Poisson points, resetting, universality and the role of the last item

Claude Godrèche

Poisson points, resetting, universality and the role of the last item

Claude Godrèche

Université Paris-Saclay, CNRS, CEA, Institut de Physique Théorique,
91191 Gif-sur-Yvette, France

Abstract. For a stochastic process reset at random times, we discuss to what extent the probabilities of some orderings of observables associated with the intervals of time between resetting events are universal, i.e., independent of the choice of the observables, and in particular, to what extent universality depends on the choice of the distribution of these intervals. For Poissonian resetting, universality relies only on a combinatorial argument and on the statistical properties of Poisson points. For a generic distribution of time intervals between resets, universality no longer holds in general.

21 April 2023

E-mail: `claude.godreche@ipht.fr`

Consider n iid (independent, identically distributed) continuous random variables $X_1, X_2, \ldots, X_n$. The probability that any of them, say the last one for example, is larger than the $n-1$ remaining ones is equal to $1/n$. This also holds if these variables are exchangeable, i.e., if a permutation of these variables does not change their joint distribution [1]. Exchangeable random variables are not necessarily independent, while iid random variables are necessarily exchangeable. Similarly, the probability that these random variables are monotonically increasing, $\mathbb{P}(X_1 < X_2 < \cdots < X_n)$, (or decreasing) is equal to $1/n!$. One can likewise consider events involving orderings of these random variables of increasing complexity, whose probability can be arbitrarily difficult to determine (see, e.g., [2]). These probabilities are independent of the common distribution of $X_1, X_2, \ldots, X_n$, as long as the latter is continuous.

If now the number of these random variables is no longer fixed but is itself a fluctuating quantity, that henceforth we denote by N_t (with a notation that will shortly appear natural), then the probability of these different events should be averaged with the probability $\mathbb{P}(N_t = n)$.

In this letter we will expand on the aforementioned topic. In particular, we will provide an alternative perspective on certain questions raised in the recent work [3]. As stated in [3], given a random process undergoing stochastic resetting at a constant rate r to a position drawn from some distribution, if the random variables X_i are dynamical observables associated with the intervals between resetting events‡, the probabilities of the events oulined above (such as the last element being greater than all preceding elements or the sequence being monotonically increasing) are 'super-universal', meaning they are independent of the specific process, the observables X_i and the distribution of the restarting position.

As we now demonstrate, this universality is a simple consequence of the properties of Poisson processes. Indeed, since the stochastic process is reset at a constant rate r§, the resetting events are Poisson points in time, with the following properties [4]:

(i) the number of points in $(0, t)$ is given by the Poisson distribution

$$\mathbb{P}(N_t = n) = \mathrm{e}^{-rt}\frac{(rt)^n}{n!}, \tag{1}$$

(ii) the $N_t + 1$ intervals between these points in $(0, t)$ are statistically equivalent, i.e., exchangeable.

The first property is well known. The second one is a consequence of a fundamental property of Poisson points, stating that, conditional on $N_t = n$, the positions of these n events in $(0, t)$ are independently and uniformly distributed on $(0, t)$.

Consider now the observables $X_1, X_2, \ldots, X_{N_t+1}$ associated with these N_t+1 time intervals between resetting events. For instance, if the stochastic process is Brownian motion, X can be the length τ of the interval between two resettings, or the maximum attained by the process in this interval, or else the area under the process, and so on. From the above, it is clear that for each of these cases the corresponding random variables are exchangeable. Hence, denoting by $Q_1(t)$ the probability that the last observable (be it the interval τ, or the maximum, or the area, etc) is larger than all previous ones, we have

$$Q_1(t) = \left\langle \frac{1}{N_t + 1} \right\rangle = \sum_{n\geq 0} \frac{\mathbb{P}(N_t = n)}{n+1} = \frac{1 - \mathrm{e}^{-rt}}{rt} \underset{t\to\infty}{\approx} \frac{1}{rt}. \tag{2}$$

‡ We refer the reader to [3] for an introduction to the literature on resetting processes.

§ For simplicity we assume that the process is reset at the origin.

Similarly, let $Q_2(t)$ be the probability of the event $\{X_1 < X_2 < \cdots < X_{N_t+1}\}$. We have

$$Q_2(t) = \left\langle \frac{1}{(N_t+1)!} \right\rangle = \sum_{n\geq 0} \frac{\mathbb{P}(N_t=n)}{(n+1)!} = \frac{\mathrm{e}^{-rt}}{\sqrt{rt}}\, I_1(2\sqrt{rt}) \underset{t\to\infty}{\approx} \frac{\mathrm{e}^{-rt+2\sqrt{rt}}}{2\sqrt{\pi}(rt)^{3/4}}, \tag{3}$$

where $I_n(z)$ is the nth modified Bessel function of the first kind. The same applies to the probabilities of more complex orderings of events.

These conclusions were reached in [3] using a more computational scheme. The analysis given above is purely combinatorial and relies on the statistical properties of Poisson points.

The natural question to ask at this stage is: what happens if the N_t+1 intervals in $(0,t)$ are no longer exchangeable? As is well known, the Poisson process is the simplest renewal process [5, 6], where the interarrival times are exponentially distributed with density $\rho(\tau) = r\mathrm{e}^{-r\,\tau}$ and where the number of events N_t in $(0,t)$ is given by the Poisson distribution (1). For any renewal process, the intervals of times between events obey the sum rule

$$\tau_1 + \tau_2 + \cdots + \tau_{N_t} + B_t = t, \tag{4}$$

where, B_t, the last interval between the observation time t and the last event to its left, named the backward recurrence time, is, in the particular case of a Poisson process, statistically equivalent to the N_t intervals $\tau_1, \ldots, \tau_{N_t}$. This property was used above to demonstrate the universality of the results (2), (3), regardless of the choice of the observable X. For a generic renewal process however, i.e., for a generic density $\rho(\tau)$ of the intervals of time between resetting events, this property no longer holds, namely the last interval is no longer statistically independent of the N_t previous ones. One is therefore naturally led to investigate what remains of the universality property mentioned above.

As a test bed, let us focus on the probability that the last observable is the largest one, applied, firstly, to the case where the observable in question is the interval length itself,

$$Q_1^{(\tau)}(t) = \mathbb{P}\big(B_t > \max(\tau_1, \ldots, \tau_{N_t})\big), \tag{5}$$

where the superscript in the notation of this probability enhances the fact that the latter depends on the choice of the observable, as we now show. The analysis of this case was given in [7], with the following result, in Laplace space,

$$\underset{t}{\mathcal{L}}\, Q_1^{(\tau)}(t) = \hat{Q}_1^{(\tau)}(s) = \int_0^\infty \mathrm{d}b\, \frac{\mathrm{e}^{-sb}\int_b^\infty \mathrm{d}\tau\, \rho(\tau)}{1-\int_0^b \mathrm{d}\tau \rho(\tau)\mathrm{e}^{-s\tau}}. \tag{6}$$

If $\rho(\tau) = r\,\mathrm{e}^{-r\,\tau}$, the result (2) is recovered [7]. If $\rho(\tau)$ is uniform $\mathcal{U}(0,1)$, it is found that (see [7])

$$Q_1^{(\tau)}(t) \underset{t\to\infty}{\approx} \frac{1}{2\,t^2}, \tag{7}$$

which is different from the naive estimate $Q_1^{(\tau)}(t) \approx \langle 1/(N_t+1)\rangle$, decaying as $1/2t$ when $t\to\infty$.

If observables $X_1, X_2, \ldots$ are attached to the intervals $\tau_1, \tau_2, \ldots$, skipping all details, (6) is generalised into

$$\underset{t}{\mathcal{L}}\, Q_1^{(X)}(t) = \int_0^\infty \mathrm{d}y\, \frac{\hat{\psi}(y,s)}{1-\int_0^y \mathrm{d}x\, \hat{\varphi}(x,s)}, \tag{8}$$

with the definitions

$$\hat{\varphi}(x,s)=\int_0^\infty \mathrm{d}\tau\,\mathrm{e}^{-s\tau}\rho(\tau)f(x|\tau),\quad \hat{\psi}(y,s)=\int_0^\infty \mathrm{d}b\,\mathrm{e}^{-sb}p(b)f(y|b),$$

where $f(x|\tau)=\frac{\mathrm{d}}{\mathrm{d}x}\mathbb{P}(X<x|\tau)$ is the conditional density of X given τ, and $p(b)=\int_b^\infty \mathrm{d}\tau\,\rho(\tau)$. One can check on these expressions that if $X\equiv\tau$, then (8) reduces to (6). On the other hand, analysing (8) in all generality is not a simple task, therefore we shall resort to a heuristic argument in order to obtain an asymptotic estimate of $Q_1^{(X)}(t)$.

Consider the N_t observables $X_1, X_2\ldots$, obtained by taking some random function of the intervals $\tau_1,\ldots,\tau_{N_t}$. For example,

$$X_1=\tau_1^a\,\zeta_1,\ldots,\ X_{N_t}=\tau_{N_t}^a\,\zeta_{N_t},$$

where a is a positive exponent and $\zeta_1,\zeta_2,\ldots$ are iid random variables, independent of $\tau_1,\tau_2,\ldots$, whose common density $\rho(\tau)$ is taken uniform $\mathcal{U}(0,1)$, as in the example leading to (7). Likewise, we associate the random variable $B_t^a\,\zeta_{N_t+1}$ to the last interval B_t. Furthermore, we define $X_{N_t+1}=\tau_{N_t+1}^a\,\zeta_{N_t+1}$, where the interval τ_{N_t+1} straddles the observation time t, and the excess time E_t by $B_t+E_t=\tau_{N_t+1}$ [5, 6]. The question is whether the probability

$$Q_1^{(X)}(t)=\mathbb{P}(B_t^a\,\zeta_{N_t+1}>\max(X_1,\ldots,X_{N_t})) \tag{9}$$

is the same as (5) or at least has the same asymptotic behaviour (7). A heuristic argument, based on extreme value statistics, and confirmed by numerical simulations, shows that, if the common distribution of $\zeta_1,\zeta_2,\ldots$ is exponential, or Gaussian, or more generally is in the Gumbel class [9], then

$$Q_1^{(X)}(t)\underset{t\to\infty}{\approx}\frac{c}{a\,t\,\ln t}, \tag{10}$$

which is quite different from the result (7). The constant c appearing in (10) depends on the choice of distribution for the random variables ζ.

The heuristic argument is as follows. Let for example $\zeta_1,\zeta_2,\ldots$ be exponentially distributed random variables with unit parameter. Let us fix $N_t=n$, and let us denote by $X_{(1)}$ the largest amongst the $n+1$ random variables $X_1,\ldots,X_{n+1}$, and by $X_{(2)}$ the second largest one. When the event $\{B_t^a\,\zeta_{n+1}>\max(X_1,\ldots,X_n)\}$ holds, then necessarily $X_{(1)}=X_{n+1}\equiv\tau_{n+1}^a\zeta_{n+1}$, which occurs with probability $1/(n+1)$, and $X_{(2)}=\max(X_1,\ldots,X_n)$. Given this holds, we have the following equivalent inequalities

$$B_t^a\,\zeta_{n+1}>X_{(2)}\iff(\tau_{n+1}-E_t)^a\zeta_{n+1}>X_{(2)}\iff\Big(1-\frac{E_t}{\tau_{n+1}}\Big)^a X_{(1)}>X_{(2)}. \tag{11}$$

We note that in order for $X_{(1)}$ and $X_{(2)}$ to be the largest observables, asymptotically, i.e., when n and t are large, the intervals attached to them are necessarily close to unity, which entails that $E_t\ll 1$, as well as $X_{(1)}\approx\zeta_{(1)}$, $X_{(2)}\approx\zeta_{(2)}$. Therefore the rightmost inequality in (11) can be rewritten as

$$1-a\frac{E_t}{\tau_{n+1}}>\frac{X_{(2)}}{X_{(1)}}\approx\frac{\zeta_{(2)}}{\zeta_{(1)}},$$

which leads finally, in the asymptotic regime, to the equivalence

$$B_t^a\,\zeta_{n+1}>X_{(2)}\iff E_t<\frac{\zeta_{(1)}-\zeta_{(2)}}{a\,\zeta_{(1)}}. \tag{12}$$

At large times the distribution of E_t becomes stationary, i.e., $E_t \to E_{\rm eq}$, with (see e.g., [8])

$$\mathbb{P}(E_{\rm eq} < x) = \frac{1}{\langle\tau\rangle}\int_0^x {\rm d}e\,(1-e) \approx \frac{x}{\langle\tau\rangle}, \tag{13}$$

because x (playing the role of the right-hand side of (12)) is small (see (15)). So

$$\mathbb{P}(B_t^a\,\zeta_{n+1} > X_{(2)}) \approx \frac{\zeta_{(1)}-\zeta_{(2)}}{a\langle\tau\rangle\zeta_{(1)}}. \tag{14}$$

Using a well-known extreme value statistics argument, we have, for large n,

$$\mathrm{e}^{-\zeta_{(1)}} \approx \frac{\xi_1}{n}, \qquad \mathrm{e}^{-\zeta_{(2)}} \approx \frac{\xi_1+\xi_2}{n},$$

where ξ_1 and ξ_2 are exponentially distributed with parameter unity, hence

$$\frac{\zeta_{(1)}-\zeta_{(2)}}{\zeta_{(1)}} \approx \frac{\ln\frac{\xi_1+\xi_2}{\xi_1}}{\ln n}. \tag{15}$$

Averaging this expression on ξ_1 and ξ_2, we finally get, after multiplication of (14) by $1/(n+1)$ and replacement of n by $t/\langle\tau\rangle$,

$$Q_1^{(X)}(t) \underset{t\to\infty}{\approx} \frac{1}{a\,t\ln t},$$

which is (10) with $c=1$. This very reasoning can be extended to other distributions of the random variables $\zeta_1, \zeta_2, \ldots$. For instance if the latter variables are uniform $\mathcal{U}(0,1)$, one obtains

$$Q_1^{(X)}(t) \underset{t\to\infty}{\approx} \sqrt{\frac{\pi}{16\,t^3}}, \tag{16}$$

which again is different from (7).

Comparing (7) to (10) and (16) shows that the degeneracy (independence in the choice of observable) holding for Poissonian resetting (see (2)) is lifted, i.e., the probability $Q_1^{(X)}(t)$ now depends on the choice of the observable X, with different expressions according to whether $X \equiv \tau$ or $X \equiv \tau^a\zeta$, as soon as $\rho(\tau)$ is no longer exponential.

The question of whether universality extends beyond the Poissonian resetting protocol, i.e., with a non-exponential distribution $\rho(\tau)$ was raised in [3], where it was anticipated that this universality would still hold, however only for large t, and provided $\rho(\tau)$ decays sufficiently fast at large τ. As demonstrated above, this does not hold true in general. A complete discussion of universality beyond this point would require a thorough investigation. At the very least, the analysis given above demonstrates the role played by the last item (interval or observable attached to the latter) for the determination of $Q_1^{(X)}(t)$. For probabilities of other orderings of $X_1, X_2, \ldots$ (such as for example $Q_2^{(X)}(t)$), the role of the last element is probably all the less important as the definitions of the corresponding events give less weight to this last element.

As a corollary of the above, a way of restoring exchangeability at any finite time consists in considering the N_t first intervals only, discarding the last one, B_t. Now, since these intervals are statistically equivalent, the same holds for any observable associated with these intervals. Hence the universality property holding with Poisson points for the complete sequence $X_1, \ldots, X_{N_t+1}$ now holds for the restricted sequence

$X_1, \ldots, X_{N_t}$ regardless of the choice of density $\rho(\tau)$. This is illustrated below on a few examples.

Consider first the probability that the last observable of the sequence is the largest one, denoted by $q_1(t)\|$. Then, for any distribution of the intervals $\rho(\tau)$, and for any choice of the observables attached to these intervals, we have

$$q_1(t) = \left\langle \frac{1}{N_t} \right\rangle,$$

(where $N_t > 0$). For example, for Poisson points, with $\rho(\tau) = \mathrm{e}^{-r\tau}$, this probability was computed in [7], with the result

$$q_1(t) = \mathrm{e}^{-rt} \int_0^{rt} \mathrm{d}u \, \frac{\mathrm{e}^u - 1}{u} \underset{t\to\infty}{\approx} \frac{1}{rt},$$

which has the same asymptotic behaviour as in (2). This asymptotic behaviour also holds for any distribution of the intervals with a finite first moment [7]. For a generic density $\rho(\tau)$, in Laplace space, we have [7],

$$\underset{t}{\mathcal{L}}\, q_1(t) = -\frac{1-\hat{\rho}(s)}{s} \ln(1-\hat{\rho}(s)).$$

As an application, let us consider a simple example of a distribution $\rho(\tau)$ with a fat tail, namely such that $\hat{\rho}(s) = \mathrm{e}^{-\sqrt{s}}$. We obtain

$$q_1(t) \underset{t\to\infty}{\approx} \frac{\ln(4t) + \gamma}{2\sqrt{\pi t}},$$

(see [7]) where γ is Euler's constant.

Likewise, the probability of the event $\{X_1 < X_2 < \cdots < X_{N_t}\}$ is, for any distribution of the intervals $\rho(\tau)$,

$$q_2(t) = \left\langle \frac{1}{N_t!} \right\rangle,$$

hence, for Poisson points, for example,

$$q_2(t) = \mathrm{e}^{-rt} I_0(2\sqrt{rt}) \underset{t\to\infty}{\approx} \frac{\mathrm{e}^{-rt+2\sqrt{rt}}}{2\sqrt{\pi}(rt)^{1/4}},$$

which decays slightly slower than (3). For a generic density $\rho(\tau)$, in Laplace space, we have

$$\underset{t}{\mathcal{L}}\, q_2(t) = -\frac{1-\hat{\rho}(s)}{s} \mathrm{e}^{\hat{\rho}(s)}.$$

For the example of $\hat{\rho}(s) = \mathrm{e}^{-\sqrt{s}}$, we obtain

$$q_2(t) \underset{t\to\infty}{\approx} \frac{\mathrm{e}}{\sqrt{\pi t}}.$$

In Laplace space, more complicated expressions would be obtained for the case of more complex orderings of events.

Other considerations on the role of the last interval in renewal processes can be found in [7]. As a final comment, the role of the last item, highlighted in the present letter, is reminiscent of the role of boundary conditions in the language of statistical physics.

$\|$ This probability was denoted by $Q^{\mathrm{III}}(t)$ in [7].

References

[1] Kingman J F C 1978 *Uses of exchangeability* Ann. Probability **6** 183
[2] Luck J M 2014 *On the frequencies of patterns of rises and falls* Physica A **407** 252
[3] Smith N, Majumdar S N and Schehr G 2023 *Striking universalities in stochastic resetting processes* arXiv preprint arXiv:2301.11026
[4] Stirzacker D 2005 *Stochastic processes and models* (Oxford: Oxford University Press)
[5] Cox D R and Miller H D 1965 *The Theory of Stochastic Processes* (London: Chapman & Hall)
[6] Feller W 1968, 1971 *An Introduction to Probability Theory and its Applications*, Volumes 1&2 (New York: Wiley)
[7] Godrèche C, Majumdar S N and Schehr G 2015 *Statistics of the longest interval in renewal processes* J. Stat. Mech. P03014
[8] Godrèche C and Luck J M 2001 *Statistics of the occupation time of renewal processes* J. Stat. Phys. **104** 489
[9] Gumbel E. J. 2004 *Statistics of Extremes* (NY: Dover)