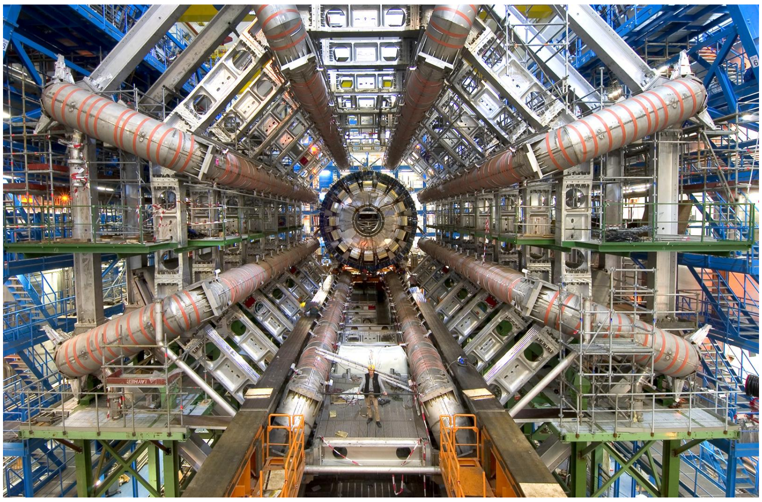
Corrections of the energies of electrons in the barrel/endcap transition region of the ATLAS electromagnetic calorimeter using Multivariate techniques

Chrysanthi Moni

Supervisor: Bruno Lenzi
Co-supervisor: Nansi Andari

August 7, 2014



Introduction

The main purpose of this study is the correction for the energy losses of the $e^{\pm}$ in the transition region between the barrel and the end-caps of the Electromagnetic Calorimeter (EMCal) of ATLAS, by using Multivariate techniques. The crack region is the one with the largest amount of material upstream the EMCal and this is the reason for which $e^{\pm}$ lose a great part of their energy as they pass through it. In this project, the contribution of the Multivariate Analysis in the correction of the E/E_{true} distribution as well as in the derivation of the Gaussian peak versus $|\eta|$ and E_T^{true}, is examined. η is the pseudorapidity used as a spatial coordinate for the description of the angle of a particle relative to the beam axis and $E_T^{true} = E_{true}/cosh(|\eta|)$, where E_{true} is the true energy of the particles. Finally, the improvement of the resolution by using MVA techniques with and without scintillator is also explored.

Electromagnetic Calorimeter

The sampling Electromagnetic Calorimeter of ATLAS is a lead liquid-argon (LAr) detector with accordion shaped kapton electrodes and lead absorber plates. Its goal is to measure the energies of the charged and neutral particles, through their electromagnetic interactions. The liquid-argon is chosen because of its intrinsic linear behavior, its intrinsic stability of the response over time and its intrinsic radiation hardness as well. The accordion geometry provides full φ coverage, where φ is the azimuthal angle rotating around the beam axis (z-axis), and symmetry without any azimuthal cracks and fast extraction of the signal. [PHD Thesis, Bruno Lenzi 2010].
The EMCal consists of a barrel in the $|\eta| < 1.475$ region and two end-caps in higher η regions, $1.375 < |\eta| < 3.2$. Between the barrel and the end-caps there is an overlap region, $1.375 < |\eta| < 1.52$, which is the region with the largest amount of material before the calorimeter. In this region particles lose part of their energy because of their interactions with the passive material at the end of the barrel and in front of the end-caps (cryostats), energy depositions in the active layers of the accordion and due to lateral and longitudinal leakages. In addition to the LAr detectors, between the barrel and the end-caps there are scintillators that absorb part of the electromagnetic showers and can be used to correct the energy of the particles as well. Finally, there are two presamplers both in front of the barrel and in front of the end-caps. The presamplers, thin active liquid-argon layers, are responsible for the correction of the energy lost as the particles pass through the passive material, such as the cryostat walls in front of the calorimeter [The Cern Large Hadron Collider:Accelerator and Experiments vol.1].
The EMCal is segmented into active layers. There are three active layers in the precision-measurement η region, $0 < |\eta| < 2.5$, two active layers in the $2.5 < |\eta| < 3.2$ region, and finally two active layers in the overlap region between the barrel and the end-caps, $|\eta| < 1.475$ and $|\eta| > 1.375$ respectively.

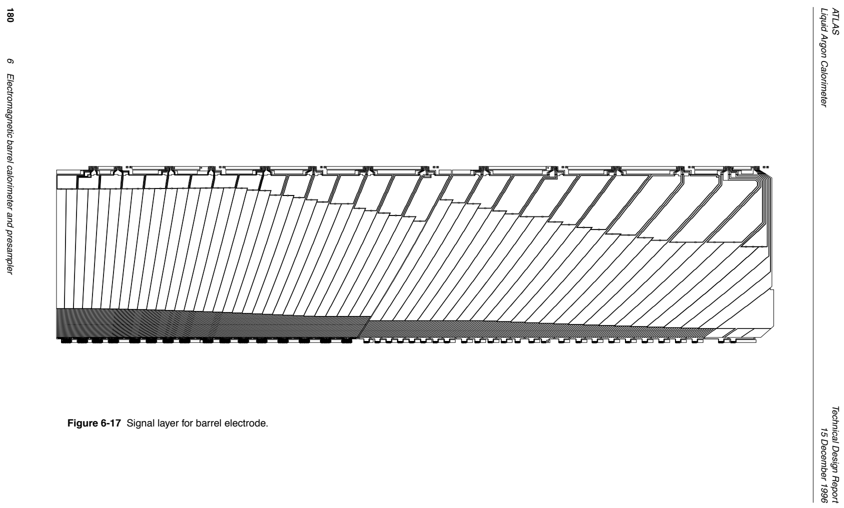


Figure 1: Signal layer for barrel electrode

In the precision-measurement η region, the fine segmentation of the front layer ("strips") leads to a precisely position measurement, with high precision in η. In addition, the middle layer receives the bulk of the electromagnetic shower as it is the largest one, and the measurements are done with the same accuracy both in η and φ. Finally, there is a back layer which collects the tail of the electromagnetic shower. In this layer the measurements are done more precisely in φ than in η [The Cern Large Hadron Collider:Accelerator and Experiments vol.1].

Multivariate (MVA) Calibration

The EMCal of ATLAS measures the energy of the particles, which is the sum of their energy deposition in the three longitudinal layers of the calorimeter and it is lower than their true energy due to energy losses caused by their interactions with the passive material in front of the barrel as well as in front of the end-caps, the deposition of their energy in the active layers of the accordion and due to lateral and longitudinal leakages, as mentioned above. The main purpose of the energy calibration in the EMCal is the correction for these energy losses as well as for the residual dependences in the energy response E/E_{true} and the determination of the true energy of the particles. The calibration can be implemented using detailed Monte Carlo simulations for the estimation of the energy losses.
For the purpose of this project Multivariate (MVA) techniques are used for the correction of $e^{\pm}$ energies, measured in the EMCal of ATLAS and the results produced by this method are compared with the results extracted by the current calibration method, denoted as Standard calibration (Std). The main goal of the MVA calibration is to estimate the true energies of the particles by using quantities measured by the detector. In order to achieve that, MVA calibration

corrects for energy losses before, inside and after the calorimeter and also for losses originating from the reconstruction. It allows the introduction of arbitrary number of input variables and it provides an easiest way to derive new set of corrections as well as an improved calibration with better resolution [Lenzi B. and Turra R.,11 Oct 2013].
The MVA calibration is derived by using the TMVA (Toolkit for MultiVariate Analysis) framework. "The TMVA provides a Root-integrated environment for the processing and the evaluation of multivariate classification and regression" [A. Hoecker et all, TMVA 4]. The MVA classification and regression are composed of two phases: the training and the evaluation. The training consists of training the sample and defining the target and the implementation of the input variables. The input variables used for the purpose of this study are the following:
1) the total energy in the accordion, E_{acc}, which is the sum of the uncalibrated energies in the accordion layers (strips, middle,back)
2) the energies in the two presamplers, one in front of the barrel and one in front of the end-caps, divided by the total energy in the accordion
3)$\eta_{cluster}$, the pseudorapidity in the ATLAS frame which takes into account the misalignment of the detector, in order to correct for the variation of the material in front of the accordion
4)η_{Calo}, the pseudorapidity with respect to the cell edge, defined as the η modulus the width of one cell of the middle layer (0.025), in the frame of the EMCal
5)the energies in the barrel's as well as in the end-caps' layers
6)$\frac{E_{s1}}{E_{s2}}$, the sum of the particles energies in the first layers of both barrel and end-caps divided by the sum of their energies in the second layers of barrel and end-caps
7)the $E_{scintillator}/E_{acc}$, the energy measured in the scintillator divided by the total energy in the accordion
8) $phiModCalo = fmod((el_{clphiCalo}) * 128/\pi, 1)$, φ with respect to the cell edge, in the calorimeter frame. It corrects for the sampling fraction as well as for the lateral leakages outside the calorimeter [Lenzi B. and Turra R.,11 Oct 2013].

The main goal of the MVA regression is to minimize the resolution, the smaller the resolution the better, behavior that tends to improve the output energy of the particles, in order to be as close to their true energy as possible.The estimation of the energy resolution in the EMCal was performed with beams of electrons and pions before putting them in the Atlas detector, and it was found to be, with the subtraction of the noise,

$$\frac{\sigma(E)}{E} = \frac{10\%}{\sqrt{E(GeV)}} \oplus 0.7\%. \tag{1}$$

The Multivariate method used in our study is the "boosted decision trees" (BDT), which works with no special configuration.

Results and Discussion

The histogram of E/E_{true}, where E is the reconstructed energy of the $e^{\pm}$ and E_{true} their true energy, ideally is a Gaussian distribution with the mean at 1. In our case, because of the energy losses mentioned above, the distribution is not Gaussian but is shifted from 1 with a tail towards low energies. The use of the MVA techniques contributes to the correction of this distribution in order to be as narrow as possible with peak at 1, for the achievement of the best resolution. Except for the E/E_{true} histogram, there are also the plots of the peak position (Gaussian peak) of this quantity versus $|\eta|$ and $E_T^{true} = E_{true}/cosh(|\eta|)$, computed in different $|\eta|$ and E_T^{true} bins. Finally, shown below are the figures of the resolution of E/E_{true}, which is quantified by the interquartile range, as a function of $|\eta|$ and E_T^{true}.
The MVA regression is implemented in different $|\eta|$ and E_T^{true}bins. The choice of the E_T^{acc} bins, (0-43,43-87,87-4e5) , for the single electrons of our study, ranges from low E_T^{acc} values up to $\sim 1TeV$, according to the plot below:

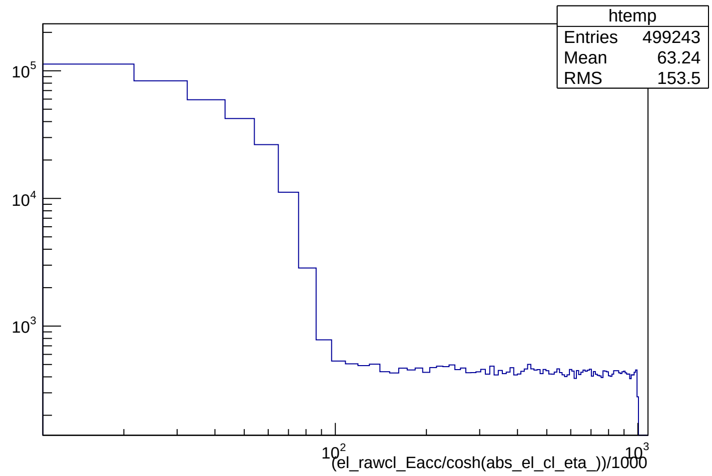


Figure 2: Histogram of E_T^{acc}. The E_T^{acc} ranges from a few GeV up to 1 TeV.

The $|\eta|$ bins are chosen from the graph of $E_{scintillator}/E_{acc}$ versus $|\eta|$. The reason for which the $|\eta|$ bins are defined from this plot is because the energy in the scintillator shows how the $e^{\pm}$ showers behave as a function of $|\eta|$ and also because it gives information about the amount of energy lost before and after the calorimeter.

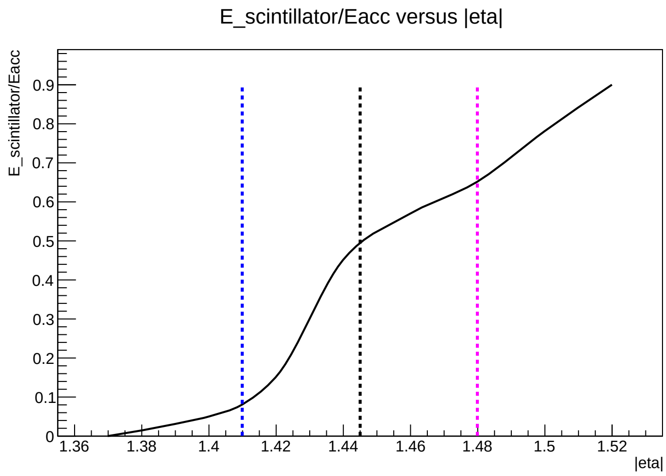


Figure 3: $E_{scintillator}/E_{acc}$ versus $|\eta|$. This figure displays the dependence of $E_{scintillator}/E_{acc}$ on $|\eta|$. As it seems, the graph is divided by dashed lines in four $|\eta|$ bins, according to its different behavior in these regions.

As it seems from the figure, the behavior of $E_{scintillator}/E_{acc}$ versus $|\eta|$ is different in the four $|\eta|$ regions divided by the dashed lines. This behavior leads us to run the regression in these four different $|\eta|$ bins (1.37-1.41,1.41-1.445,1.445-1.48,1.48-1.52).
The best results extracted from the performance of this process are the following:

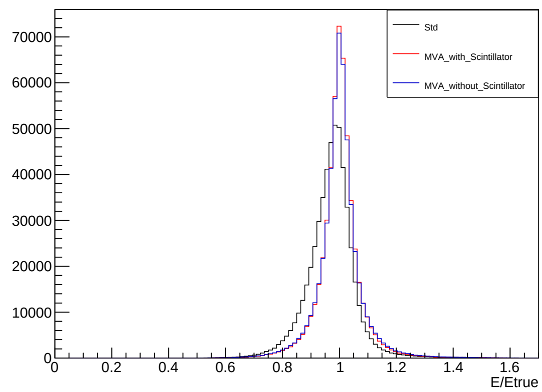


Figure 4: Distribution of E/Etrue comparing the standard calibration, Std (in black) and the MVA calibration with (in red) and without (in blue) the use of the scintillator.

Figure 2 depicts the histogram of E/E_{true} computed with the Std as well as with the MVA calibration. There are two MVA calibrations, one with scintillator as an input variable and another one without it. As it seems from the plot, the use of the MVA calibration, in both cases, results to narrower distributions of E/E_{true} which peak at unity in contrast to that computed with the Std which is more spread and shifted from 1. The comparison of the two MVA calibrations, results that the model with the scintillator produces a slightly narrower distribution than the model without scintillator.

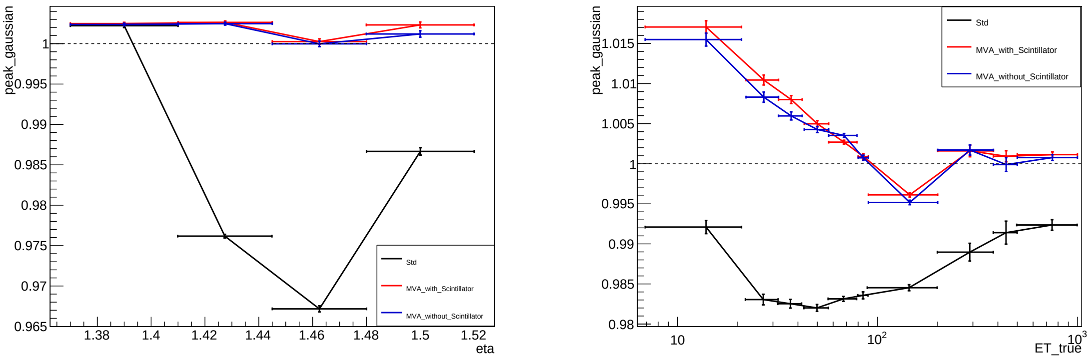


Figure 5: Gaussian peak of E/E_{true} versus $|\eta|$ (left) and versus E_T^{true} (right). In the first plot the peak position computed with both the MVA calibrations is close to 1 between 1.44 and 1.48, sustaining small deviations from it in the other regions. However the deviation of the peak position, derived with the Std is higher than the previous in all η regions. In the second plot the peak position calculated with the MVA calibrations is almost unity in higher E_T^{true} bins in contrast to that computed with the Std, while in lower it deviates from that because of lack of events.

In figure 3 the graph on the left shows the Gaussian peak of E/E_{true}, which is the peak position, as a function of $|\eta|$ while that one on the right shows the Gaussian peak of E/E_{true} as a function of E_T^{true}. As it seems from the figure, the peak position of E/E_{true} versus $|\eta|$ computed with both the MVA calibrations in low $|\eta|$ bins, between 1.37 and 1.41, displays a very small deviation from 1, about 0.2% and between 1.44 and 1.48 it almost reaches unity. This results that in these $|\eta|$ bins the output energy of the particles is almost the same with their true energy. In higher $|\eta|$ bins the Gaussian peak calculated without scintillator as input variable approaches more to unity than the other one, which is strange as it is the opposite of the expected and needs further study. On the other hand, the peak position computed with the Std deviates more from 1 everywhere, with its highest distance from it at $|\eta| = 1.46$, about 3.3%. Concerning the graph on the right, the peak position of E/E_{true} versus E_T^{true} is close to unity in the highest E_T^{true} bins, between 400 and 1000 GeV. As it seems from the figure, in lower E_T^{true} bins, between 0 and 12 GeV, its deviation from 1 is the highest, about 1.78% and 1.7% for the MVA calibration performed with scintillator as input variable and that one without scintillator, respectively, which are higher than the deviation of the Std from 1. This happens because the MVA calibration corrects for the mean but not for the peak and also because we did not apply the shifts. After 12 GeV it starts reducing until 100 GeV, where it is almost unity.

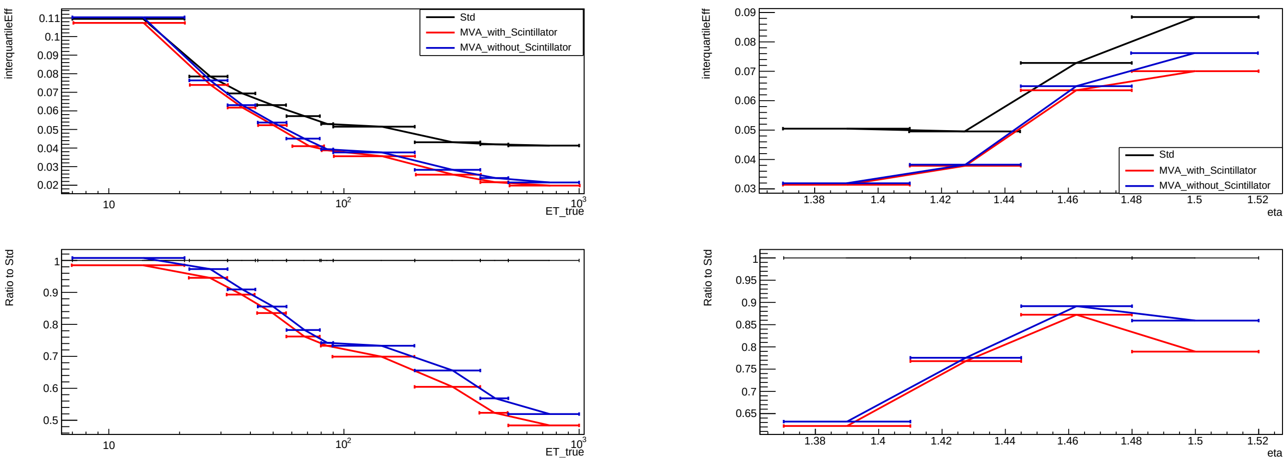


Figure 6: Resolution of E/E_{true} versus E_T^{true} (left) and versus $|\eta|$ (right) . In the first plot, versus E_T^{true} the resolution gets better as it moves from lower to higher E_T^{true} bins. As it seems from the second plot versus E_T^{true} the amelioration of the resolution is about 50% and 46% for that produced by the MVA with scintillator and for that calculated with the MVA without it, respectively. In the graph of interquartileEff versus $|\eta|$ the best resolution exists in lower $|\eta|$ bins and it is found to be $\sim 40\%$ for both MVA calibrations. In higher $|\eta|$ bins the scintillator improves the resolution about 13% more than the MVA without it.

Figure 4 shows the resolution of E/E_{true} versus $|\eta|$ (left) and versus E_T^{true} (right). The resolution is quantified by using the interquartile range. The interquartile range is a measure of statistical dispersion of a distribution, which equals to the subtraction of the lower quartiles by the upper quartiles. The lower and the upper quartiles refrain 25%, each of them, from the median of the distribution. As mentioned above, the calculation of the histogram of E/E_{true} took place in different $|\eta|$ and E_T^{true} bins. For each $|\eta|$ and E_T^{true} bins there is one histogram with three distributions, one computed with the Std calibration, one with the MVA calibration with scintillator and one with the MVA calibration without it, and for each of these histograms the calculation of the interquartile range is performed. Thereafter, the combined interquartile ranges of every distribution lead to the production of the interquartileEff plot versus $|\eta|$ as well as versus E_T^{true}. The interquartileEff is equal to the interquartile range/1.349, where 1.349, is the number with which the transition from the interquartile range to the standard deviation is implemented in a Gaussian distribution. Thus in these plots the interquartileEff represents the resolution of E/E_{true} versus E_T^{true} and $|\eta|$ [Upton, Graham; Cook, Ian (1996) and Zwillinger, D., Kokoska, S.(2000)]. In the first plot, versus E_T^{true} the resolution ranges between 2% and 12% while in the second one versus $|\eta|$ it ranges between 3.2% and 7.8%. This happens because the resolution versus $|\eta|$ depends on the E_T^{acc} distribution (Figure 2). As it seems from the figure of interquartileEff versus E_T^{true} the best resolution is achieved in higher E_T^{true} bins and it is about 2% for both the MVA calibrations. However, the resolution produced by the Std in higher E_T^{true} bins is 4.5%, which is worse than the resolution computed by the two MVA calibrations. Concerning, the second plot versus E_T^{true}, the division of the distributions of the first plot with that obtained with the Std calibration leads to the comparison of the MVA calibrations with that in order to find the improvement of the resolution . As is shown from the figure, the resolution gets better as it goes from lower to higher E_T^{true} bins, and in higher E_T^{true} bins the gain for the resolution computed with the MVA calibration with scintillator as input variable is 50% while for the resolution calculated with the other MVA calibration is almost 46%.
In the figure of intequatrileEff versus $|\eta|$, the best resolution is achieved in lower $|\eta|$ bins and is almost 3.2% for both the MVA calibrations while that one computed with the Std calibration is worse than the other two not only in these $|\eta|$ bins, where it is about 5%, but also in the whole $|\eta|$ range. As it seems from the second plot versus $|\eta|$, the optimization of the resolution derived with the MVA calibration with scintillator is 38% and that one of the resolution obtained from the other MVA calibration without the scintillator is 37%, in lower $|\eta|$ bins. In higher $|\eta|$ bins the contribution of the scintillator is quite important as the MVA calibration with scintillator improves the resolution $\sim$ 13% more than the MVA calibration without scintillator.

Conclusions

The main purpose of this project was to derive a correction for the energies of electrons in the transition region between the barrel and the end-caps of the ATLAS EMCal, using multivariate techniques. This was done for the first time and greatly improves the resolution compared to the existing methods Additionally, the comparison between the two multivariate techniques used in our analysis, one with the energy of the scintillator as input variable and one without it, shows that in higher $|\eta|$ region the scintillator improves the resolution $\sim$ 13% more than the MVA calibration without it.

References

[The Cern Large Hadron Collider:Accelerator and Experiments vol.1] The Cern Large Hadron Collider:Accelerator and Experiments vol.1, Cern Geneva 2009

[PHD Thesis, Bruno Lenzi 2010] Search for the Higgs boson decaying to four leptons in the ATLAS detector at LHC and studies of muon isolation and energy loss, Bruno Lenzi, 2010

[Lenzi B. and Turra R.,11 Oct 2013] Monte Carlo calibration update for electrons and photons using multivariate techniques, Lenzi B.,Turra R., 11 Oct 2013

[A. Hoecker et all, TMVA 4] arXiv:physics/0703039 [Data Analysis, Statistics and Probability] TMVA 4,Toolkit for Multivariate Data Analysis with ROOT, A.Hoecker et all

[Upton, Graham; Cook, Ian (1996) and Zwillinger, D., Kokoska, S.(2000)] Understanding Statistics, Oxford University Press. p. 55 and CRC Standard Probability and Statistics Tables and Formulae, CRC Press