depleted in $b \rightarrow e$

• Not sure that I have much smart to say about $d_0$ and conversion removal
  – Except that $d_0$ is sensitive to both $b \rightarrow e$ and residual conversions
Comments about choices of FO

• V2 (extrapolation in $d_0$ and ID)
  – Expect that it minimizes the jet dependence
  – Expect to be most appropriate to BG where jets have similar HF content as jets in QCD, ie, W+jets

• V1 (like V2 but adding isolation extrapolation)
  – Should still work for W+jets, should work better than V2 for BG dominated by top.
  – Expect more jet dependence
  – Numerically smaller FR than V2, which is an advantage

• V3 (extrapolation in $d_0$ and isolation)
  – Should work for top
Sample & FR binning

Derive the FR from QCD_Pt30_Summer09-MC_31X_V3_7TeV-v1/V03-00-35/

• No special handling of leading jet, trigger bias, ...
  – These are systematic studies for later

• Binning similar to Jμ-Oli-Ingo (JHI)
  – 3 bins on $P_T$ (10-20  20-60  >60)
  – 2 bins in $|\eta|$ (0-1.5  1.5-2.4)
    • 1.5 is also the boundary between barrel & endcap hardwired in the electron selection code. JHI bin boundaries is 1.4xx (close)
  – Should probably consider finer binning, not for now
**Comments**

- PT variations not too large  
  (good, maybe we do not need too many more bins)
- eta variations significant, but the cuts are explicitly different in barrel and endcap
- V2 and V3 fake rates are a bit higher than we would like
Checks on Wjets and top (1)

• Can we predict the fakes in W and top events?
  – Fakes = true fakes or electrons from heavy flavors
  – W events are the background for WW and ttbar analyses
  – Top events are the background for SUSY analyses

• Start from $\mu e$ standard hypotheses in our ntuple
  – Reminder of reason for $\mu e$ later

• Truth match the $\mu$ to $W \rightarrow \mu$ or $W \rightarrow \tau \rightarrow \mu$ or $t \rightarrow \mu$ or $t \rightarrow \tau \rightarrow \mu$
  – Use leptonIsFromW() function
  – Similar in spirit to what $J_\mu$-Oli-Ingo have done in the past, but different in small details
Checks on Wjets and top (2)

- Require the electron in the hypothesis to NOT be from $W \rightarrow e$ etc
  - Again using the leptonIsFromW() function
  - Necessary for top sample, redundant for $W$ sample
- Make sure that electrons is not from $W \rightarrow \mu\gamma\nu$ followed by $\gamma$ conversion
  - This is a source of BG, but it is not a source of BG that the FR method is designed for
  - Code: if(els_mc_id() [iele] == 22 && abs(els_mc_motherid() [iele]) == 13) continue
- Require that electron and muon be separated by 0.4
  - So that muon does not affect the electron isolation
  - Not a big deal (I don’t think)
Checks on Wjets and top (3)

• Require that the electron be either
  1. A good electron passing the numerator cuts
  2. A denominator electron failing the numerator cuts
• For the 2nd type, weight by FR/(1-FR)
• Compare the (ele) $P_T$ distribution of (1) with the (weighted) $P_T$ distribution of (2)
  – The former is “the observed” the latter is “the prediction”
  – Do it both for SS and OS
• If the FR is “good”, the two distributions should agree
Aside: why $\mu e$ (reminder)

- We could have also selected ee events, with one of the electrons truth matched, and then studied the $2^{nd}$
- We would have $\sim$ doubled the statistics!
- Problem is that technically $W \rightarrow e\gamma$ followed by $\gamma$ conversion is not easy to identify in our ntuples
- To get rid of this issue, we only look at truth matched $W \rightarrow \mu\nu$ events
More details

**MC samples are Madgraph:**
TTbarJets-madgraph_Summer09-MC_31X_V3_7TeV-v2/V03-00-35
WJets-madgraph_Summer09-MC_31X_V3_7TeV-v1/V03-00-35

Plots are number of raw events out of MC files

Did not bother to propagate the errors on the fake rate predictions.
In the top sample, below 20 GeV we underestimate by about a factor of two. Above 20 GeV it is not so bad, within about 30%. The undershoot in the W OS below 20 GeV is worrisome.
The results in the W sample are pretty good. Probably better than V1.
The top sample is more problematic.
The point of V3 (no extrapolation in ID) is to do better on the top. Not clear that we succeeded.
Observations

• No matter what we do, we can get it to a factor of 2. Getting better than 20% is problematic
• The ttbar and WW analyses have Wjets as the main background and require $P_T > 20$.
  – All attempts (V1,V2,V3) give reasonable results in this regime (within ~ 30%?).
  – But careful because below 20 starts to break down
  – V2 is probably best, but the FR is a bit large
    • Can we stand it? If not, use V1, or perhaps try a V4 with a tighter isolation, ie, intermediate between V1 and V2
• The SUSY analyses have ttbar as a BG and require $P_T > 10$
  – They are much tougher to get right
  – In some cases, the total can be OK, but the $P_T$ distributions are sick (undershoot below 20 and overshoot above 20 can compensate)
    • Not clear what to recommend at this point
Check of jet dependence

• Use the FR calculated in QCD_Pt30 to predict the number of electrons seen in QCD_Pt80.
• We hope that the FR from QCD_Pt30 still works, more or less
• This is actually a pretty severe test. Almost a factor of 3 in jet $P_T$!
• Next page: plots of observed vs expected
As anticipated, the V2 Fake Rate is the most stable wrt jet $P_T$ variations.

The V1 and V3 Fake Rates show large variations, especially for $P_T>20$. This is a bit disconcerting.
Concluding Remarks, Status

• For the ttbar and WW analyses (dominated by Wjets BG) we could use the V2 FR which works well for $P_T > 20$ and is quite stable wrt jet $P_T$ variations.
• The problem with the V2 FR is that numerically it is a bit large (see pg 8)
• Will try a V4 which is intermediate btw V2 and V1, ie, starts from V2, adding a smaller isolation extrapolation (say 0.2 or 0.3 instead of 0.4)
  – The hope will be that we can get small numerical values of FR but keeping the jet dependence under control
• The SUSY analyses (top background) are more problematic
  – Need to do some more thinking
• Want to do some thinking to see what ID cut would make sense to play with