

Elementary Statistics on Trial (the case of Lucia de B.)

Richard D. Gill* and Piet Groeneboom †

January 31, 2009

Abstract

It has often been noticed that, in observing the number of incidents that nurses experience during their shifts, there is a large variation between nurses, and over time. We propose a simple statistical model to describe this phenomenon and apply this to the Lucia de Berk case.

Still to be done:

Add the three 2x2 tables (before, after data-cleaning);

Add some summary statistics (hypergeometric, binomial);

Condense Appendix;

Delocalize.

1 Introduction

We model the incidents that a nurse experiences as a homogeneous Poisson process on the positive halfline, with a nurse-dependent intensity λ . As is well-known, a Poisson process is used to model incoming phone calls during non-busy hours, fires in a big city, etc. Since we believe incidents to be rare, a Poisson process is also a natural choice for modeling the incidents that a nurse experiences.

Our model is parametric, and we take as the distribution of the intensity λ over nurses the Gamma($\rho, \rho/\mu$) distribution. Using this model, our sample consists of realizations of the random variable

$$(L, T, N),$$

*Mathematical Institute, Leiden University; <http://www.math.leidenuniv.nl/~gill>

†DIAM, Delft University; <http://ssor.twi.tudelft.nl/~pietg>

where L has a Gamma($\rho, \rho/\mu$) distribution, and N , conditionally on $L = \lambda$ and $T = t$, has a Poisson distribution with parameter λt . The random variable T represents the time interval in which incidents take place (for a particular nurse). So, if there are n nurses, we deal with a sample

$$(L_1, T_1, N_1), \dots, (L_n, T_n, N_n),$$

of independent random variables, all having the same distribution as (L, T, N) . The random variable N_i represents the number of incidents nurse i experiences in the time interval T_i . As a somewhat arbitrary choice, we will take $\rho = 1$. This means, among other things, that it can easily happen that one nurse has twice the incident rate of another nurse.

The statistical problem boils down to the estimation of the parameter μ .

We use Derksen and de Noo's revised data set, taking account of incidents among the other nurses (which formerly were taken to be, by definition, *not suspicious*), and removing incidents and deaths for which Lucia was deemed innocent (not charged with murder or attempted murder, presumably because these events were medically speaking "expected to happen, when they actually did").

We will also give results for the data set obtained when we do not treat Lucia so generously: a number of incidents removed by order of the courts (because Lucia was not charged with murder even though she was present) and those removed by Derksen (because Lucia was not found guilty by the court) are put back. We will see that though this has a big effect on the number of incidents in her shifts, its effect on our final conclusion is small.

2 The numbers

Combining the Juliana Children's Hospital and the two wards of the Red Cross Hospital, Lucia had 201 shifts, 7 incidents. It is not clear whether this combination works out pro or contra Lucia (this depends on whether she did proportionately more or less shifts at the different wards, and whether the overall mean incident rate is larger or smaller at each ward).

We do need to do a combined analysis; the alternative to quick and dirty "just add everything up" is a much more complicated analysis and certainly requires making more, unsupportable, assumptions. We'll take the overall probability of an incident per shift to be the ratio of total incidents to total shifts, $\mu = 23/1734$. If we take a shift to be our unit time interval, then this would be a moment estimate of the mean intensity of incidents EL . This means, that, conditionally on $T = 201$, the number of incidents for Lucia follows a mixture of Poisson random variables with parameter $201L$, where

the intensity L has a $\Gamma(1, 1/\mu)$ distribution, which is in fact the exponential distribution on $[0, \infty)$ with first moment μ . Thus on average, an innocent Lucia would experience $201 \cdot \mu = 201 \cdot 23/1734 \approx 2.66609$ incidents.

The probability of having 7 or more incidents is given by:

$$\frac{1}{\mu} \int_0^\infty \mathbb{P}\{N \geq 7 | L = \ell, T = 201\} e^{-\ell/\mu} d\ell$$

It is well-known that, for a random variable N , which is distributed according to a Poisson(λ) distribution, we have:

$$\mathbb{P}\{N \geq n\} = \frac{1}{(n-1)!} \int_0^\lambda e^{-x} x^{n-1} dx.$$

So we find:

$$\begin{aligned} & \frac{1}{\mu} \int_0^\infty \mathbb{P}\{N \geq 7 | L = \ell, T = 103\} e^{-\ell/\mu} d\ell \\ &= \frac{1}{6! \mu} \int_0^\infty \left\{ \int_0^{201\ell} e^{-y} y^6 dy \right\} e^{-\ell/\mu} d\ell \\ &= \frac{1}{6!} \int_0^\infty e^{-\{1+1/(201\mu)\}y} y^6 dy \approx 0.1096182. \end{aligned}$$

This is just a bit smaller than *one in nine*.

A picture of the probabilities $\mathbb{P}\{N \geq k | T = 203\}$, $k = 1, 2, \dots$ is shown in Figure 1.

A more unfavourable conclusion might be obtained for Lucia, if we do not follow first the court and then Derksen in removing incidents from the statistics when they occur in her shifts, but she is not charged, or found guilty of causing them. But there is another correction to be made, if the unbiasedness of the data goes before everything else: Lucia was charged with murder or attempted murder on two occasions when the death occurred in the shift after hers, while nothing special was reported during her shift. These two ‘‘incidents’’ were counted as incidents in Lucia’s shifts in the statistical analyses done so far. Finally, yet another reanimation outside of Lucia’s shifts has been found in the hospital records.

All these corrections result in a total of 14 incidents in Lucia’s shifts, instead of 7, out of a grand total of 32 incidents in 1734 shifts. Taking again $\rho = 1$ we find the probability 0.036429 or *one in twenty-five*.

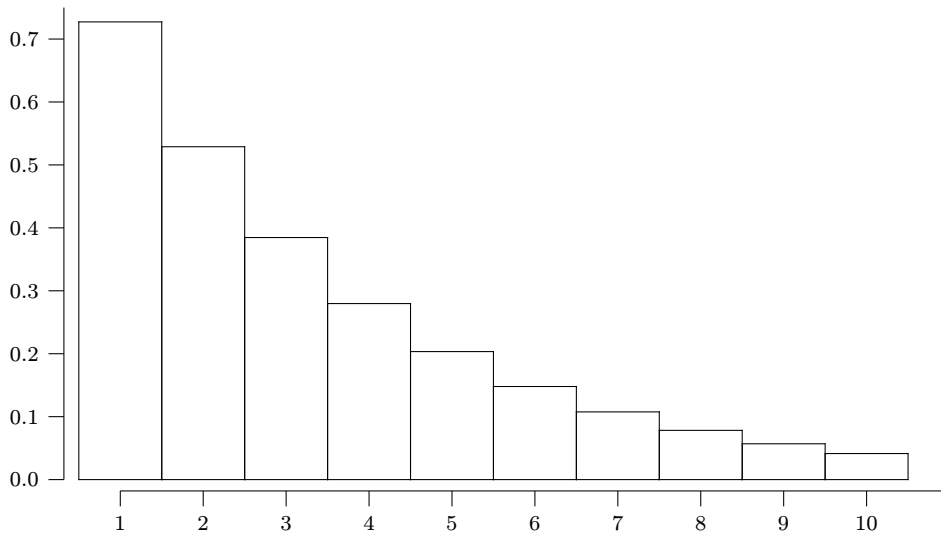


Figure 1: Probabilities (in the model) that the number of incidents in 203 shifts for one nurse is at least 1,2,3,..., if $\mu = 23/1681$. The probabilities are given by the heights of the columns above 1, 2, 3, ..., respectively.

3 Conclusion: One in twenty-five will go to jail

A modest amount of variation makes the chance that an innocent nurse experiences at least as many incidents as the number Lucia actually did experience, the somewhat unremarkable *one in nine*. Making some less favourable choices to her in the data cleaning process, only decreases this chance to *one in twenty-five*.

The fact that a modest amount of heterogeneity turns an almost impossible occurrence into something merely mildly unusual, is strong evidence that it does truly exist: the intuition of medical specialists needs to be challenged and the experience of nurses and nursing specialists is supported.

4 Appendix: extended discussion of heterogeneity

4.1 Preliminary remarks

The null-hypothesis tested by the court statistician H. Elffers in the case of Lucia de Berk is supposed to mean that incidents on a ward, and shifts of a particular nurse, are independent of one another. In the minds of lawyers or medical specialists, as in that of the man in the street, independence means: lack of causality. Causality can be “measured” by performing the thought experiment: suppose that Lucia worked on a particular shift, and that an incident occurred: would the same incident have happened if Lucia had been magically exchanged for another nurse? The idea in this thought experiment is that “everything else that might be relevant is kept the same”, so that we compare strictly comparable situations: with and without Lucia, everything else being unaltered. In a randomized double-blind clinical trial we do keep everything relevant the same, by randomization. In observational studies we are unable to do this, so instead we are forced to take explicit account of anything which could be relevant, in one way or another, if we want to conclude causality. This requires prior knowledge concerning the mechanisms underlying the phenomenon under study.

4.2 Are nurses interchangeable?

According to many medical specialists we have spoken to, nurses are indeed completely interchangeable with respect to the occurrence of medical emergencies among their patients: nurses merely carry out the instructions given to them by the medical staff, and they do this according to standard practices of proper care, so it can make no difference at all to replace one nurse by another. However according to nursing staff we have consulted, this is not the case at all. Different nurses have different styles and different personalities, and this can and does have a medical impact on the state of their patients. Especially regarding care of the dying, it is folk knowledge that terminally ill persons tend to die preferentially on the shifts of those nurses with whom they feel more comfortable. (This might apply to the Red Cross Hospital, where Lucia worked on two adjacent wards for terminally ill aged patients). As far as we know there has been no statistical research on this phenomenon.

4.3 Definition of incidents

There is another respect in which nurses can have an impact on “incidents”. In the Lucia case, incidents were never formally defined. However, if medical doctors were expressly called to the bed of the patient by nursing staff, then that soon qualified as an incident, especially if Lucia was somehow involved. Who decides if

the doctors should be alerted? The nurses on duty, themselves, of course. It seems that several of Lucia's incidents were created by herself in situations where she was uneasy about the patient who appeared to be developing some new and, to her, alarming symptoms. The nurse who keeps a closer eye on her patients, and who is less prepared to take risks, will generate in this way incidents on her shift, which otherwise might be postponed to the next shift or even fail to materialise at all. According to medical specialists, nurses do not have a choice in such situations: they have been trained to make the right decision and every nurse in the same situation will make the same decision. According to nurses however, this is just not true. Nurses do have to make their own decisions and though they should always be able to justify their choices, this does not mean that every individual will make the same choice in the same circumstances.

4.4 Inadequacy of the hypergeometric distribution as a model and spurious correlations

Above we have mentioned two ways in which a particular nurse could have a causal but "innocent" influence on the occurrence of an incident (since if she is replaced by another nurse it happens later or not at all). The underlying cause is unmeasured, and indeed perhaps unmeasurable, heterogeneity between nurses.

Next we discuss sources of correlation which correspond to indirect rather than direct causation: we speak then of spurious-correlation, correlation which can be explained by confounding factors, by common causes. In order to do this let us reconsider the two possible motivations for the hypergeometric distribution which so far has been used implicitly or explicitly by almost all researchers in the field.

One can derive the null-hypothesis assumption in two ways: either by taking the incidents on a given ward as fixed, and the shifts of a nurse as random, or by taking the shifts to be fixed, while the incidents are considered to be randomly occurring events. Let us consider in turn keeping one of the processes fixed, and seeing the other as the only source of randomness. For ease of exposition consider a single calendar year on a given ward of a given hospital – about 1100 shifts, of which a full-time permanently employed nurse might work about 150.

In the first picture – fixed incidents, random shifts – the picture to have in mind is that during the year, patients come and go, incidents occur at times dictated by the patients' day to day medical histories during the year. There may be patterns or regularities in this process but we simply take the actually occurring incidents as given at the times they did. The idea is that the incidents would have happened anyway, exactly when they did, independently of which nurses were on duty. Now we have to allocate shifts to nurses. The use of the hypergeometric distribution corresponds to determining the 150 shifts of our given nurse on the first of January by randomly selecting 150 lottery tickets from a large box of 1100 tickets, one for each shift of the year.

In actual fact, shifts are allocated in a dynamic process during the year, and

they follow rather evident patterns. That a particular nurse could equally likely have been assigned any 150 of the 1100 shifts in a year, is certainly not a *realistic* way to model the idea that which nurse works which shifts is somehow arbitrary. Whether it is an *adequate* way to model this situation depends on our purposes. The adequacy should be investigated, not taken for granted.

In the second picture – random incidents, fixed shifts – we think of the shifts of a given nurse as being fixed in advance. Incidents on the ward occur according to a completely random process, independent of shifts. This is the picture employed by the court’s statistician Elffers who speaks of *a random distribution of the incidents over the shifts*. He says *random* but he means *uniform random*. The idea that incidents are critical but rare medical crises might lead one to imagine that an incident can occur in any given shift of the year, with the same tiny chance. But the fact that one cannot a priori state when incidents are more likely to happen, does not mean that they do not have causes or contributing causes, and there can be time patterns in those causes.

We can think of three plausible reasons why incidents on a particular ward – though taken to be random – will not occur with constant risk throughout a whole year.

Firstly, one should realise that if we focus on a single ward on a hospital such as the Juliana children’s hospital, hospital administration and hospital policy influences how many and what kinds of patients are in that ward at each particular moment. Policy may change from time to time. Suppose that in order to economize, an intensive care ward is closed down (this happened during Lucia’s time working at the JKZ). Does this have no influence at all on the severity of the cases in the different wards, and the numbers of children in different wards (intensive care, and in medium-intensive care)? We would guess that it does have influence. Potential patients can be referred away to other hospitals, if the remaining wards are full. The hospital will presumably adjust its policy regarding admission (especially admission for a serious operation whose exact timing is to some extent a matter of choice of the doctors, patient and family), and to some extent adjust the transfer and discharge policy, both inside the hospital (from intensive to medium care, for instance) and outside, to adjust to the new capacity of the hospital.

In the case of the Juliana Children’s hospital there is another rather sensitive matter of policy: whether very ill children, who are not going to live for very long, should die at home or in the hospital wards. We understand that this policy did change once at the JKZ in the period of interest. Presumably a change in policy concerning where the hospital wanted children to die, would be implemented by changing admission, transfer, and discharge policy on individual patients in individual wards of the hospital. One can imagine that an intended hospital policy change would lead to adjustments in ward policy, and these adjustments would lead to changes in incident rates on particular wards. Those who are going to die, are going to die anyway, but the time and place where they die is altered by changing the kind of care they are given. Furthermore, it cannot be excluded that the

intended and the actual effects of policy changes differ, given the complexity and sensitivity of the situation.

Secondly, possibly the time of year has some influence on the rate at which incidents happen on a ward. We are talking, at JKZ, about severely ill young children who are afflicted with multiple medical problems caused by multiple genetic defects. Will it make a difference if the same child is in the hospital in winter or in summer? Will the same kinds of children be in hospital in summer and winter, anyway? One might imagine that thanks to central heating and air-conditioning, the climate inside the hospital is identical in summer and winter, but thinking of influenza and common cold viruses, hay-fever from pollen in the spring, smog from differing traffic intensities and differing weather at different times of year, it would seem that there is no reason to assume that the risk of incidents is exactly constant during the whole course of a year. Even with air-conditioning, the climate inside the hospital's wards is not the same in summer and winter.

Thirdly, incidents occur during the treatment of specific patients, and certain patients could have much higher risk of incidents than others. One patient at JKZ was responsible for three incidents, another for two; in both cases, within relatively short time periods.

So, taking shifts as fixed, we might consider incidents as being random and rare events – but this falls far short of the assumption that they have a constant probability throughout the year. Falsely assuming a constant risk is convenient, but not necessarily realistic. One might hope that the *arbitrariness* of which shifts are given to a particular nurse, and the *arbitrariness* of the shifts in which incidents happen, taken together, might alleviate the consequences of the incorrectness of either assumption of *uniformity* or *constancy*. But this would be wishful thinking. We argue that the combined effect of both departures from uniformity is to aggravate, not to neutralize, their negative effects.

In actual fact the shifts of a given nurse follow a rather systematic pattern. The nurse takes one of the three shifts (morning, evening, night) every day for a week or two. Recall that there have to be nurses on duty throughout the night and throughout the weekends, while the medical specialists tend to have “normal working hours”. After a run of a certain kind of shift, with or without “weekend breaks” (which need not be in the weekend at all), the nurse will maybe have a week's break. Occasionally she will go on a training course. She takes free days and vacations, just like everyone else. Sometimes she herself is at home on sick leave.

Shifts are allocated on a week to week or monthly basis, but from time to time there are alterations to the schedule. Vacation and courses are planned when they can be accommodated without difficulty into the hospital regime. Because of sickness or other events, nurses may swap shifts with one another, in informal agreements among themselves and their supervisors. Some nurses might prefer to avoid shifts which can be expected to be particularly difficult, others might on the contrary like to experience the challenges. The nurses working on a particular ward have

various degrees of qualification and varying characteristics: full-time or part-time, fully qualified or in training, permanent staff or temporary staff. Sometimes a nurse is temporarily “lent out” to another ward. Presumably not all shifts are the same as far as whether or not incidents can be expected to occur; and which nurses are given which shifts, is not random at all. There are requirements concerning the presence always of a nurse with certain qualifications. There are less nurses on duty during the night than during the day (mealtimes, washing, medication is preferably done during the day).

Taken together, even if we consider both the shifts of a given nurse as a random process, and the incidents on a ward as a random process, and even if we consider the two processes as stochastically independent of one another, the assumption of constant intensities of either is a guess, not based on any evidence or argument. There may be patterns in the risk of incidents and there are certainly patterns in the shifts of nurses. These patterns may be correlated, through the process by which shifts are shared over the different nurses according to their different personal situations, their different wishes for particular kinds of shifts, their different qualifications, and the changing situation on the ward.

4.5 Concluding remarks on the effects of heterogeneity

In the body of this paper we have shown the dramatic effect a modest amount of heterogeneity can have on tail probabilities, the probability that one nurse would experience a strikingly large amount of incidents. We have two main sources of heterogeneity: time variation, and nurse variation. As we have sketched, there may be correlation between the two. Time variation can be separated into medium term (week to week or month to month variation) and to the variation within weeks and even days - morning, evening, night shifts; weekdays and weekends. In the case of Lucia de Berk we have no data whatsoever about the shifts of other nurses. We do know that at Lucia’s ward on JKZ there were many temporary and part-time staff, including many trainees. Lucia was one of a small number of permanent full-time staff, as can be observed by the imbalance between the maximum number of nurses (27) who according to the staffing plan could be employed on the ward at any particular time, and the number of shifts which Lucia worked, and the number of nurses working on the ward within any particular shift (three or four during the day; one or two at night). Regarding time variation in shifts and incidents we do have some data: we know the actual time-schedule and we could in principle take account of medium term (week to week, month to month) time variation by stratification. The rapid increase in the number of parameters and the small actual number of incidents would have an enormous impact on the tail probabilities.

Heterogeneity of any kind increases the variation in the number of incidents experienced by a randomly chosen nurse over a given period of time (given number of shifts). From the well-known relations

$$E(X) = E(E(X|Y)),$$

$$\text{var}(X) = \text{E}(\text{var}(X|Y)) + \text{var}(\text{E}(X|Y)),$$

it follows that whereas for a Poisson distributed random variable (a suitable model for small number of independent rare events) variance and mean are equal, for a mixture of Poisson's (with different conditional means), variance is larger than mean. So whether a nurse has differing proportions of shifts in different time periods, while the incident rate varies over time, or whether different nurses take or are allocated different types of shifts from one another, or whether some nurses experience more or less incidents than other for "innocent" reasons as mentioned above, in all cases the end-result is *overdispersion* caused by *heterogeneity*.

The regular pattern of shifts of a given nurse, and the lack of a definition of "incident with a shift" can easily aggravate this situation in the context of a murder investigation triggered by the observation of an unlikely coincidence. Since a full-time nurse works one shift of every three for a longish period, almost all shifts with an incident are adjacent to one of her shifts, if not actually one of her shifts. If hospital investigators suspect this nurse and have latitude in how to assign an incident to a shift, as well as in what they consider an incident at all, they have a lot of latitude to *locate* incidents in the shifts of that nurse. In the case of Lucia de Berk, a formal definition of *incident* never seems to have been made, let alone of *shift with an incident*. At some times attention was restricted to reanimations (whether successful or unsuccessful). Since there was no central registration of reanimations, whether or not a reanimation occurred was determined by memory of hospital staff and then checking incomplete ward-log-books and patient-case-files which had never been intended for this purpose. The hospital administrators were later anxious to adopt a much wider notion of incident. They were being asked to objectively support their already publicly declared observation that Lucia had been present at reanimations far too often. The prosecution and later the judges also wanted to support their claim that Lucia had murderous inclinations by regarding any medical emergencies in which she was thought to be involved as murder attempts, however minor the incident.

Even if we restrict attention to the more or less objectively defined (though not centrally registered) event *reanimation*, we still have to assign this event to a shift. Perhaps we should focus on the time of the events which immediately lead up to the attempted reanimation. Perhaps we should focus on events during the reanimation attempt. In the case of an unsuccessful reanimation, the time of death which a doctor wrote on a death certificate is presumably not the relevant time. This introduces a major subjective element which must be avoided in any serious statistical investigations.

In the case of Lucia de Berk, there was no formal research protocol which laid down an intersubjective definition of "shift with an incident" which could be objectively verified by recourse to existing hospital documentation. Instead the hospital relied on memory and subjective assignments, made by medical staff and hospital administrators who were all too aware of the purpose of the investigation.

In at least two cases a death happening in the shift after Lucia's was turned into an incident in her shift, even though nothing was noticed at the time.