

# Extending Poisson Distribution to trending datasets

**Source:** <http://sci.tech-archive.net/Archive/sci.stat.math/2004-10/0004.html>

---

**From:** Fred Chen (*flipsu5\_at\_comcast.net*)

**Date:** 09/30/04

Date: 30 Sep 2004 00:30:36 -0700

Richard Ulrich <Rich.Ulrich@comcast.net> wrote in message  
news:<e3nml0907bevsqh5rm5c740kivo8g64qbc@4ax.com>...

> On 29 Sep 2004 10:58:51 -0700, *flipsu5@comcast.net* (Fred Chen) wrote:

>

>> *What care must I take to apply Poisson statistics to data that  
>> normally exhibits random behavior from one data point to another, but  
>> could reflect a systematic behavior in the big picture, i.e., exhibit  
>> a long-term trend?*

>

> *Do you intend to say here that there is autocorrelation  
> across time? Or does 'systematic behavior' only describe  
> the fact that there is a trend across time?*

>

> *Finding a long-term trend, in the way I think of it, is  
> what Poisson statistics are used for; when the criterion  
> is assumed to have Poisson error.*

>

>

>>

>> *In case the Poisson assumption breaks down, what statistics can be  
>> used for the analysis?*

>

> *The main 'breakdown' of the Poisson assumption, for data  
> of the sort that I'm accustomed to, happens when the events  
> are correlated, either at one time (events occurring in batches),  
> or across time (autocorrelation, or, say, accidentally looking at  
> prevalence instead of incidence). Then you usually have to  
> frame your question rather differently. - OLS regression will  
> be similarly distorted. Effectively, the question has fewer degrees  
> of freedom than the N.*

>

> *These two 'breakdowns' can show up as overdispersion  
> or underdispersion of the variances. There are statistical  
> cures for some dispersion errors. That is, you might simply  
> fit another distribution, when the under- or over-dispersion  
> arises because the distribution had too many zeroes, or was*

- > *log-normal instead of Poisson.*
- >
- > *It all depends on what you have in mind for how \*your\**
- > *Poisson assumption could 'break down' – which might be*
- > *entirely different from my cases.*

Rich, thanks for your reply. The scenario I am thinking of is the data shows a trend which is explainable at this point as due to a phenomenon that is affected by the accumulated number of flagged items. However, each data point still has its inherent randomness, so that for a given time interval, the number of flagged items detected is still rare and can occur at different instants in that interval.

Does Poisson analysis breakdown here? If so, how can I calculate the standard deviation of the number of flagged items as a whole?

Fred

P.S. I am thinking this may be an inhomogeneous Poisson process since the occurrence of flagged items in one interval do not directly affect those in another.