J/ψ inclusive selection using the Electron Weight

Claudia Patrignani — Genova

1 Introduction

Inclusive decays to J/Ψ are one of the few channels in which we can detect charmonium formation, and we need efficient and selective criteria to recognize the high energy electron pairs coming from J/Ψ decay.

This memo describes the criteria used to select the $J/\psi + X$ candidates with the "Electron Weight" (EW).

The EW (already used in E760 but never described in detail) is essentially a modified version of the Neyman-Pearson test with some assumptions and approximations needed to get an estimation of the probability distribution also with a large number of discriminating variables.

2 Neyman-Pearson test or method of Likelihood-Ratios

A standard method[1] to discriminate between two hypoteses is to compare the value of the likelihood obtained in each case:

$$\mathcal{L}(hyp\,1) \quad vs \quad \mathcal{L}(hyp\,2)$$

A selection criterium is defined in terms of the ratio between the values of the likelihood functions for the different hypoteses:

$$\rho = \frac{\mathcal{L}(hyp\,1)}{\mathcal{L}(hyp\,2)}.$$

To apply the method it is necessary to define what the likelihood functions are for the given hypoteses, or equivalently, to write the "true" probability

[1] See for instance: Frodesen,Skieggestad,Tofte: "Probability and Statistics in Particle Physics", Universitetforlaget, Oslo (Chapter 14)

probability density distribution $f(x_1, \cdots, x_n)$ as a function $\{x_i\}$, since the likelihood for each hypotesis can then be written as:

$$\mathcal{L} = f(x_1, \cdots, x_n),$$

This test is known as Neyman-Pearson test and is well known in classification problems: the optimal classification (minimum number of fake-assigments) is obtained by assignign each event to the class with the higher probability (i.e. when $\rho = 1$ also known as Bayes decision rule).

When dealing with a classification problem on events which pass some preliminary selection, the optimal classification on these events is obtained using the conditional probabilities:

$$\rho = \frac{f_{hyp\,1}(x_1, \cdots, x_n|prel.sel.)}{f_{hyp\,2}(x_1, \cdots, x_n|prel.sel.)} \frac{Prob_{hyp\,1}(prel.sel.)}{Prob_{hyp\,2}(prel.sel.)}$$

From the "bookmaker's" point of view this is the most natural choice (to bet on the one which has the highest probability in the given condition).

One interesting feature that can be derived from the Bayes rule (often overlooked) is that even if the probability for the preliminary selection is unknown we can still classify events using an other classifier ρ' defined as

$$\rho' = \frac{f_{hyp\,1}(x_1, \cdots, x_n|prel.sel.)}{f_{hyp\,2}(x_1, \cdots, x_n|prel.sel.)}$$

which minimizes mis-identification fraction in the sample of events passing the preliminary selection:

$$F = \frac{N_{12}}{N_1} + \frac{N_{21}}{N_2}$$

3 Application of the method in E835

EW can be seen as an example of application of the Neyman-Person test (and of its effectiveness in discrimination problems).[2]

[2] Another example was the "Electron Quality" (EQ) proposed by R. Cester: a particular version of this test comparing the electron hypotesis versus the null hypotesis, in which the probability density funtion was normalized to to its maximum value:

$$EQi = \frac{f(hyp\,1)}{\max(f(hyp\,1))} \qquad (\ 0 \leq EQi \leq 1\ )$$

Actually the "optimal classification" is not what we really want, because it takes into account the absolute number of mis-assigment, while we are more interested in the signal-to-noise ratio. In this case the method gives no prescription on the cut value for ρ, which must be chosen empirically. Moreover we don't want to use the absolute probabilities (which depend on the cross section): in our case it is better to discriminate events based on the value of ρ', choosing a cut value as a compromise between efficiency (correct assignment fraction on signal) and background cross section (proportional to the correct assignemnt fraction for background).

The two hypoteses are of course "electron" and "background", and the choice of the variables is limited to the ones provided by E835[3] detector exacly with the aim of giving a characteristic signature for high energy electrons, i.e. dE/dx in hodoscopes, Čerenkov pulse height and energy deposit in the calorimeter:

$H1\ mip$: dE/dx in H1 (corrected by θ dependence)

$HP\ mip$: dE/dx in H2' (corrected by θ dependence)

$H2\ mip$: dE/dx in H2 (corrected by θ dependence)

Č ele : Čerenkov pulse height (correcting by θ dependence and different mirrors photoelectron yield)

$s.m.(r,w)$: Cluster second moments along rings and wedges:

$$s.m.(\text{ring}) = \frac{\sum_{r,w=1}^{3} E(r,w) * (r - r_0) * *2}{\sum_{r,w=1}^{3} E(r,w)}$$

$$s.m.(\text{wedge}) = \frac{\sum_{r,w=1}^{3} E(r,w) * (w - w_0) * *2}{\sum_{r,w=1}^{3} E(r,w)}$$

[3]The only difference between E760 and E835 EW definition is in the variable choice (we had a different inner detector) and also in the Čerenkov performances.

$3\times 3/5\times 5$: Ratio of the energy deposit in a 3×3 divided by that in a 5×5 grid surrounding the highest energy block:

$$3\times 3/5\times 5 = \frac{\sum_{r,w=1}^{3} E(r,w)}{\sum_{r,w=1}^{5} E(r,w)}$$

$2\times 2/4\times 4$: Ratio of the energy deposit in a 2×2 defined by the 2 highest energy blocks divided by that in the 4×4 grid enclosing them:

$$2\times 2/4\times 4 = \frac{\sum_{r,w=max1}^{max2} E(r,w)}{\sum_{r,w=max1-1}^{max2+1} E(r,w)}$$

Cl.mass : Cluster mass, defined as the invariant mass of the energy deposits in each block.

Note that in these definitions all calorimeter blocks surrounding the central one are taken into account, even if they are associated to a different cluster.

In order to apply the method we need the conditional probability distribution, which have been determined from the data.[4]

From the practical point of view there are two possible ways of doing so:

a) assume an "a priory" knowledge (modulo some parameter) of the probability density distribution (and their correlations) and use a maximum likelihood method to define the value of these parameters;

b) make no assumption on the functional form and use the (normalized) histogram of the variables as a probability distribution.

[4] While nobody would ever complain for the use of conditional probabilities, some people may object to the fact that we are using the data to define "a posteriori" their probability distribution function. However the idea of "training" an algorithm with real data is used in most non parametric classification approaches (e.g in Artificial Neural Networks or Nearest Neighbours) and has a solid theoretical ground in the Bayes definition of probability.

The second approach is the one used in the definition of the EW. It has the advantage that no assumpion is made on the background probability distribution, but poses severe problems on the accuracy of the derived probability distribution due to the limited size of the event sample.

The major disadvantage is that it is practically impossible to determine the $n-$dimensional probability distribution and the variables must be considered as independent.

This is not a real disadvantage of this approach: for real electrons we know that the source of correlation is in the calorimeter granularity and the correlation is small; for the background the kind and amount of correlation depends on the particular source of background, i.e. is expected to be different in the case of gamma conversions or hadronic showers, and it would have been difficut, also with the other approach, to give some reasonable assumption.

In a certain way our inability to determine the correlation matrix in the second approach matches our "a priori" ignorance on the background origin, and the best descripion of our ignorance can be expressed by assuming the variables as independent.

From the practical point of view the question was never addressed, since in E760 we could perform a sufficiently effective test even discarding variable correlations on background and in E835 the improved Čerenkov ph.e. yield results in an even better performance.

If the variables are independent the conditional probabilities can obviously be written as

$$f^{el,bck}(x_1,\ldots,x_n) = \prod_i f^{el,bck}(x_i)$$

in this case also the parameter ρ' could be written as

$$\rho' = \frac{\prod_i f_i^{el}(x_i)}{\prod_i f_i^{bkg}(x_i)} = \prod_i \frac{f_i^{el}(x_i)}{f_i^{bkg}(x_i)} = \prod_i W_i \equiv \mathrm{EW}$$

The EW is given by the product of the individual weights for each variable.

The condition imposed to the events before extracting the probability density distribution are the preliminary selection and obviously the trigger.

Based on the discussion above, the EW is meaningful only for events which satisfy those conditions. The preliminary selection used to define the

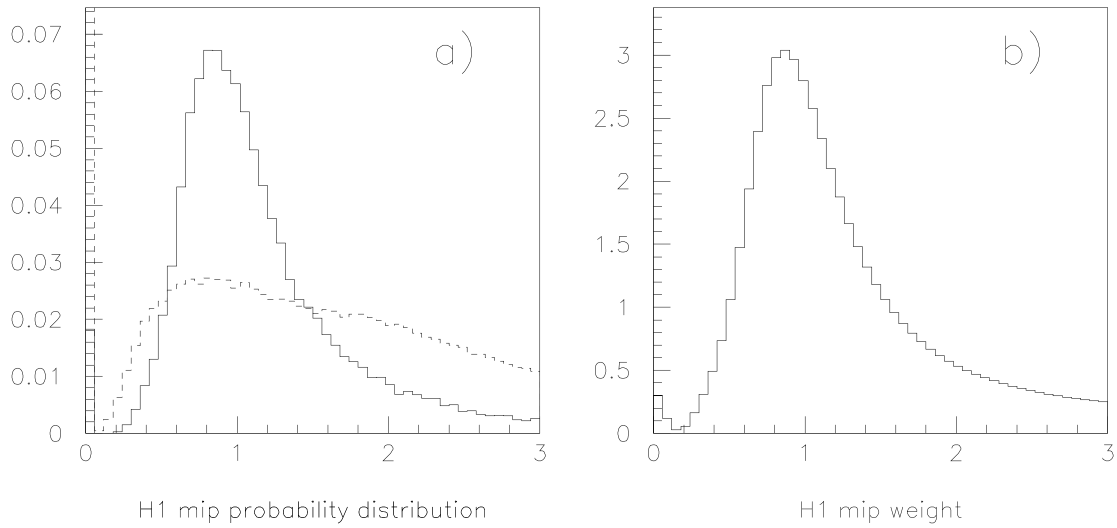

Figure 1: *a)* Probability distribution for electrons (solid) and background (dashed); *b)* Weight

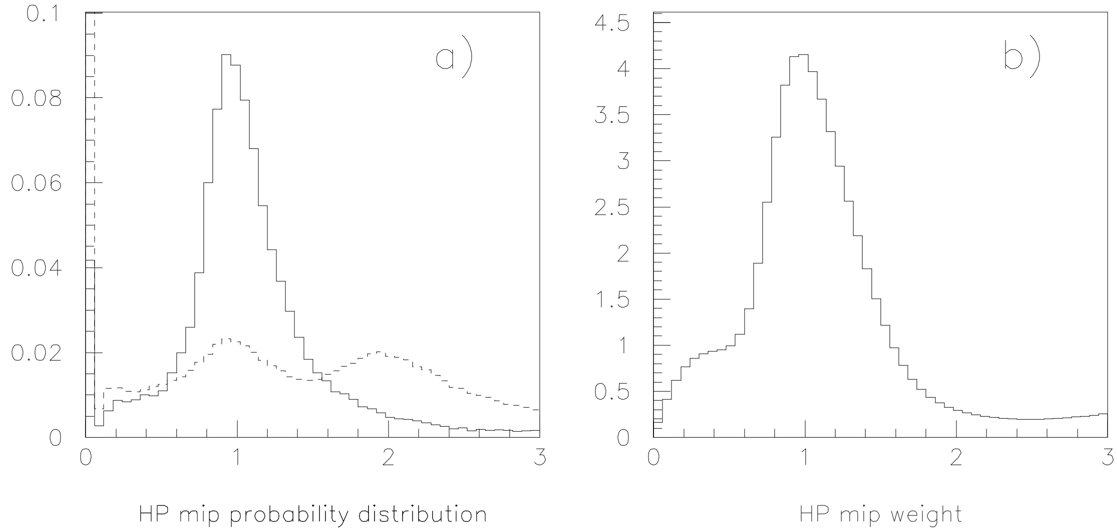

Figure 2: *a)* Probability distribution for electrons (solid) and background (dashed); *b)* Weight

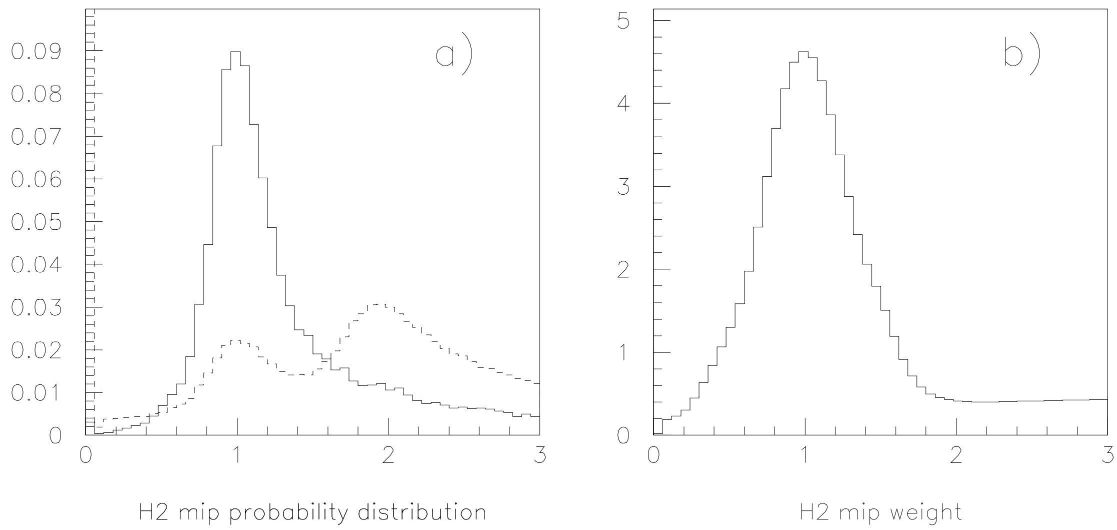

Figure 3: *a)* Probability distribution for electrons (solid) and background (dashed); *b)* Weight

EW is similar to that used in E760 and requires that the CCAL cluster be associated to some ADC signal (above pedestal) in at least 2 hodoscopes and to some ADC signal (above pedestal) in the proper Čerenkov mirror. Since in E835 there is no inefficiency in the septum region, no events without a Cerenkov are accepted.

A "good" event must have a pair of genuine electrons, therefore it is natural to discriminate on the product EW1*EW2.

The probability density functions of these variables (fig.1a to fig.9a) have been determined on a clean sample of real electrons (from J/Ψ and χ_2 scans, requiring only preliminary selection, kinematical fit and event topology) and on a clean sample of background events (simply given by all events passing the preliminary selection from the η_c' scan[5]). The weight for each variable is plotted in figures 1b to 9b.

[5]The few real J/Ψ events in these runs cannot be considered a contamination...!

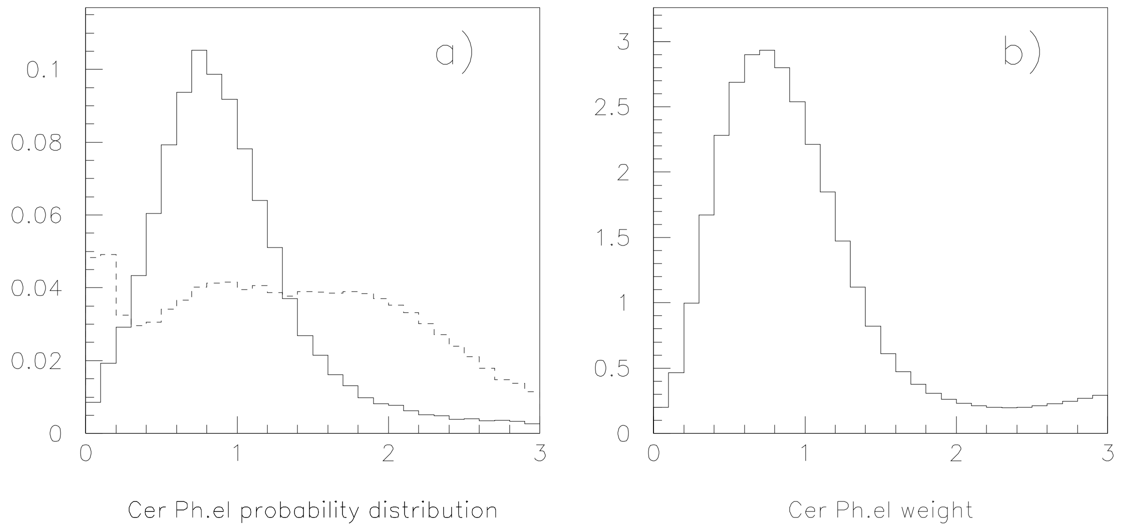

Figure 4: *a)* Probability distribution for electrons (solid) and background (dashed); *b)* Weight

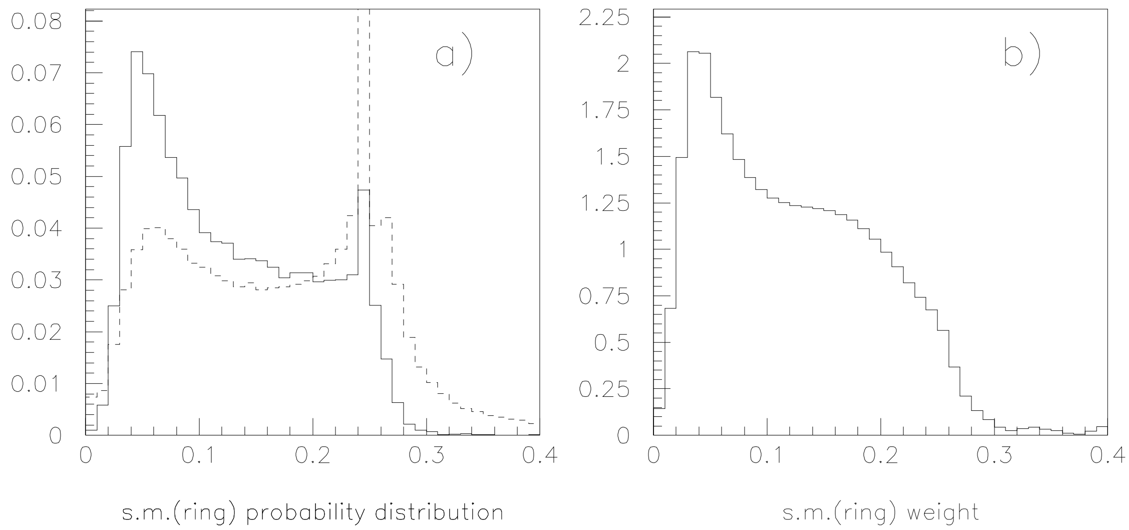

Figure 5: *a)* Probability distribution for electrons (solid) and background (dashed); *b)* Weight

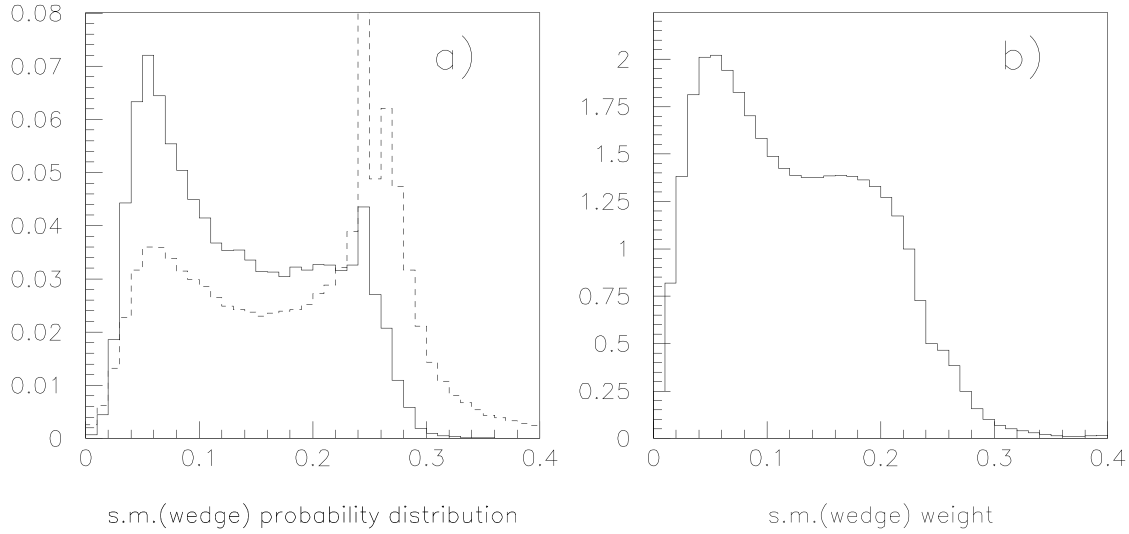

Figure 6: *a)* Probability distribution for electrons (solid) and background (dashed); *b)* Weight

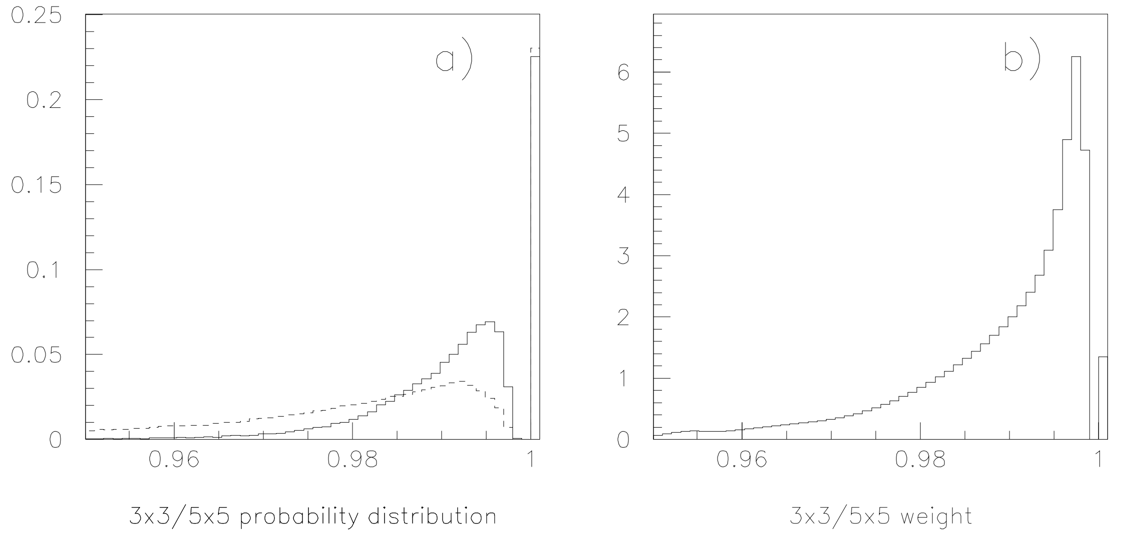

Figure 7: *a)* Probability distribution for electrons (solid) and background (dashed); *b)* Weight

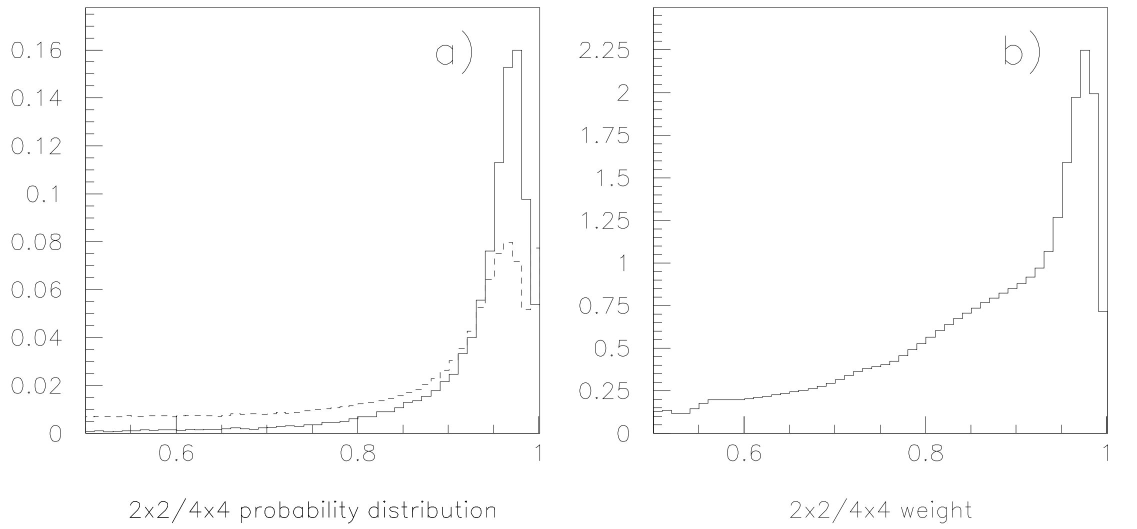

Figure 8: *a)* Probability distribution for electrons (solid) and background (dashed); *b)* Weight

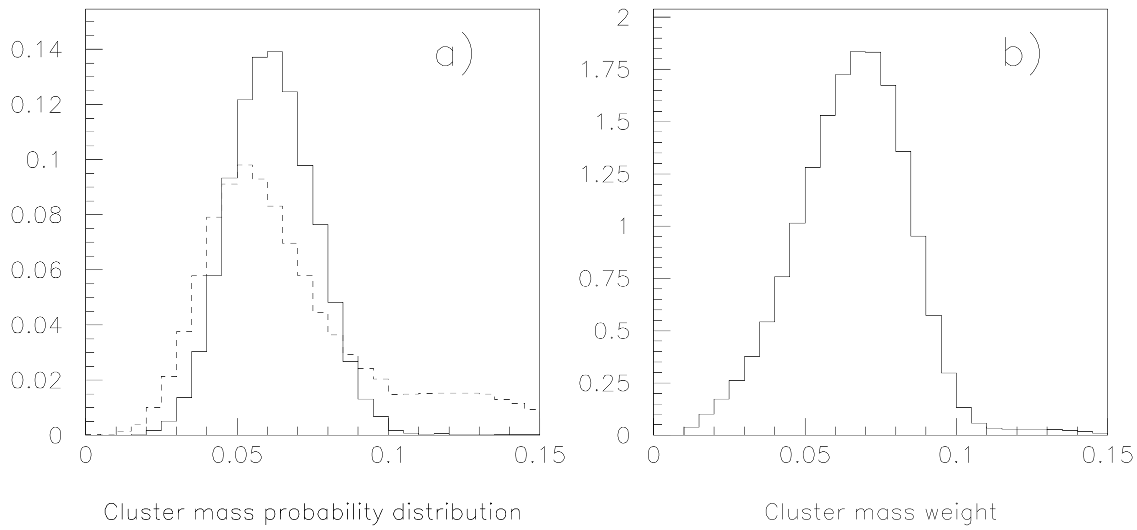

Figure 9: *a)* Probability distribution for electrons (solid) and background (dashed); *b)* Weight

4 Efficiency and background rejection

The distribution of EW1*EW2 for J/Ψ and for background events (after the preliminary selection) are plotted in fig.10 (arbitrary scale).

Since the method doesn't give any prescription for a cut value on EW, we can only make a choice based on the background rejection compared to the efficiency.

The test efficiency (including that of the preliminary selection which is performed before evaluating EW) has been measured on a clean sample of J/Ψ events from the "gold" data selected using only kinematical fit and topology (the efficiency of the preliminary selection used to define the EW is shown in the plot as the upper limit to the EW efficiency) :

$$\epsilon = \int_{T_{cut}}^{\infty} f(T) dT$$

were $f(T)$ is the probability density distribution of the test variable, EW1*EW2.

The fraction of background rejected by the EW1*EW2 cut (on events already passing the preliminary selection) is given by

$$\mathcal{R} = \int_{-\infty}^{T_{cut}} f(T) dT$$

The efficiency and background rejection vs the value of EW1*EW2 are plotted in fig 11.

The last plot shows the e^+e^- invariant mass of events selected requiring EW1*EW2> 10 during the '96 search of the η_c' ($\approx$ 10.9 pb^{-1}. The J/ψ inclusive peak can be clearly seen. A rough estimate of the background cross section under the peak yields a value of $\sim 6\,pb$, while the efficiency of the whole analysis (i.e. neglecting trigger efficiency and geometrical acceptance) is $\sim 90\%$ (assuming it is not very different from the efficiency determined on $J/\psi \to e^+e^-$).

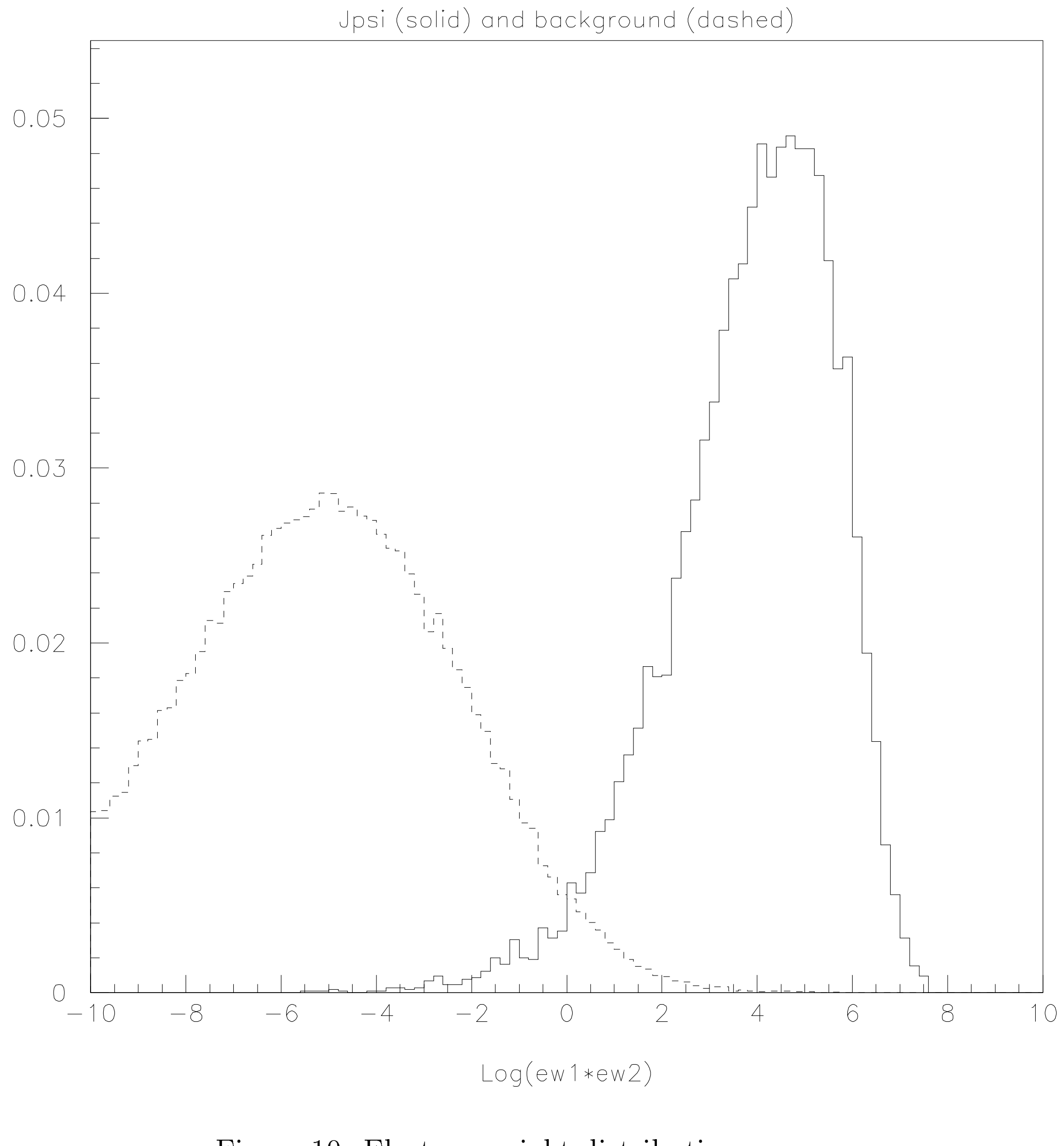

Figure 10: Electron weight distribution

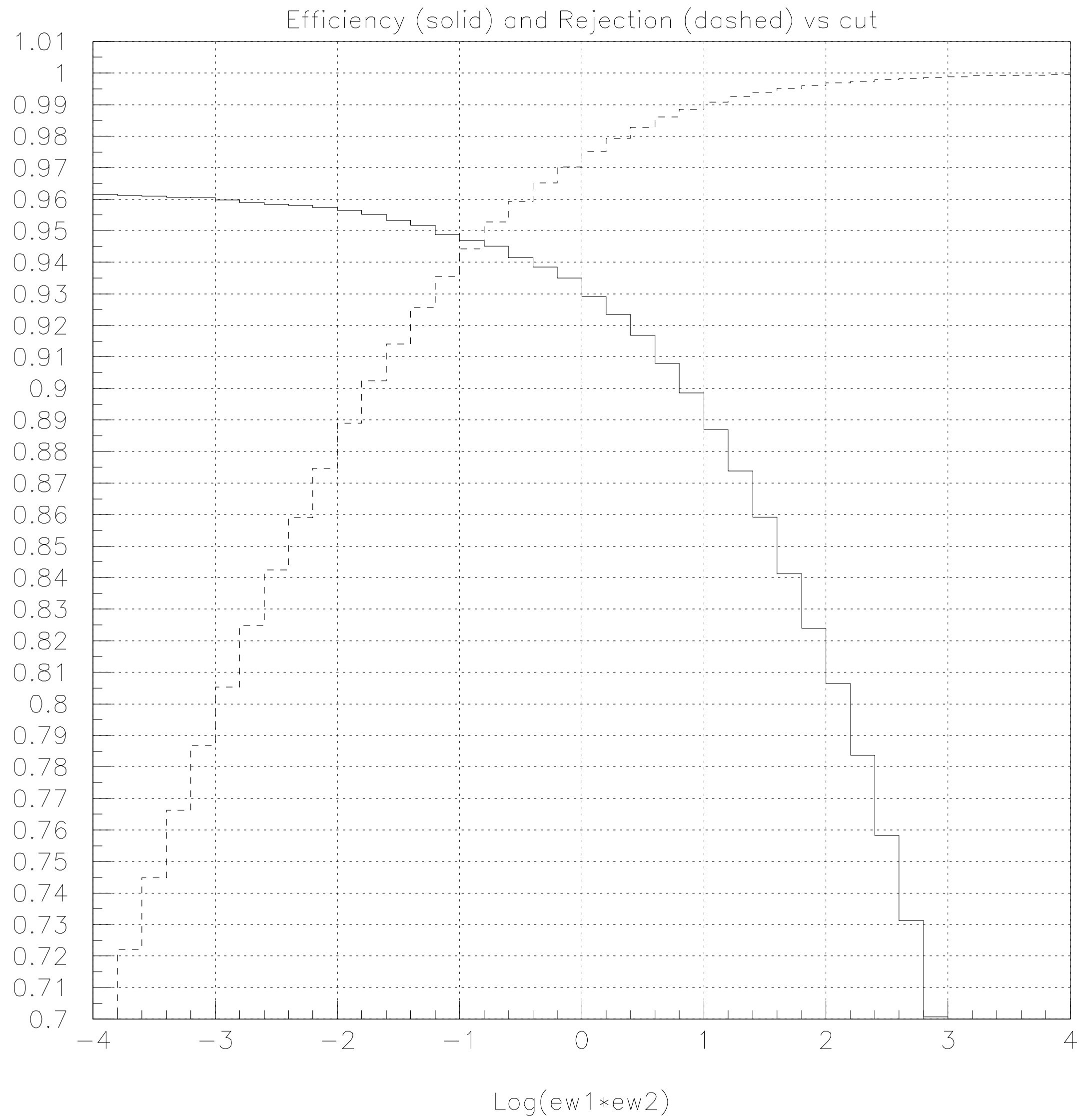

Figure 11: Electron weight efficiency and Background rejection

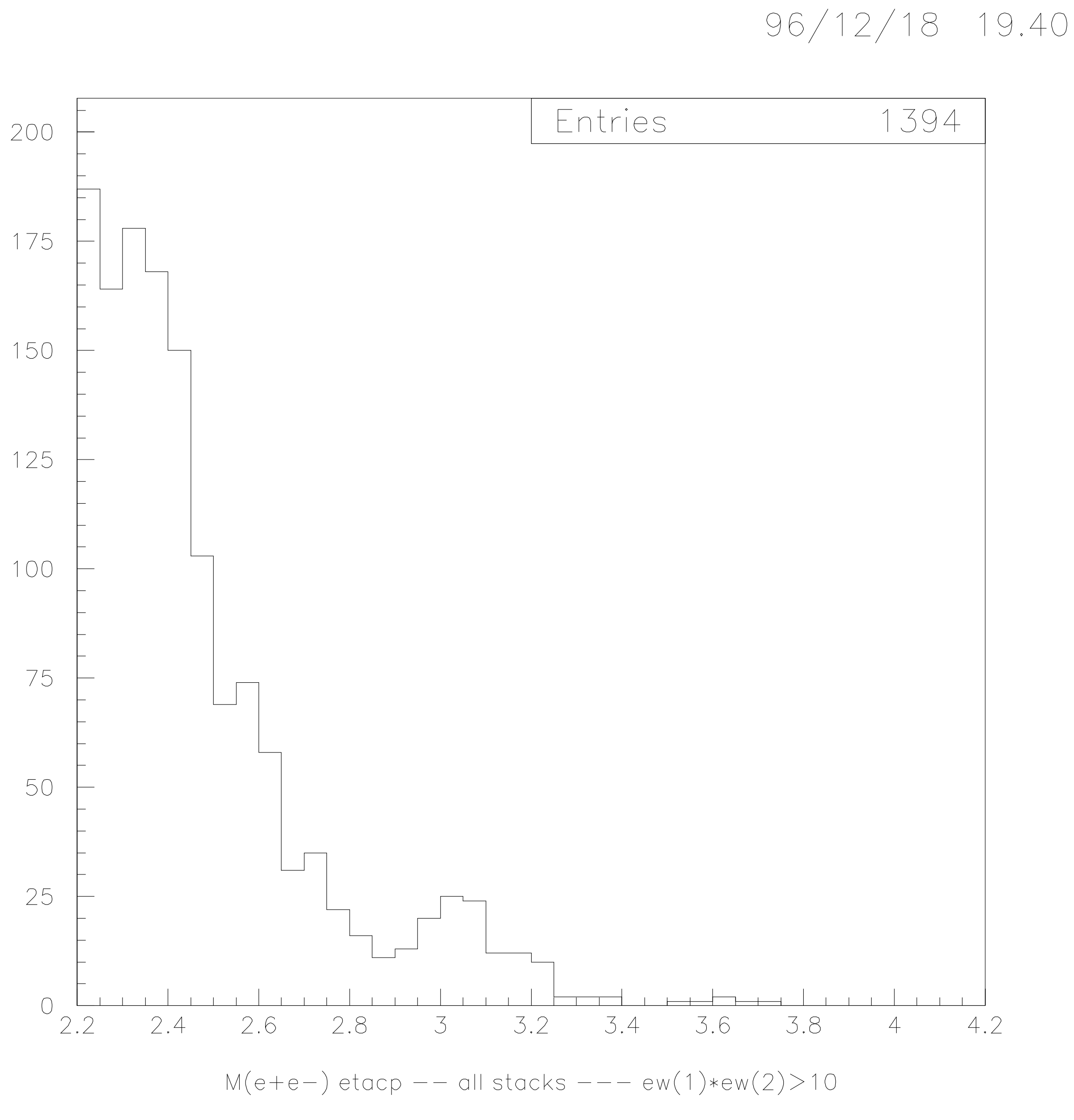

Figure 12: e^+e^- invariant mass obtained requiring EW1*EW2> 10 on $10.9\,pb^{-1}$ collected during the η_c' search