History

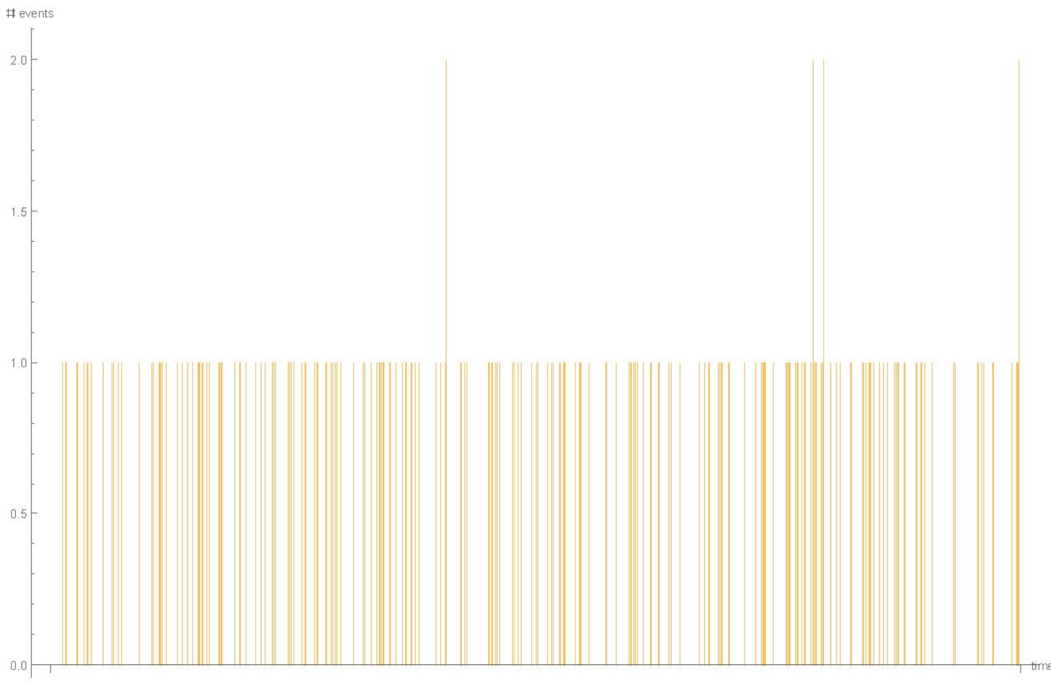
https://twitter.com/nntaleb/status/742819547095420930

Homogeneous Poisson

Randomness does not look random. One should be careful in declaring a “trend” from purely random Poisson processes, particularly for narrow windows. What we have below is a random exponential arrival time, which shows periods with no event, and periods with a concentration of events. This is why one should never “p-hack” arrival times: you test from the structure. **S. Pinker and the other unsophisticated coauthor Spagat** (to be polite) accepted our statement on homogeneous Poisson but wondered why we didn’t look at “subtrends” in data. An analysis of dependence structure (say correlation functions) **does take into account** subtrends in data.
We simulate 10,000 years of “history”.

```
tab = Table[RandomVariate[PoissonDistribution[.019]], {10000}];
BarChart[tab, AxesLabel → {time, "# events"}]
```

In[ ]:=



Checking recovery

```
edist = EstimatedDistribution[tab, PoissonDistribution[λ]]
```

```
PoissonDistribution[0.0186]
```

$$C F(h) = \sum_{i=1}^{n-h} (x_i - \hat{\mu}) (x_{i+h} - \hat{\mu}) \Big/ \sum_{i=1}^{n} (x_i - \hat{\mu})^2$$

```
Table[CorrelationFunction[tab, n], {n, 2, 6}] // N
```

```
{-0.00799983, -0.00252348, 0.00295286, 0.0084292, -0.00252917}
```

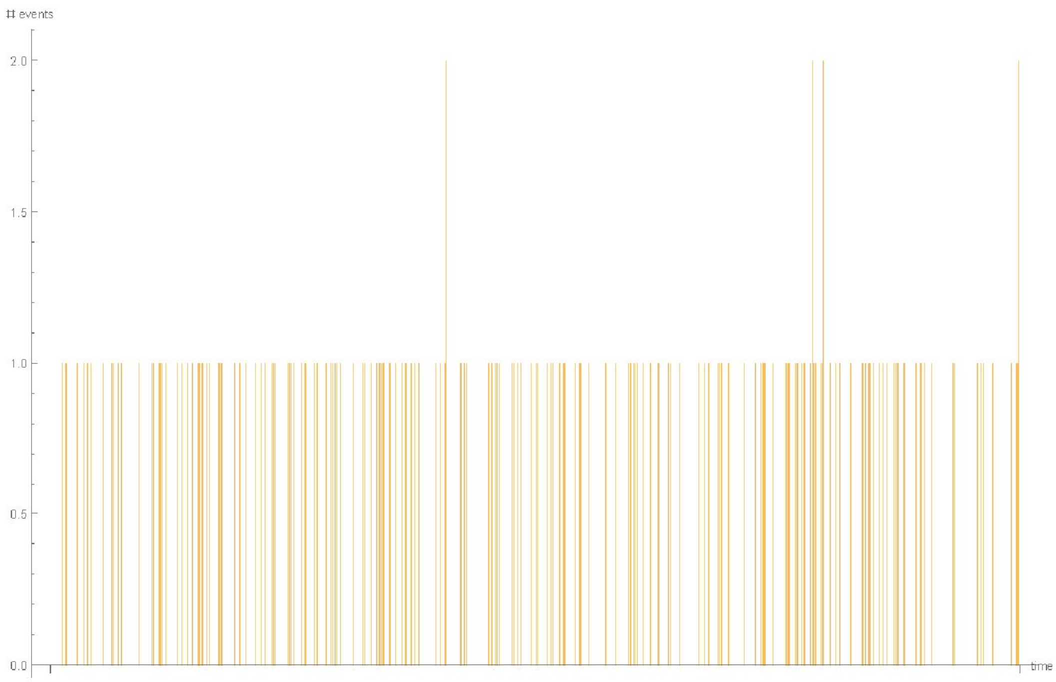
So although there appears to be there is no dependence structure from the top. An additonal test is to reshuffle and compare time windows.

Out[ ]=

Homogeneous Poisson

Randomness does not look random. One should be careful in declaring a "trend" from purely random Poisson processes, particularly for narrow windows. What we have below is a random exponential arrival time, which shows periods with no event, and periods with a concentration of events. This is why one should never "p-hack" arrival times: you test from the structure. **S. Pinker and the other unsophisticated coauthor Spagat** (to be polite) accepted our statement on homogeneous Poisson but wondered why we didn't look at "subtrends" in data. An analysis of dependence structure (say correlation functions) **does take into account** subtrends in data.
We simulate 10,000 years of "history".

```
tab = Table[RandomVariate[PoissonDistribution[.019]], {10000}];
BarChart[tab, AxesLabel → {time, "# events"}]
```



Checking recovery

```
edist = EstimatedDistribution[tab, PoissonDistribution[λ]]
```

```
PoissonDistribution[0.0186]
```

$$C\,F(h) = \sum_{i=1}^{n-h} (x_i - \hat{\mu})(x_{i+h} - \hat{\mu}) \Big/ \sum_{i=1}^{n} (x_i - \hat{\mu})^2$$

```
Table[CorrelationFunction[tab, n], {n, 2, 6}] // N
```

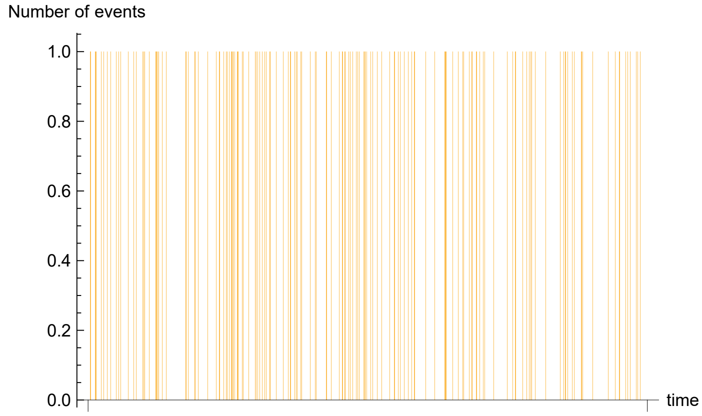
```
{-0.00799983, -0.00252348, 0.00295286, 0.0084292, -0.00252917}
```

So although there appears to be there is no dependence structure from the top. An additonal test is to reshuffle and compare time windows.

In[ ]:=
```
tab = Table[RandomVariate[PoissonDistribution[0.019]], {10000}];
```

In[ ]:=
```
BarChart[tab, AxesLabel → {time, "Number of events"}]
```

Out[ ]=



You can check the recovery of the Poisson distribution.
I did not know this is in deed the case.

In[ ]:=
```
edist = EstimatedDistribution[tab, PoissonDistribution[λ]]
```

Out[ ]=
```
PoissonDistribution[0.0169]
```

What is the correlation function? But yes the values are very low.

```
? CorrelationFunction
```

In[ ]:=
```
Table[CorrelationFunction[tab, n], {n, 2, 6}] // N
```

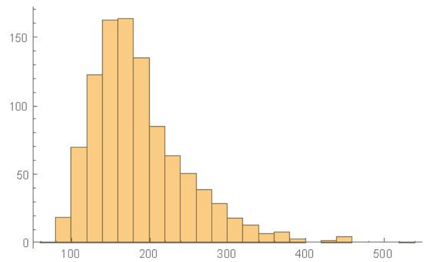
Out[ ]=
```
{0.000862679, -0.00515792, 0.00687812, 0.024933, -0.00516308}
```

Another approach to check nonrandomness from the dist of the maximum

Now we need the distribution of the maximum of uninterrupted runs

```
tableofruns =
 Table[SequenceCases[ta = Table[RandomVariate[PoissonDistribution[.019]], {1000}], {p : Repeated[0]} :> Length[{p}]] // Max,
  {1000}]
```

```
h1 = Histogram[tableofruns, "Probability"]
```




```
Mean[tableofruns] // N
```

```
187.284
```

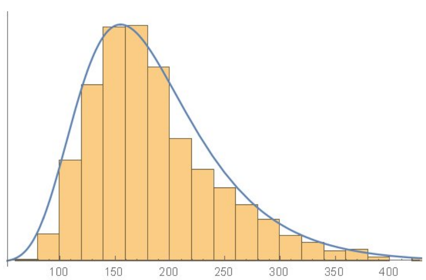
In[•]:=

Analytically, we can derive it as follows. Consider the interrival times of a Poisson process ($\tau_1, \tau_2, \ldots \tau_n$) where τ_i is the elapsed time between arrival i and $i+1$. We have $P(\tau_1 > t) = P(n(t) = 0) = e^{-\lambda t}$, hence the CDF $P(\tau_1 \leq t) = 1 - e^{-\lambda t}$. The density is that of the exponential distribution $\phi(t) = e^{-t\lambda}\lambda,\ t \geq 0$, and the mean is $\frac{1}{\lambda}$. Now over many trials m we have the distribution of the longest period:

$$f_{\max} = \frac{\partial\left(1 - e^{-\lambda t}\right)^m}{\partial x} = \lambda\, m\, e^{-\lambda t}\left(1 - e^{-\lambda t}\right)^{m-1}$$

We cheat with the heuristic that over a period T, $m \approx T\lambda$, which is of o(.). Hence over a long period

$$f_{\max}(T) = e^{-t\lambda}\left(1 - e^{-t\lambda}\right)^{-1+T\lambda}\, T\lambda^2$$



$$\int_0^\infty \lambda^2\, t\, T\, e^{\lambda(-t)}\left(1 - e^{\lambda(-t)}\right)^{\lambda T-1} dt = \frac{\text{HarmonicNumber}[T\lambda]}{\lambda}$$

```
HarmonicNumber[1000 × .019] / .019
```

```
186.723
```

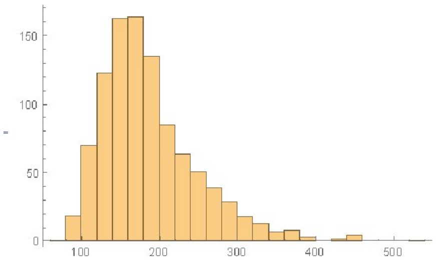
Out[ ]=

Another approach to check nonrandomness from the dist of the maximum

Now we need the distribution of the maximum of uninterrupted runs

```
tableofruns =
 Table[SequenceCases[ta = Table[RandomVariate[PoissonDistribution[.019]], {1000}], {p : Repeated[0]} :> Length[{p}]] // Max,
  {1000}]
```

```
h1 = Histogram[tableofruns, "Probability"]
```




```
Mean[tableofruns] // N
```

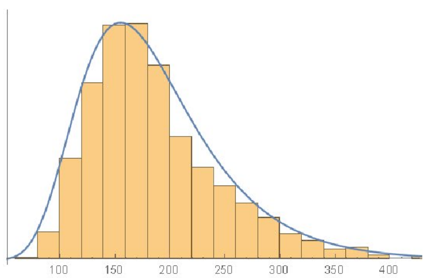
```
187.284
```

Analytically, we can derive it as follows. Consider the interrival times of a Poisson process (τ_1, τ_2, ... τ_n) where $\tau_i$is the elapsed time between arrival i and $i+1$. We have $P(\tau_1 > t) = P(n(t) = 0) = e^{-\lambda t}$, hence the CDF $P(\tau_1 \leq t) = 1 - e^{-\lambda t}$. The density is that of the exponential distribution $\phi(t) = e^{-t\lambda}\lambda$, $t \geq 0$, and the mean is $\frac{1}{\lambda}$. Now over many trials m we have the distribution of the longest period:

$$f_{\max} = \frac{\partial\left(1 - e^{-\lambda t}\right)^m}{\partial x} = \lambda\, m\, e^{-\lambda t}\left(1 - e^{-\lambda t}\right)^{m-1}$$

We cheat with the heuristic that over a period T, $m \approx T\lambda$, which is of o(.). Hence over a long period

$$f_{\max}(T) = e^{-t\lambda}\left(1 - e^{-t\lambda}\right)^{-1+T\lambda} T\lambda^2$$



$$\int_0^\infty \lambda^2\, t\, T\, e^{\lambda(-t)}\left(1 - e^{\lambda(-t)}\right)^{\lambda T - 1} dt = \frac{\text{HarmonicNumber}[T\lambda]}{\lambda}$$

```
HarmonicNumber[1000 × .019] / .019
```

```
186.723
```

In[ ]:=
```
tableofruns =
 Table[SequenceCases[ta = Table[RandomVariate[PoissonDistribution[0.019]], {1000}],
    {p : Repeated[0]} :> Length[{p}]] // Max, {1000}]
```

Out[ ]=
```
$Aborted
```

```
h1 = Histogram[tableofruns, "Probability"]
```

```
Mean[tableofruns] // N
```

The density follows a exponential distribution.

```
HarmonicNumber[1000 * 0.019] / 0.019
```

```
? :>
```