

# Derivation of Sigmoid Function from Bayes Inference

Andriy Zatserklyaniy, zatserkl@fnal.gov

March 8, 2018

## 1 Bayes' Theorem

Let us assume that there are only two classes(hypotheses) of events:  $C_1$  and  $C_2$ . Event  $x$  has been occurred. What the probability that  $x$  belongs to the class  $C_1$ ?

Conditional probabilities in general:

$$P(x|C)P(C) = P(x \cdot C) = P(C|x)P(x)$$

For our two hypotheses

$$\begin{aligned} P(C_1|x)P(x) &= P(x|C_1)P(C_1) \\ P(C_2|x)P(x) &= P(x|C_2)P(C_2) \\ \frac{P(C_1|x)}{P(C_2|x)} &= \frac{P(x|C_1) P(C_1)}{P(x|C_2) P(C_2)} \\ P(C_1|x) &= P(C_2|x) \cdot \frac{P(x|C_1) P(C_1)}{P(x|C_2) P(C_2)} \\ &= \frac{\cancel{P(x|C_2)}\cancel{P(C_2)}}{P(x)} \cdot \frac{P(x|C_1) P(C_1)}{\cancel{P(x|C_2)}\cancel{P(C_2)}} \\ &= \frac{P(x|C_1)P(C_1)}{P(x)} \\ &= \frac{P(x|C_1)P(C_1)}{P(x|C_1)P(C_1) + P(x|C_2)P(C_2)} \end{aligned}$$

rewrite as

$$P(C_1|x) = \frac{1}{1 + \frac{P(x|C_2) P(C_2)}{P(x|C_1) P(C_1)}}$$

denote the second term in the denominator as  $e^{-t}$  where (NB: reciprocal because of minus sign)

$$t = \log \frac{P(x|C_1) P(C_1)}{P(x|C_2) P(C_2)}$$

and

$$P(C_1|x) = \frac{1}{1 + e^{-t}}$$

Note that  $P(C_1|x)$  is a value between 0 and 1.

## 2 Discussion

The term  $P(C_1|x)$  is a *posterior* probability that improves the value of the *a priori* probability  $C_1$  after the event  $x$ .

I guess that after the some number of events the posterior probability will have almost the same value independent of the initial value  $P(C_1)$ .

My understanding is that the hidden layers provide iterations to the final value of posterior probability of hypothesis  $C_1$ . Hence, more hidden layers – more iterations.

### 3 LaTeX template cont'd

Let's consider histogram with bin width of 1/2 units (MeV on the figure), Fig.1.

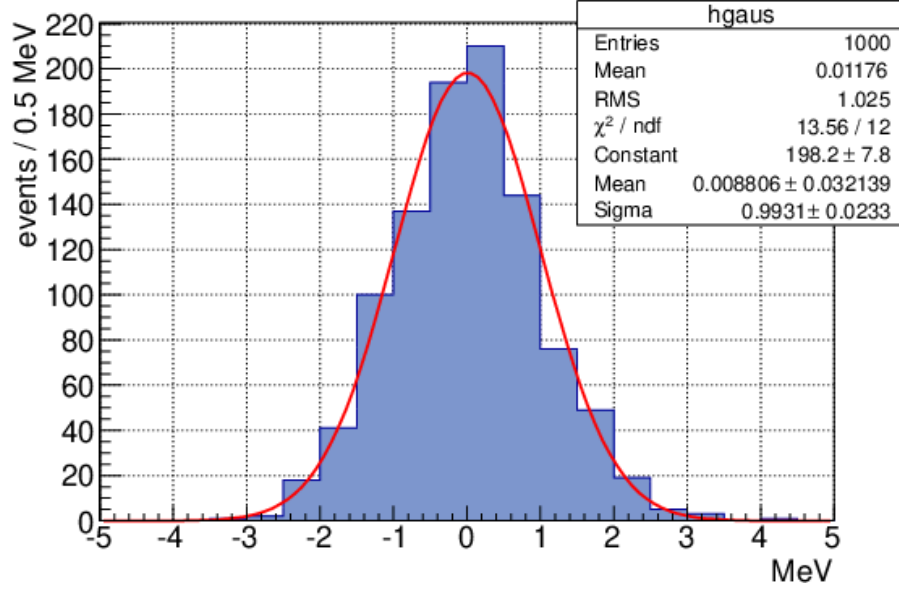


Figure 1: Histogram and Gaussian fit

Calculate the number of events in the histogram. The number of events is proportional to the histogram area

$$S = N \cdot s_1$$

where  $s_1$  is area which corresponds to one event. The area of every histogram bin is a product of the bin width  $w$  and the number of events in the bin. Therefore, the area which corresponds to one event is a product of the bin width  $w$  and 1:

$$s_1 = w \cdot 1 = w$$

To calculate an area of the Gaussian-shape histogram in the Fig.1 let's approximate it by the area of the Gaussian

$$g(x) = Ae^{-\frac{(x-x_0)^2}{2\sigma^2}}$$

Fit results are:

$$A = 198.2$$

$$\sigma = 0.993$$

Calculate the area under the Gaussian

$$\begin{aligned}
 S &= \int_{-\infty}^{+\infty} A e^{-\frac{(x-x_0)^2}{2\sigma^2}} \\
 &= \sqrt{2\pi}\sigma A \underbrace{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-x_0)^2}{2\sigma^2}}}_{=1} \\
 &= \sqrt{2\pi}\sigma A \\
 &\approx 2.5 \cdot \sigma \cdot A
 \end{aligned}$$

Express this area in terms of the number of events  $N$  and the area of the one event  $s_1$ :

$$\begin{aligned}
 S &= N \cdot s_1 \\
 &= N \cdot w
 \end{aligned}$$

hence

$$N = S/w$$

or

$$N = \sqrt{2\pi} \cdot \sigma \cdot A/w$$

Because  $\sqrt{2\pi} \approx 2.5066$

$$N \approx 2.5 \cdot \sigma \cdot A/w$$

The histogram on the Fig.1 was generated for 1000 events Gaussian-distributed with  $\sigma = 1$ . In our approximation

the area under the Gaussian

$$S \approx 2.5 \cdot 198.2 \cdot 0.993 = 492.0$$

the number of events for  $w = 0.5$  MeV

$$N \approx 2.5 \cdot 198.2 \cdot 0.993/0.5 = 984.1$$