

# Calculation of total length of follow up when comparing two Poisson processes

TOL := 0.000000001      Set tolerance for calculation. This is a system parameter

## Introduction

Suppose that we observe two independent Poisson processes over time and suppose that we observe each for the same length of time. How long does the observation time have to be to establish that there is a difference between processes?

## Meaning of parameters

$r_1$  the rate per unit time in the first process

$r_2$  the rate per unit time in second process

Obs<sub>time</sub> the length of follow up

$\lambda_1$  the expected number of events in the first Poisson process

$\lambda_2$  the expected number of events in the second Poisson process

$\alpha$  the type one error rate

$\beta$  the type two error rate

$k = 1, 2$  the number of tails for the test

$\Delta$  the standardised non-centrality parameter

## Definition of functions for parameters

$\lambda(r, \text{Obs}_{\text{time}}) := r \cdot \text{Obs}_{\text{time}}$       general function defining expected number of events as a function of rate and time

$\varepsilon := \frac{1}{2}$       Continuity correction

$\text{Signal}(r_1, r_2, \text{Obs}_{\text{time}}) := \lambda(r_1, \text{Obs}_{\text{time}}) - \lambda(r_2, \text{Obs}_{\text{time}}) - \varepsilon$       The expected difference in number of events adjusted by the continuity correction

$V_{\text{diff}}(r_1, r_2, \text{Obs}_{\text{time}}) := \lambda(r_1, \text{Obs}_{\text{time}}) + \lambda(r_2, \text{Obs}_{\text{time}})$       The variance of the difference in the number of events using the fact that a Poisson has variance equal to mean and that the two processes are independent

$Z(r_1, r_2, \text{Obs}_{\text{time}}) := \frac{\text{Signal}(r_1, r_2, \text{Obs}_{\text{time}})}{\sqrt{V_{\text{diff}}(r_1, r_2, \text{Obs}_{\text{time}})}}$       The standardised difference between the observed number of events for the two processes

$$\Delta(\alpha, \beta, k) := \text{qnorm}\left(1 - \frac{\alpha}{k}, 0, 1\right) + \text{qnorm}(1 - \beta, 0, 1)$$

Non-centrality parameter  
as a function of type I and type II  
errors

NB *qnorm* is a built in Mathcad function giving the quantiles of the Normal distribution.

## An example

### The problem posed

It is desired to examine whether it is true that the risk of injury per hour in team sports is greater than in solo sports. The rate parameters below are supposed to apply

$$r_1 := \frac{0.00175}{\text{hr}} \quad \text{rate per hour team sports} \qquad r_2 := \frac{0.00093}{\text{hr}} \quad \text{rate per hour solo sports}$$

A two tailed test at the 5% level will be carried out and 80% power is targetted. What is the length of follow up?

Set parameters for hypothesis test

$$\alpha := 0.05 \qquad \beta := 0.2 \qquad k := 2$$

### Solution

The approach to finding a solution uses a Normal approximation. The required standardised non-centrality parameter is calculated as a function of  $\alpha, \beta$  and  $k$ . The actual non-centrality parameter is calculated using the rates and the follow-up. Subtracting one from the other defines an equation whose right hand side is 0. The root of this equation is the desired follow up.

$$\text{Obs}_{\text{time}} := 1000 \cdot \text{hr} \quad \text{Guess value for solution}$$

$$\Delta(\alpha, \beta, k) = 2.802 \quad \text{Required non-centrality given type I and type II errors}$$

$$\text{Follow}(r_1, r_2, \alpha, \beta, k) := \text{root}(Z(r_1, r_2, \text{Obs}_{\text{time}}) - \Delta(\alpha, \beta, k), \text{Obs}_{\text{time}})$$

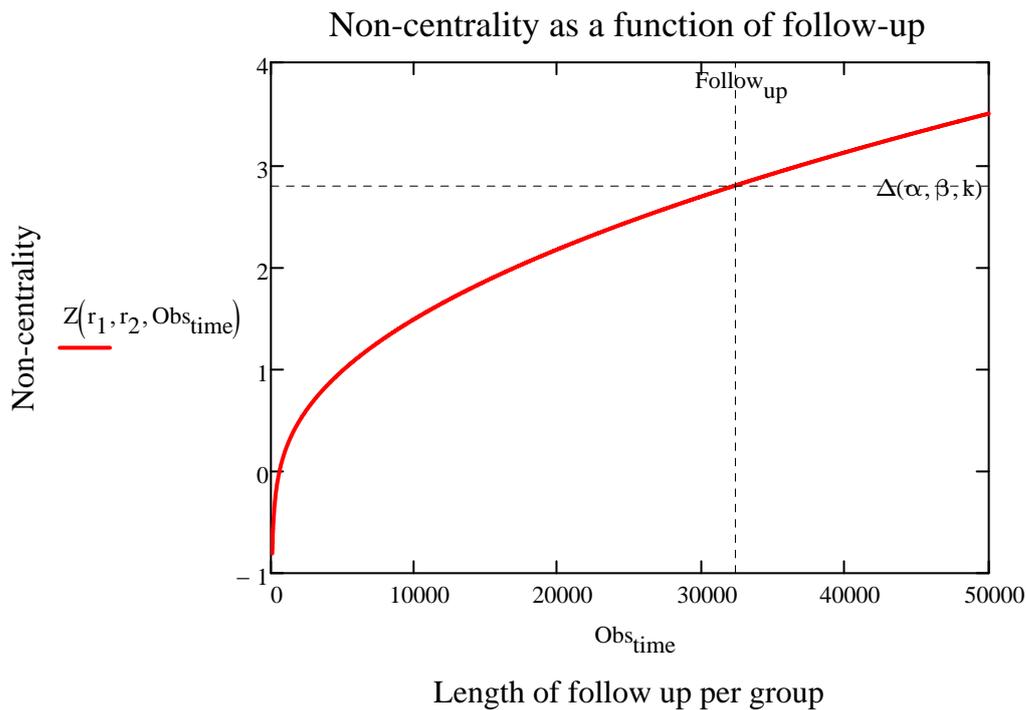
Function for root of the equation

$$\text{Follow}_{\text{up}} := \text{Follow}(r_1, r_2, \alpha, \beta, k) \quad \text{Obtain solution}$$

$$\text{Follow}_{\text{up}} = 32492 \cdot \text{hr} \quad \text{Display solution}$$

### Illustrate solution

Obs<sub>time</sub> := 100.hr, 110hr .. 50000·hr



### Calculate expected number of events in each group

$$\lambda_1 := \lambda(r_1, \text{Follow}_{\text{up}})$$

$$\lambda_2 := \lambda(r_2, \text{Follow}_{\text{up}})$$

$$\lambda(r_1, \text{Follow}_{\text{up}}) = 56.86$$

$$\lambda(r_2, \text{Follow}_{\text{up}}) = 30.217$$

$$\lambda_1 = 56.86$$

$$\lambda_2 = 30.217$$

### Check solution using alternative test

We use the fact that conditional on a given total number of events over both groups the number in a given group will have a binomial distribution. If the null hypothesis is true the probability that a given event should be in one group rather than the other will be 0.5.

## Define binomial P-value

$$Pval(x_1, x_2) := pbinom(x_2, x_1 + x_2, 0.5)$$

We use *pbinom* here, the built in Mathcad function for a cumulative binomial probability to calculate the probability that there are  $x_2$  or fewer events in group 2 given the total in both groups. This thus becomes a P-value.

$$Sig(x_1, x_2) := pbinom(x_2, x_1 + x_2, 0.5) \leq \frac{\alpha}{k} \quad \text{This defines an indicator function that takes on the value 1 if the P-value is less than } \alpha/k$$

## Define power function

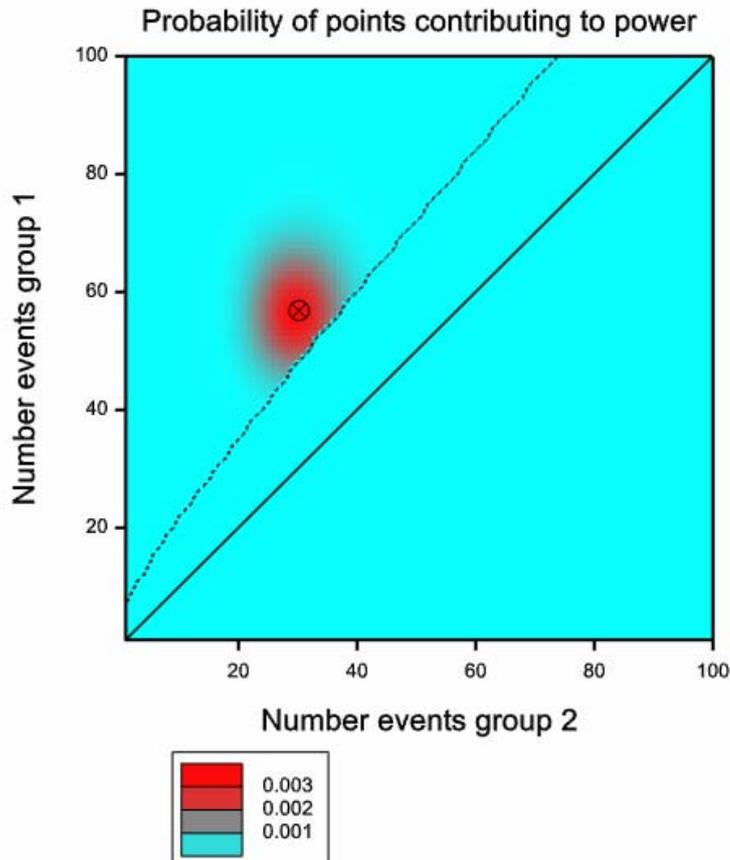
Now define a power function. We assume, as is usual that we are only interested in significance in one direction. We let  $x_1$  take on all possible values that are greater than  $x_2$ , since we will never conclude significance unless  $x_2$  is less than  $x_1$ . Thus  $x_1$  must be at least 1. Then we let  $x_2$  range over all possible values less than  $x_1$ . For each allowable pair of points we calculate the probability of the point occurring. Finally we sum the probability of all such points multiplying by the indicator function. In this way we only count points for which the result is significant. In practice we must choose some upper limit for the Poisson. Here 400 has been chosen although the probability of this number of events being seen in practice is minute.

Note that *dpois* is a built-in Mathcad function giving probabilities for the Poisson distribution

$$\text{Power} := \sum_{x_1=1}^{400} \left[ dpois(x_1, \lambda_1) \cdot \sum_{x_2=0}^{x_1-1} (dpois(x_2, \lambda_2) \cdot Sig(x_1, x_2)) \right]$$

Power = 0.7926      Compare this to       $1 - \beta = 0.8$

## Plot power calculation



GensStat(R) graph illustrating power calculation. The contours of the joint probability are illustrated. However only those points above and to the left of the jagged boundary would be counted as significant so the points below and to the right have had their probability set to zero. Also shown is the line of exactly equality and the point corresponding to the mean number of events in group 1 and group 2. The power is the sum of the probabilities of the points above and to the left of the dashed line.

## Discussion

It should be noted that for most practical problems the solution given is optimistic. This is because there are very few pure Poisson processes (PPP) and in practice data usually display extra Poisson variation.

In the example considered, if a PPP applied it would follow that studying 10 athletes for 1000 hours each would give the same information as studying 1000 athletes for 10 hours each. This is very unlikely to be the case and would only apply if they were all equal risk. In practice there would probably be a personal element of risk that would vary from athlete to athlete even for those athletes following the same kind of sport.

A popular model under such circumstances is a gamma-Poisson mixture. That is to say each athlete's personal risk of injury would be regarded as being drawn at random from a gamma distribution whose parameters depended on whether he was participating in group or solo sports. Conditional on the personal risk the actual probability of injury would then be assumed to follow a negative binomial. This leads to a negative binomial. For the negative binomial the variance is greater than the mean unlike for the Poisson for which it is equal to the mean.

In practice a data-set consisting of two groups of persons being compared with different follow-up per individual could be analysed as a generalised linear model with a log link for a negative binomial with an offset for length of follow-up. However, there *might* be an issue if the event is censoring as might be the case with our example, where injury might lead to sporting activity being suspended.