<u>Example</u>: Suppose independent tests are conducted on mice to develop a vaccine. If the prob of success is $1/3$ in each trial, what is the prob that at least 5 trials are needed to get the first success?

$X \rightarrow$ no of trials needed to get the first success.

$$P(X=k) = \left(\frac{2}{3}\right)^{k-1} \cdot \frac{1}{3}, \quad k=1,2,\cdots$$

$$P(X \geq 5) = \sum_{k=5}^{\infty} \left(\frac{2}{3}\right)^{k-1} \cdot \frac{1}{3}$$

$$= \left(\frac{2}{3}\right)^4 \cdot \frac{1}{3}\left[1 + \frac{2}{3} + \left(\frac{2}{3}\right)^2 + \cdots\right]$$

$$= \left(\frac{2}{3}\right)^4 \cdot \frac{1}{3} \cdot \frac{1}{1-\frac{2}{3}} = \left(\frac{2}{3}\right)^4$$

$$= \frac{16}{81}.$$

Memoryless Property of Geometric

Dist^n: $$P(X > m) = \sum_{k=m+1}^{\infty} q^{k-1} p$$

$$= q^m p \left(1 + q + q^2 + \cdots \right)$$

$$= q^m \cdot p \cdot \frac{1}{(1-q)} = q^m$$

So P(success has not been achieved in m trials) $= q^m$.

$$P(\underbrace{X > m+n}_{E} \mid \underbrace{X > n}_{F}) = \frac{P(E \cap F)}{P(F)}$$

$$= \frac{P(E)}{P(F)} = \frac{P(X > m+n)}{P(X > n)}$$

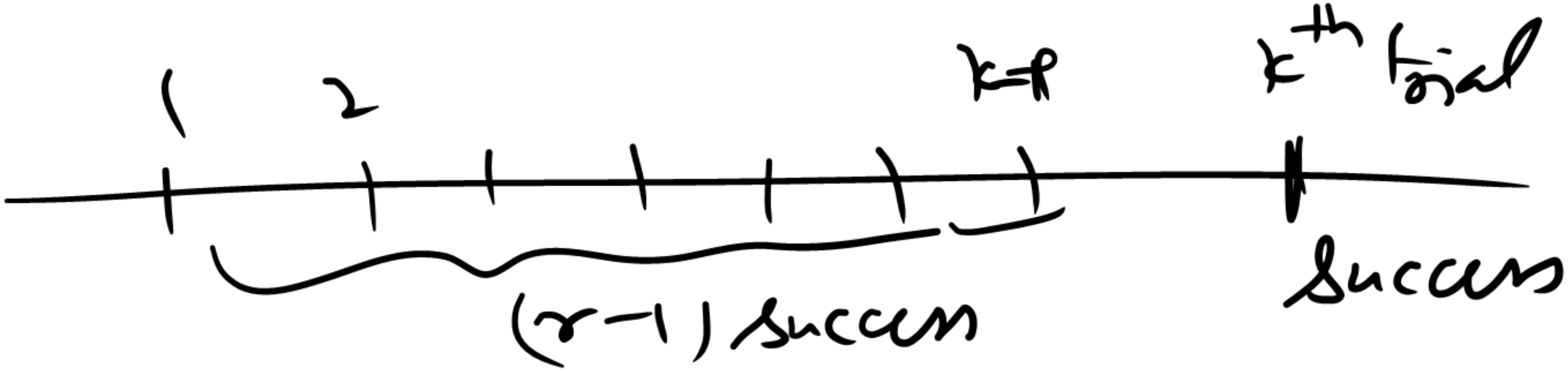
$$= q^{m+n}/q^{n} = q^{m} = P(X > m)$$

Suppose independent Bernoulli trials are performed under identical conditions (with prob of success p in each trial) until r^{th} success is achieved.

$X \to$ the no. of trials required

$X \to r, r+1, r+2, \ldots$

$$p_X(k) = P(X=k) = \binom{k-1}{r-1} p^{r-1} q^{k-r} \cdot p$$



$$= \binom{k-1}{r-1} q^{k-r} p^{r}, \quad k = r, r+1, \ldots$$

Negative Binomial / Inverse Bin. distn.

$$\sum_{k=r}^{\infty} \binom{k-1}{r-1} q^{k-r} p^{r} = p^{r} \cdot \frac{1}{(1-q)^{r}}$$

$$= 1$$

$$E(X) = \sum_{k=r}^{\infty} k \cdot \binom{k-1}{r-1} q^{k-r} p^{r}$$

$$= \frac{r}{p}\left\{ \underbrace{\sum_{k=r+1}^{\infty} \frac{k!}{r!\,(k-r)!}\, q^{k+1-\overline{r+1}}\, p^{r+1}}_{1} \right\}$$

$$= \frac{r}{p}$$

$$E(X^2) = E\,\underbrace{X(X-1)} + E(X) \qquad X$$

$$V(X) = \frac{rq}{p^2}$$

$$M_X(t) = E\left(e^{tX}\right)$$

$$= \sum_{k=r}^{\infty} e^{tk} \binom{k-1}{r-1} q^{k-r} p^{r}$$

$$= p^{r} e^{tr} \sum_{k=r}^{\infty} \binom{k-1}{r-1} \left(qe^{t}\right)^{k-r}$$

$$= \left(pe^{t}\right)^{r} \Big/ \left(1-qe^{t}\right)^{r}$$

$$= \left(\frac{pe^t}{1-qe^t}\right)^r, \quad qe^t < 1$$

$$\text{or } t < -\log_e q$$

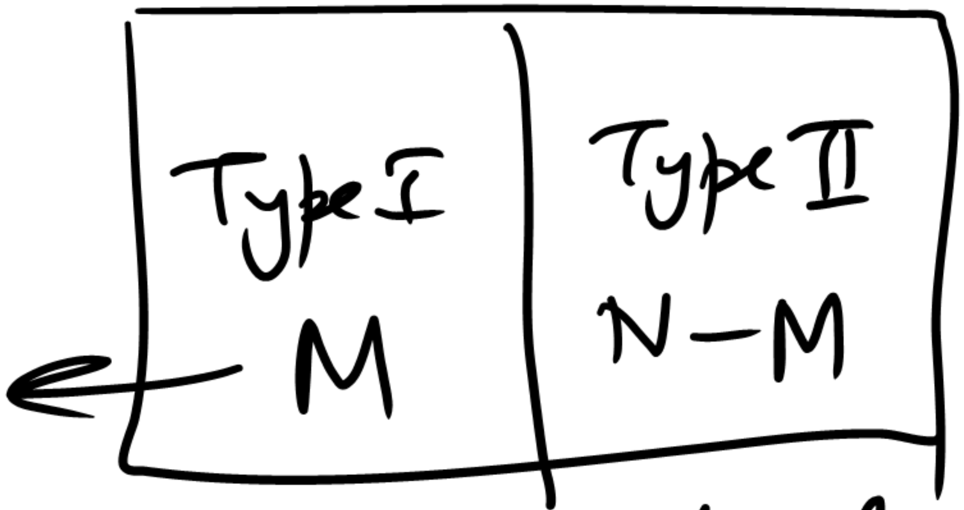
__Example__: Suppose a heavy machine has several independent components each with failure probability p. The machine fails if 4 components fail. In how many epochs the machine will stop functioning $\longrightarrow$ X

$X \to 4, 5, \ldots$

$$P(X=k) = \binom{k-1}{3} q^{k-4} p^4, \quad k = 4, 5, \ldots$$

Hypergeometric Distribution

Suppose a population has N items



Suppose we randomly select n items without replacement.

$X \rightarrow$ the no. of items of Type I in the sample

$X \rightarrow 0, 1, \ldots, n$

$$p_X(k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, \dots, n$$

s.t. $k \le M$, $\quad n-k \le N-M$

$$\sum_{k=0}^{n} \binom{M}{k}\binom{N-M}{n-k} = \binom{N}{n} \qquad \circledast$$

Consider the coefficient of x^n in

the expansion of $(1+x)^N = (1+x)^M (1+x)^{N-M}$

$$E(X) = \sum_{k=\cancel{0}\,1}^{n} \frac{k \cdot \binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} = \frac{Mn}{N}$$

$$E\,X(X-1) = \frac{M(M-1)\,n(n-1)}{N(N-1)}$$

$$E(X^2) = E\{X(X-1)\} + E(X) = \frac{nM\,(nM - n - M + N)}{N(N-1)}$$

$$V(X) = E(X^2) - (E(X))^2$$

$$= \frac{nM(N-n)(N-M)}{N^2(N-1)}$$

$$= \left(\frac{N-n}{N-1}\right) \frac{nM}{N} \left(1 - \frac{M}{N}\right)$$

Theorem: Let X have hypergeometric dist^n with parameters (N, M, n).

As $N \to \infty$, $M \to \infty$ $\ni$ $\frac{M}{N} \to p$,

the $P(X = x) \longrightarrow \binom{n}{x} p^x q^{n-x}$

ie hypergeometric can be approximated by a binomial distⁿ for large populations.

Proof ⊛

Hypergeometric distⁿ has wide applications

in estimating items of one category /
estimating population size etc. /

Suppose there is need for survey to estimate amount of a costly mineral in a mine area. Here

$$E(X) = \frac{nM}{N}$$

$$X \approx \frac{nM}{N} \Rightarrow M \approx \frac{NX}{n}$$

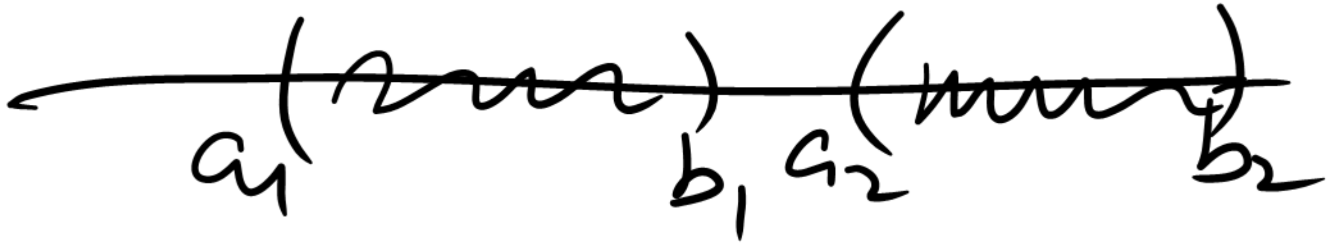
We can also estimate total population size of type I items are known

$$N \approx \frac{nM}{X}$$

<u>Poisson Process</u> : Events/ occursences/ happennings / observations over time/ area/ space are said to follow a Poisson process if they

satisfy the following assumptions:

1. The number of occurrences in disjoint time intervals are indept



2. The probability of a single occurrence in a small time interval is proportional to the length of interval.

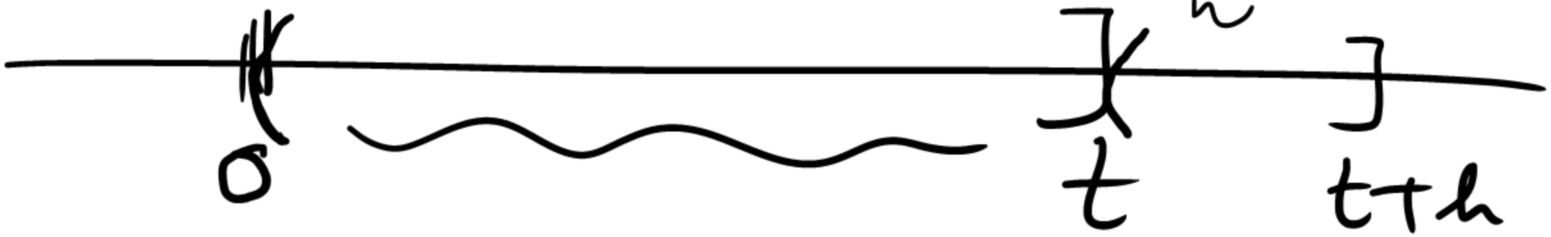
$$P_1(h) = \lambda h$$

3. The probability of more than one occurrence in a small time interval is negligible.

$$P_2(h)+P_3(h)+\cdots = o(h)$$

$$\Rightarrow 1-P_0(h)-P_1(h) = o(h)$$

$$\frac{o(h)}{h}\to 0$$



$X(t) \to$ number of occurrence in $(0,t]$

(in an interval of length t)

$$P(X(t)=n) = P_n(t)$$

$$= P(\text{ n occurrences in an interval of length } t)$$

Under assumptions (1) - (3), the distⁿ of $X(t)$ is given by

$$P_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \quad n = 0, 1, 2 \ldots \quad \ldots (1)$$

<u>Proof</u>: First we write assumptions (2) & (3) in a mathematical form

$$P_1(h) = \lambda h + \boxed{o(h)} \quad \cdots (2)$$

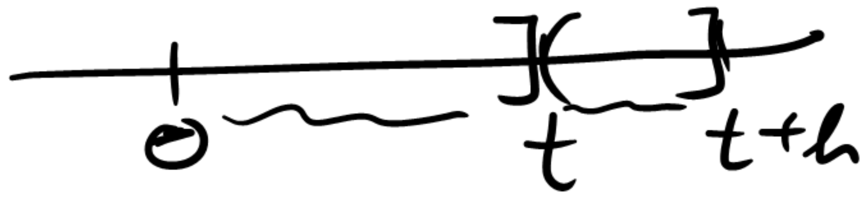
$$1 - P_0(h) - P_1(h) = o(h)$$

$$\Rightarrow \quad P_0(h) = 1 - \lambda h + o(h) \quad \cdots (3)$$

In order to prove (1), we use

the principle of induction.

$\boxed{n=0}$. Consider



$$P_0(t+h) = P(\text{no occurrence in } (0,t+h])$$

$$= P(\{\text{no occurrence in } (0,t]\} \cap \{\text{no occurrence in } (t,t+h]\})$$

$$= P(\text{no occurrence in } (0,t])\; P(\text{no occurrence in } (t,t+h])$$

as the no. of occursences in disjoint intervals are independent

$$= P_0(t)\, P_0(h)$$

$$= P_0(t)\left(1 - \lambda h + o(h)\right)$$

$$\Rightarrow \quad P_0(t+h) - P_0(t) = -\lambda h P_0(t) + o(h) P_0(t)$$

Divide by h on both sides and take

$h \to 0$,

$$P_0'(t) = -\lambda P_0(t)$$

This is first order ODE (variable separable form) with solution

$$P_0(t) = c\, e^{-\lambda t}$$

We can use initial condition

$P_0(0) = 1$, then we get $c = 1$

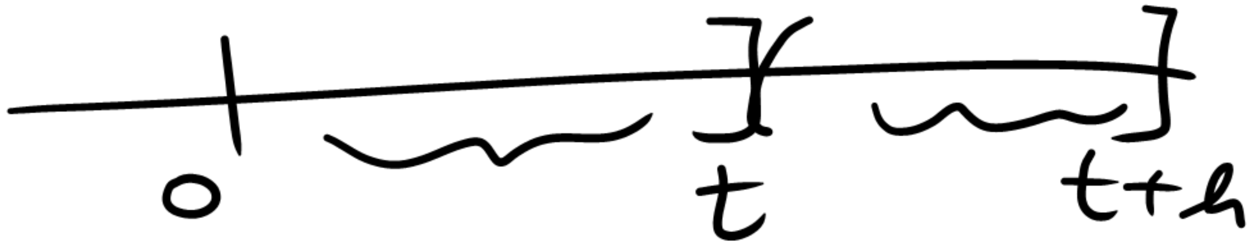
So the solution is $P_0(t) = e^{-\lambda t}$

So the statement (1) is proved for

$n = 0$.

Now we take $n = 1$. Consider

$P_1(t+h)$



$= P\Big(\text{one occurrence in } (0, t+h] \Big)$

$= P\Big(\{\text{one occurrence in } \underline{(0, t]}\} \cap$

$$\{\text{no occurrence in } (\underline{t, t+h}]\})$$

$$+P(\{\text{no occurrence in } (0,t]\} \cap \{\text{one occurrence in } (t,t+h]\})$$

$$= P(\text{one occurrence in } (0,t])\, P(\text{no occurrence in } (t,t+h])$$

$$+P(\text{no occurrence in } (0,t])\, P(\text{one occurrence in } (t,t+h])$$

$$= P_1(t)P_0(h) + P_0(t)P_1(h)$$

$$= P_1(t)\left(1-\lambda h + o(h)\right)$$

$$+ e^{-\lambda t}\left(\lambda h + o(h)\right)$$

$$\Rightarrow \frac{P_1(t+h) - P_1(t)}{h} = -\lambda P_1(t) + \lambda e^{-\lambda t} + \frac{o(h)}{h}\left(P_1(t) + e^{-\lambda t}\right)$$

Taking limit as $t \to 0$, we get

$$P_1'(t) = -\lambda P_1(t) + \lambda e^{-\lambda t}.$$

This is again a first order linear ODE and has solution

$$P_1(t) = \lambda t e^{-\lambda t} + c_1$$

Using initial condition $P_1(0) = 0$,

we get $c_1 = 0$. So

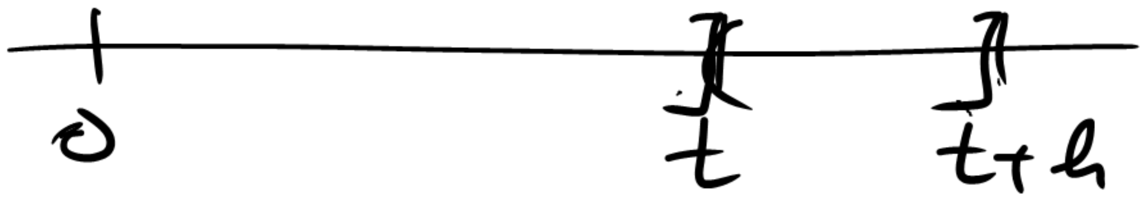
$$P_1(t) = \lambda t e^{-\lambda t}$$

So the statement (1) holds for

$$n = 1.$$

Next we assume that (1) holds for $n \le k$. Now consider

$$n = k+1.$$



$$P_{k+1}(t+h) = P\big((k+1) \text{ occurence in } (0, t+h]\big)$$

$$= P\Big(\{(k+1) \text{ occurrence in } (0,t]\} \cap \{\text{no occurrence in } (t,t+h]\}\Big)$$

$$+ P\Big(\{k \text{ occurs in } (0,t]\} \cap \{\text{one occur in } (t,t+h]\}\Big)$$

$$+ \sum_{j=1}^{k} P\Big(\{(k-j) \text{ occur in } (0,t]\} \cap \{(j+1) \text{ occur in } (t,t+h]\}\Big)$$

$$= P_{k+1}(t)\, P_0(h) + P_k(t)\, P_1(h)$$

$$+ \sum_{j=1}^{k} P_{k-j}(t)\, P_{j+1}(h)$$

$$= P_{k+1}(t)\left(1-\lambda h + o(h)\right)$$

$$+ \frac{e^{-\lambda t}(\lambda t)^k}{k!}\left(\lambda h + o(h)\right)$$

$$+ \quad o(h)\left(\sum_{j=1}^{k} P_{k-j}(t)\right)$$

$$\Rightarrow \frac{P_{k+1}(t+h) - P_{k+1}(t)}{h}$$

$$= -\lambda P_{k+1}(t) + \lambda \frac{e^{-\lambda t}(\lambda t)^{k}}{k!}$$

$$+ \frac{o(h)}{h} \left[\sum_{j=1}^{k+1} P_j(t) \right]$$

Taking limit as $h \to 0$, we get

$$P'_{k+1}(t) = -\lambda P_{k+1}(t) + \frac{\lambda^{k+1} t^k}{k!} e^{-\lambda t}$$

This is a linear first order ODE.

The general solution is

$$P_{k+1}(t) = \frac{e^{-\lambda t} (\lambda t)^{k+1}}{(k+1)!} + c_2$$

Taking initial conditions $P_{k+1}(0) = 0$, we get $c_2 = 0$. So the solⁿ is

$$P_{k+1}(t) = \frac{e^{-\lambda t} (\lambda t)^{k+1}}{(k+1)!}$$

So statement (1) holds for $n = k+1$.

So by the Principle of Mathematical Induction $P_n(t)$ holds for all $n = 0, 1, 2, \ldots$.