

# Chapter 1

## Statistical Derivation of the AMS

For a classifier  $h$  with a selection region  $\mathcal{H} = \{\mathbf{x} : h(\mathbf{x}) = s\}$ , the number of events  $n = |\mathcal{H}|$  in the selection region of a classifier follows a Poisson distribution as its mean can be expressed by the sum of two Poisson random variables  $\mu_b + \mu_s$ , where  $\mu_b$  is the expected number of background events in the selection region and  $\mu_s$  is the expected number of signal events in the selection region.

A Poisson probability for  $n$  events with an average rate of occurrence of  $\mu_b + \mu_s$  is given by,

$$P(n|\mu_b, \mu_s) = \frac{(\mu_b + \mu_s)^n}{n!} e^{-(\mu_b + \mu_s)} \quad (1.1)$$

The null hypothesis we are testing for is  $H_0 : \mu_s = 0$  and the alternative hypothesis is  $H_1 : \mu_s > 0$ . We compute the likelihood ratio defined as,

$$\Lambda = \frac{P(n|0, \mu_b)}{P(n|\mu_s, \mu_b)} = \left(\frac{\mu_b}{\mu_s + \mu_b}\right)^n e^{\mu_s} = \left(\frac{\mu_b}{n}\right)^n e^{n - \mu_b} \quad (1.2)$$

where  $\mu_s = n - \mu_b$  is the maximum likelihood estimator for  $\mu_s$ . Next, we need the sampling distribution of the test statistic  $\Lambda$  to derive a  $p$ -value.

By Wilk's theorem the test statistic  $-2\ln(\Lambda)$  follows a  $\chi^2$  distribution in the large sample limit as  $n \rightarrow \infty$ . This means that we can compute likelihood  $\Lambda$  and use the statistic  $-2\ln(\Lambda)$  under a  $\chi^2$  cumulative distribution function to compute the  $p$ -value.

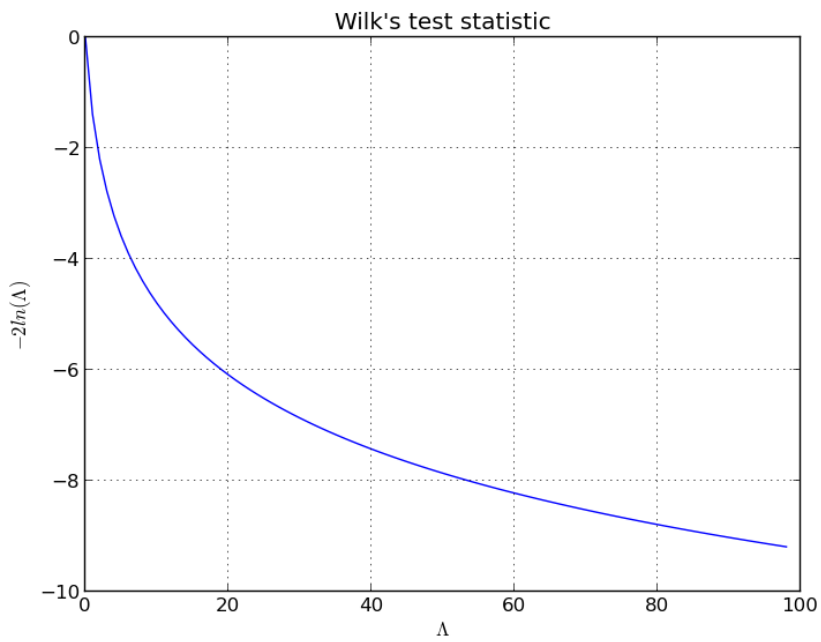
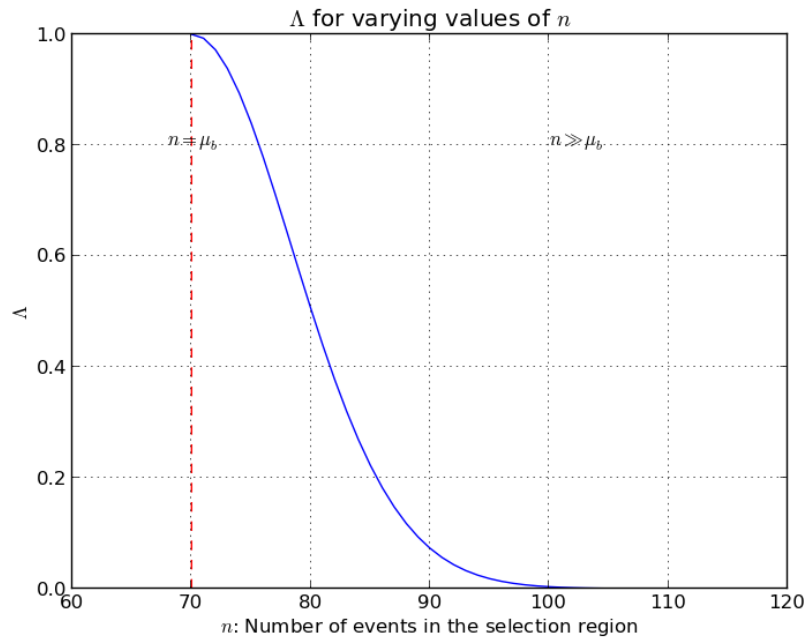
$$p = 1 - Q(-2\ln(\Lambda)) \quad (1.3)$$

$Q$  is the  $\chi^2$  cumulative distribution function.

The test statistic,

$$q_0 = \begin{cases} -2\ln(\Lambda), & n > \mu_b \\ 0, & \text{otherwise} \end{cases} \quad (1.4)$$

serves as the basis of the statistical test. In order to see why, a significant upward fluctuation in the data constitutes evidence that a background only hypothesis ( $\mu_s = 0$ ) is not supported in the data. Thus, higher values of the test statistic -  $q_0$  point to increasing evidence of incompatibility with the null hypothesis. If the data fluctuates downward such that one finds events fewer than that expected by the background process then this also points to incompatibility with the null but not because of presence of signal events but rather some other systematic error hence in this case we assign  $q_0 = 0$ .



Next, we need to define the pdf (probability density function) of  $q_0$ . Note that  $q_0$  when  $n > \mu_b$  follows a  $\chi^2$  distribution in the large sample limit. The pdf of a  $\chi^2$  distribution is,

$$\frac{1}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} \quad (1.5)$$

where  $k$  is the degrees of freedom (since a  $\chi^2$  random variable is the sum of  $k$  independent standard normal variables. If  $Z_1, \dots, Z_k$  are standard normal then,  $Q = \sum_{i=1}^k Z_i^2$  is  $\chi^2$  distributed with  $k$  degrees of freedom) and  $\Gamma$  is the gamma function,  $\Gamma(n) = (n-1)!$

Using the pdf of the  $\chi^2$  distribution and the fact that  $k = 1$  for  $q_0$  and  $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$  we deduce that the pdf of  $q_0$  is,

$$f(q_0) = \frac{1}{2} \delta(q_0) + \frac{1}{2} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{q_0}} e^{-\frac{q_0}{2}} \quad (1.6)$$

where  $\delta$  is the Dirac delta function and we use the result that when a random variable takes  $n$  discrete values  $x_1, \dots, x_n$  with associated probabilities  $p_1, \dots, p_n$  the probability density function is,

$$f(t) = \sum_{i=1}^n p_i \delta(t - x_i) \quad (1.7)$$

The cumulative distribution function of  $q_0$  is,

$$F(q_0) = \Phi(\sqrt{q_0}) \quad (1.8)$$

where  $\Phi$  is the cumulative distribution function (cdf) of the standard normal variable.

To derive this from first principles consider the expression of the cdf of the random variable  $q_0$  as defined above,

$$\begin{aligned} F_{q_0}(x) &= \int_{-\infty}^x f_{q_0}(t) dt \\ &= \frac{1}{2} \int_{-\infty}^x \delta(t) dt + \frac{1}{2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \frac{1}{\sqrt{t}} e^{-t/2} dt \\ &= \frac{1}{2} \left[ 1 + \operatorname{erf}\left(\sqrt{\frac{x}{2}}\right) \right] \\ &= \Phi(\sqrt{x}) \end{aligned}$$

where  $\operatorname{erf}(\bullet)$  is the gaussian error function defined as,

$$\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^x e^{-t^2} dt$$

This is closely related to the standard normal cumulative distribution function through,

$$\Phi(x) = \frac{1}{2} \left[ 1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \right]$$

and

$$\Phi(x) = \int_{-\infty}^x e^{-t^2/2} dt$$

The  $p$ -value and significance is from eq. 1.8 is therefore,

$$p = 1 - \Phi(\sqrt{q_0})$$

$$Z = \Phi^{-1}(1 - p)$$

This leads to the result,

$$Z = \sqrt{q_0} = \sqrt{2 \left( n \ln \left( \frac{n}{\mu_b} \right) - n + \mu_b \right)} \quad (1.9)$$

if  $n > \mu_b$  and  $Z = 0$  otherwise. The quantity  $Z$  measures significance in terms of units of sigmas from the mean, a  $Z$  value of 5 would correspond to a  $5\sigma$  effect and in physics is regarded as necessary to claim a discovery of a new particle.