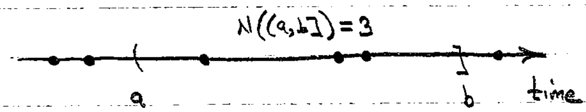
THE LAW OF RARE EVENTS.

Let $N((a,b])$ count the number of events that occur in the time interval $(a,b]$:



Assume:

① #'s of events in disjoint intervals are independent random variables:

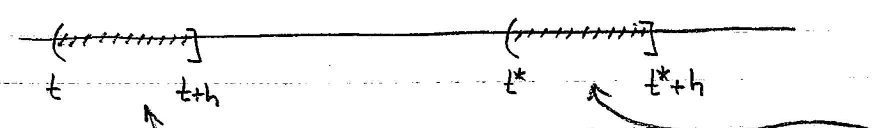
for $t_0 = 0 < t_1 < t_2 < \cdots < t_m$,

$N((t_0,t_1])$, $N((t_1,t_2])$, ..., $N((t_{m-1},t_m])$ are independent.

In particular, $P\left(N((t_0,t_1])=k \overset{\text{"and"}}{,} N((t_1,t_2])=\ell\right)$

$$= P\left(N((t_0,t_1])=k\right) P\left(N((t_1,t_2])=\ell\right).$$

② For any time t and any $h>0$, the probability distribution of $N((t,\, t+h])$ depends on h but not on t.



# of events in here has the same probability dist'n as # in here

③ $P\left(N((t,\, t+h])=1\right) = \lambda h + o(h)$ as $h \to 0$,

"little oh of h" means $\frac{o(h)}{h} \to 0$ as $h \to 0$.

where $\lambda > 0$ is a rate or intensity. Units of λ are events/time. Ex: calls/minute, deaths/year. λ is not a probability.

④ $P\left(N((t,\, t+h]) \geq 2\right) = o(h)$ as $h \to 0$. Events are rare; in a small time interval it is unlikely that 2 or more will occur.

By ②, we can consider $N((0,t])$ for arbitrary t instead of looking at $N((s, s+t])$ for arbitrary s and t.

Shorthand notation: $p_m(t) = P\left(N((0,t]) = m\right)$ for $m = 0, 1, 2, \ldots$ events.

If we can figure out $p_m(t)$, we will have completely described this "rare event" process.

Start at $m=0$:
$$p_0(t+h) = p_0(t)\, p_0(h) \quad \text{by ①}$$
$$= p_0(t)\left(1 - \lambda h + o(h)\right)$$
$$\Rightarrow \frac{p_0(t+h) - p_0(t)}{h} = -\frac{\lambda h\, p_0(t)}{h} + \frac{o(h)}{h}\, p_0(t)$$
$$p_0'(t) = -\lambda\, p_0(t) + 0$$

Solution is $p_0(t) = c e^{-\lambda t}$, but $p_0(0) = 1$

so $p_0(t) = e^{-\lambda t}$.

Try $m=1$:
$$p_1(t+h) = p_0(t)\, P\left(N((t,t+h]) = 1\right) + p_1(t)\, P\left(N((t,t+h]) = 0\right)$$
$$= p_0(t)\left(\lambda h + o(h)\right) + p_1(t)\left(1 - \lambda h + o(h)\right)$$
$$\Rightarrow \frac{p_1(t+h) - p_1(t)}{h} = \frac{\lambda h\, p_0(t)}{h} - \frac{\lambda h\, p_1(t)}{h} + \frac{o(h)}{h}$$
$$p_1'(t) = \lambda\, p_0(t) - \lambda\, p_1(t).$$

For general m:
$$p_m'(t) = \lambda\, p_{m-1}(t) - \lambda\, p_m(t).$$

Solution (check): $p_m(t) = \dfrac{(\lambda t)^m e^{-\lambda t}}{m!}$ for $m = 0, 1, 2, \ldots$

the POISSON DISTRIBUTION with mean λt

$N((s,t])$ is called a POISSON POINT PROCESS

and $X(t) = n_0 + N((0,t])$ is called a POISSON COUNTING PROCESS

fixed # events, occurred at $t \le 0$

or just a POISSON PROCESS.

Ex. Suppose prairie dog colonies are illegally poisoned according to a PP with rate = 5 illegal poisonings / year. What is the probability that 6 colonies will be poisoned in the next 3 months?

$m = 6, \quad \lambda = 5, \quad t = \frac{3}{12}, \quad P\left(N\left(\left(0, \tfrac{3}{12}\right]\right) = 6\right) = \dfrac{\left(\frac{15}{12}\right)^6 e^{-15/12}}{6!} = 0.001518.$

The waiting time, W_1, until the first event is a continuous random variable.

What is its distribution?

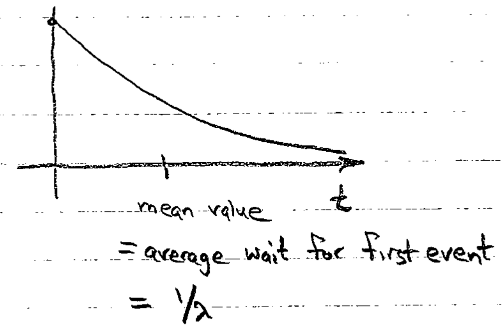
$$
\begin{aligned}
P(W_1 \le t) &= P(\text{wait no more than } t \text{ time units for the first event}) \\
&= P(\#\text{ events in } (0,t] \text{ is at least one, maybe more}) \\
&= P(N((0,t]) \ge 1) \\
&= 1 - P(N((0,t]) = 0) \\
&= 1 - e^{-\lambda t},
\end{aligned}
$$

the EXPONENTIAL cumulative dist'n fct. (cdf).

probability density function (pdf) of W_1 is then

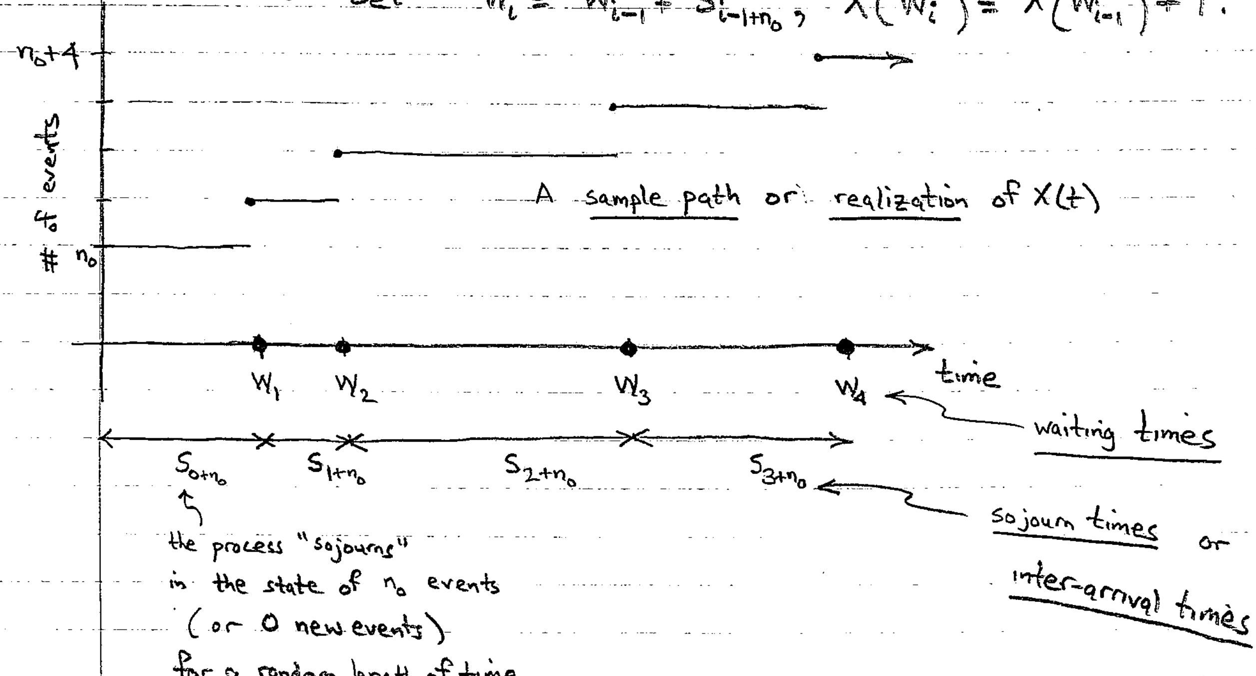
$$\frac{d}{dt}\left(1 - e^{-\lambda t}\right) = \lambda e^{-\lambda t}$$

$$\text{mean} = E[W_1] = \frac{1}{\lambda}, \quad \text{variance} = Var(W_1) = \frac{1}{\lambda^2}$$



SIMULATION OF A POISSON PROCESS, $X(t)$

- Start at time 0 with $X(0) = n_0$. Set the random number seed.
- Simulate $S_{0+n_0} \sim Exp(\lambda)$. Set $W_1 = S_{0+n_0}$, $X(W_1) = X(0) + 1$
- But by independence, the process effectively starts over at time W_1, so...
- For $i = 2, 3, \ldots, n$
  - Simulate $S_{i-1+n_0} \sim Exp(\lambda)$, independent of $S_{0+n_0}, \ldots, S_{i-2+n_0}$
  - Set $W_i = W_{i-1} + S_{i-1+n_0}$, $X(W_i) = X(W_{i-1}) + 1$.



INFERENCE FOR POISSON PROCESS

Data: $S_0, S_1, S_2, \dots, S_{n-1}$. Total of $n_0 + n$ events have occurred at time $W_n = S_0 + S_1 + \dots + S_{n-1}$. WLOG take $n_0 = 0$.

Now $S_0, S_1, S_2, \dots, S_{n-1}$ iid $\text{Exp}(\lambda)$, so

$$\text{likelihood of } \lambda = \mathcal{L}(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda S_{i-1}}$$

joint dist'n of the data, regarded as a function of any unknown parameters.

$$\text{and log-likelihood} = \ln \mathcal{L}(\lambda) = \sum_{i=1}^{n} \ln \lambda - \sum_{i=1}^{n} \lambda S_{i-1}$$

$$= n \ln \lambda - \lambda \sum_{i=1}^{n} S_{i-1}$$

$$0 \stackrel{\text{set}}{=} \frac{d \ln \mathcal{L}(\lambda)}{d\lambda} = \frac{n}{\lambda} - \Sigma S_{i-1} \Rightarrow \hat{\lambda}_{MLE} = \frac{1}{\sum_{i=1}^{n} S_{i-1}/n}$$

maximum likelihood estimator

Recall mean sojourn time $= \frac{1}{\lambda}$; MLE has invariance property

$$\widehat{\left(\frac{1}{\lambda}\right)}_{MLE} = \frac{1}{\hat{\lambda}_{MLE}} = \frac{\sum_{i=1}^{n} S_{i-1}}{n} = \text{sample mean sojourn time}$$

For large n, $\frac{1}{\hat{\lambda}_{MLE}} \stackrel{\cdot}{\sim} N\left(\frac{1}{\lambda}, \frac{1}{\lambda^2 n}\right)$ by central limit theorem.

So an approximate 95% confidence interval for $1/\lambda$ is

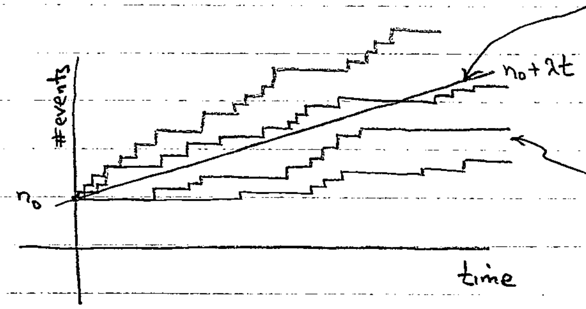
$$\frac{\Sigma S_{i-1}}{n} \pm 1.96 \sqrt{\left(\frac{\Sigma S_{i-1}}{n}\right)^2 \frac{1}{n}}$$

RELATION TO DETERMINISTIC MODEL

Poisson process $X(t) = n_0 + N((0,t])$, $N((0,t]) \sim \text{Poiss}(\lambda t)$.

$E[X(t)] = n_0 + \lambda t$; <u>linear growth</u> on average

$\text{Var}(X(t)) = \lambda t$.



Variability increases over time

Now the line is the solution to $dy = \lambda\, dt$ subject to $y(0) = n_0$.

with certainty; dy is <u>fixed</u>; dy is <u>always</u> λdt.

The random realizations can be thought of as

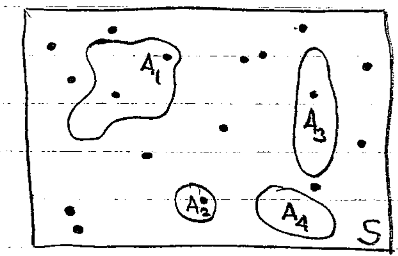
previously $\lambda h + o(h)$ for small time window h

$$\underline{\text{random}}\ \text{"}dy\text{"} = \begin{cases} 1 & \text{with probability } \lambda dt \\ 0 & \text{with probability } 1-\lambda dt \end{cases}$$

$$\text{so average "}dy\text{"} = (1)(\lambda dt) + (0)(1-\lambda dt) = \lambda dt$$

Poisson process discussed so far is <u>homogeneous</u> because λ does not change over time. Extends readily to <u>space</u>: consider 2-D.

Let $S \subset \mathbb{R}^2$ and let $A_j \subset S$.



$N(A_j) = \#$ events in A_j.

For a <u>HOMOGENEOUS POISSON POINT PROCESS</u>

with intensity $\lambda > 0$, $P\left(N(A_j) = k\right) = \dfrac{e^{-\lambda |A_j|} \left(\lambda |A_j|\right)^k}{k!}$ for $k = 0, 1, 2, \ldots$

area of A_j

and, for $A_1, A_2, \ldots, A_n$ <u>disjoint</u> areas,

$$P\left(N(A_1) = k_1, N(A_2) = k_2, \ldots, N(A_n) = k_n\right) = \prod_{j=1}^{n} P\left(N(A_j) = k_j\right)$$

(ie, # of events in disjoint sub-areas are independent.)

We can also consider <u>NONHOMOGENEOUS POISSON PROCESSES</u> in time or space.

In time, λ is replaced by $\lambda(t) > 0$.

Variation in $\lambda(t)$ might be explained by <u>covariates</u>,

such as $\ln \lambda(t) = \underset{\sim}{z}^T(t)\, \underset{\sim}{\beta}$ ← unknown coefficients that can be estimated via <u>POISSON REGRESSION</u>

Known covariates, like temperature/precip. in Brad B.'s human plague model

<u>Another example:</u> host-parasitoid interaction.

Suppose there are $H(t)$ hosts and $P(t)$ parasitoids.

(Total # encounters)/(unit time) = aHP

Assume parasitoids encounter a particular host "at random"; ie, according to a PP with rate = (# encounters)/(unit time) = $aHP/H = aP$,

so $P\left(\begin{matrix}\text{host escapes parasites} \\ \text{in } (t, t+1]\end{matrix}\right) = \dfrac{(aP \cdot 1)^0 \, e^{-aP \cdot 1}}{0!} = e^{-aP}$.

Note that for the homogeneous Poisson process we can rewrite some of our "Law of Rare Events" postulates as

① $P\left(X(t+h)-X(t)=1 \mid X(t)=K\right) = \lambda h + o(h)$ as $h \to 0$.

these events are independent for Poisson process

note that this does not depend on K

② $P\left(X(t+h)-X(t)=0 \mid X(t)=K\right) = 1-\lambda h + o(h)$.

③ $X(0) = n_0$. ← (often we take $n_0 = 0$ for convenience.)

Why not let λ depend on the number of events that have already occurred? Accordingly, let $\{\lambda_K\}$ be a sequence of positive real numbers. Define a <u>PURE BIRTH PROCESS</u> as a Markov process (future is conditionally independent of the past, given the present) satisfying the postulates:

① $P\left(X(t+h)-X(t)=1 \mid X(t)=K\right) = \lambda_K h + o_{1,K}(h)$

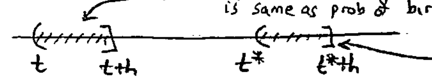
② $P\left(X(t+h)-X(t)=0 \mid X(t)=K\right) = 1-\lambda_K h + o_{2,K}(h)$

③ $P\left(X(t+h)-X(t)<0 \mid X(t)=K\right) = 0$

④ $X(0) = n_0$ ← often take $n_0=0$ so that $X(t) = \#$ births in $(0,t]$.

As before, this process has <u>stationary</u> transition probabilities in that they depend only on the width of the time window, h, <u>not</u> on its location, t.

prob of birth in here, given $X(t)=K$ is same as prob of birth in here given $X(t^*)=K$



As before, define $p_m(t) = P\big(X(t)=m\big)$ (assuming $n_0=0$).

Then, as before, derive

$$p_0'(t) = -\lambda_0\, p_0(t)$$

$$p_m'(t) = -\lambda_m\, p_m(t) + \lambda_{m-1}\, p_{m-1}(t) \qquad \text{for } m=1,2,\dots$$

with boundary conditions $p_0(0)=1$, $p_m(0)=0$ for $m\geq 1$.

As before, $p_0(t) = e^{-\lambda_0 t}$,

and so $P(S_0 \leq t) = P\big(\text{wait no more than } t \text{ for first birth}\big)$

$= P\big(\#\text{ births in } (0,t] \text{ is at least one, maybe more}\big)$

$= P\big(X(t)\geq 1\big) = 1 - P\big(X(t)=0\big)$

$= 1 - p_0(t) = 1 - e^{-\lambda_0 t}$, $\text{Exp}(\lambda_0)$.

But once you have the first birth, the process starts over in waiting for the second birth, though the rate has changed to λ_1

Thus $S_1 \sim \text{Exp}(\lambda_1)$, independent of S_0

and in fact $S_k \sim \text{Exp}(\lambda_k)$, with the S_k's mutually indep.

SIMULATION OF A PURE BIRTH PROCESS

- Start at time 0 with $X(0) = n_0$. Set the random number seed.
- Simulate $S_{0+n_0} \sim \text{Exp}(\lambda_0)$. Set $W_1 = S_{0+n_0}$, $X(W_1) = X(0)+1$
- For $i = 2, 3, \ldots$
    - Simulate $S_{i-1+n_0} \sim \text{Exp}(\lambda_{i-1})$, indep. of $S_{0+n_0}, S_{1+n_0}, \ldots, S_{i-2+n_0}$
    - Set $W_i = W_{i-1} + S_{i-1+n_0}$, $X(W_i) = X(W_{i-1}) + 1$.

Pure birth is a <u>very general</u> model: $\{\lambda_k\}$ are arbitrary positive numbers, so birth rates can speed up or slow down in any way. The model is <u>too general</u> for statistical inference; there are as many parameters $\{\lambda_k\}$ as recorded data points, the times between births $\{S_{k-1}\}$.

Need to impose <u>some structure</u> on $\{\lambda_k\}$ to get a useful model.

Simplest structure (beyond $\lambda_k \equiv \lambda$, a constant) is $\lambda_k = \beta k$, a line.