Probability Basic

Yanzhong (Eric) Huang.
2025

Ref: A first course in probability (ROSS).

1. Combinatorial Analysis

1.1 Permutations

distinct / different.

Suppose we have n objects, their are

$$n(n-1)\cdots 3\cdot 2\cdot 1 = n! \quad \text{permutations.}$$

When we have $n_1, \cdots, n_r$ are alike:

$$\frac{n!}{n_1!\, n_2! \cdots n_r!}$$

Example: PEPPER $\rightarrow n=6,\ n_1=3,\ n_2=2 \rightarrow \frac{6!}{3!\,2!}$

1.2 Combinations

From n objects pick r $\rightarrow$ the first r pemutations.

$$\binom{n}{r} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{(n-r)!\,r!}$$

$$* \binom{n}{0} = \binom{n}{n} = 1 = \frac{n!}{n!\,0!} = 1.$$

$$\binom{n}{i} = 0, \text{ if } i<0 \text{ or } i>n.$$

* Identities:

$$\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$$

proof: $$= \frac{(n-1)!}{(r-1)!(n-1-r+1)!} + \frac{(n-1)!}{r!(n-1-r)!}$$

$$= \frac{(n-1)!\,r}{(n-r)!\,r!} + \frac{(n-1)!\,(n-r)}{(n-r)!\,r!}$$

$$= \frac{n!}{(n-r)!\,r!} = \binom{n}{r}$$

* Binomial theorem

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

* Multinomial theorem.

$$\binom{n}{n_1, n_2 \cdots n_r} = \frac{n!}{n_1!\,n_2! \cdots n_k!}$$

$$(x_1 + x_2 + \cdots + x_r)^n = \sum_{\substack{(n_1, \cdots, n_r): \\ n_1 + \cdots + n_r = n}} \binom{n}{n_1, \cdots, n_r} x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r}.$$

known as multinomial coefficient

2. Axioms of Probability

Sample space $\rightarrow$ Ω or S. all possible outcome.

Event $\rightarrow$ E or A, subset of Ω.

Axioms:

1) $0 \le \mathbb{P}(A) \le 1$
2) $\mathbb{P}(\Omega) = 1$
3) For any sequence of mutually exclusive $A_1 A_2 \cdots$

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

2.1 conditional probability and independence

$$P(A \text{ given } B) = P(A|B) = \frac{P(AB)}{P(B)}$$

$*$ $P(A) = P(A|B)P(B) + P(A|B^c)\,P(B^c).$

Independence. $P(AB) = P(A)P(B).$

equivelent: $P(A|B) = P(B)$, $P(B|A) = P(A)$

2.2 Random Variables

The quantites of numerical experiment outcomes:

<u>cdf</u>: $F(x) = \mathbb{P}\{X \leq x\} = \mathbb{P}\{\text{outcome} \leq x\}$.

<u>pmf</u> (finite/<u>discrete</u>) → countable values

$$p(x) = \mathbb{P}\{X = x\} \rightarrow \sum_{i=1}^{\infty} p(x_i) = 1$$

<u>continuous</u> r.v. if exists a non-negative $f(x)$ (<u>pdf</u>).

$$P\{X \in C\} = \int_C f(x)\, dx.$$

$$\downarrow$$

$$F(x) = P\{X \in (-\infty, x]\} = \int_{-\infty}^{x} f(s)\, ds.$$

* Multivariable distribution.

$$F_{X,Y}(x,y) = P\{X \leq x, Y \leq y\} = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\, du\, dv$$

Independence: $X \perp Y$

$$P(X \in C, Y \in D) = P(X \in C)\, P(Y \in D)$$

$$\int_{C,D} f_{X,Y}(x,y)\, dx\, dy = \int_C f_X(x)\, dx \int_D f_Y(y)\, dy.$$

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y).$$

2.3 Expectation

discrete: $\mathbb{E}[X] = \sum_i x_i\, p(x_i) = \sum_i x_i\, P\{X = x_i\}.$

continuous: $\mathbb{E}[X] = \int_{-\infty}^{+\infty} x\, f(x)\, dx$

$$\mathbb{E}[g(X)] = \sum_i g(x_i)\, p(x_i).$$

or

$$= \int_{-\infty}^{+\infty} g(x) f(x)\, dx.$$

2.4 Variance

$$Var(X) = \mathbb{E}\left[(X-\mathbb{E}(X))^2\right]$$
$$= \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

$$Cov(X,Y) = \mathbb{E}\left[(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\right].$$
$$= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

$$Var(X+Y) = Var(X) + Var(Y) + 2cov(X,Y).$$

$$Corr(X,Y) = \frac{cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}}$$

2.5 Chebyshev's & Markov's Inequality

<u>Markov inequality</u>: If X is nonnegative, $\forall a>0$.

$$\mathbb{P}\{X \geq a\} \leq \mathbb{E}[X]/a$$

proof

Let $Y = \begin{cases} a, & \text{if } X \geq a \\ 0, & \text{if } X < a. \end{cases} \rightarrow X \geq Y$

$$\mathbb{E}[X] \geq \mathbb{E}[Y] = a\,\mathbb{P}\{X \geq a\}.$$

Corollary of Markov's inequality $\rightarrow$ Chebyshev's inequality

if X is a r.v. have mean μ and variance σ^2, for any $k > 0$,

$$P\{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2}$$

proof

$$Y = \frac{(X-\mu)^2}{\sigma^2} \rightarrow \mathbb{E}[Y] = \frac{\mathbb{E}[(X-\mu)^2]}{\sigma^2} = 1.$$

$$P\{Y \geq k^2\} \leq \frac{1}{k^2}$$

$$\Downarrow$$

$$P\{(X-\mu)^2 \geq k^2\sigma^2\} \leq \frac{1}{k^2}$$

$$\Downarrow$$

$$P\{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2}$$

Chebyshev's inequality could further $\rightarrow$ weak Law of Large Numbers

Theorem Weak LLNs

Let $X_1, X_2, \cdots$ be a sequence of i.i.d. r.v.

Then for any $\varepsilon > 0$,

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \varepsilon\right\} \to 0 \quad \text{as } n \to \infty$$

proof with extra assumption finite σ^2.

$$\mathbb{E}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \mu.$$

$$Var\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{1}{n^2}\left[Var(X_1) + \cdots\right] = \frac{\sigma^2}{n}.$$

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \frac{k\sigma}{\sqrt{n}}\right\} \leq \frac{1}{k^2}$$

for any $\varepsilon > 0$, let $\varepsilon = \frac{k\sigma}{\sqrt{n}}$. $\to \frac{\sqrt{n}\,\varepsilon}{\sigma} = k$.

$$P\{\cdots\} \leq \frac{\sigma^2}{n\varepsilon^2} \qquad n \to \infty,\ \frac{\sigma^2}{n\varepsilon^2} \to 0$$

Strong $\quad \lim_{n \to \infty} \frac{X_1 + \cdots + X_n}{n} = \mu.$

3. Some Discrete Distributions

3.1 Binomial

X the number of successes in n trials, with success propability p in each trial $X(n,p)$.

pmf: $P_i \equiv P\{X=i\} = \binom{n}{i} p^i (1-p)^{n-i}$, $i = 0, 1, \cdots n$

$$\Updownarrow$$

$$X = \sum_{i=1}^{n} X_i, \text{ where } X_i = \begin{cases} 1 & \text{if ith success.} \\ 0 & \end{cases}$$

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = n\mathbb{E}[X_i] = \underline{\underline{np}}.$$

$$\begin{aligned} Var[X] &= Var\left[\sum_{i=1}^{n} X_i\right] = n\, Var[X_i] \\ &= n\left[\mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2\right] \\ &= n(p-p^2) = \underline{\underline{n\,p(1-p)}} \end{aligned}$$

3.2 Poisson r.v.

A r.v. takes on one of the values $0, 1, 2, \cdots$ is said to be a Poisson r.v. with parameter λ, $\lambda > 0$.

parameter : λ

pmf : $p_i = P\{X = i\} = e^{-\lambda} \cdot \frac{\lambda^i}{i!}$, $i = 0, 1, \cdots$

It could apporximate numer of success in large number of trials. (could think as $n \to \infty$ in binomial, $\lambda = np$)

$$p_i = \binom{n}{i} p^i (1-p)^{n-i}$$

$$= \frac{n!}{(n-i)!\,i!} \left(\frac{\lambda}{n}\right)^i \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^i}$$

$$= \underbrace{\frac{n!}{(n-i)!\,n^i}}_{\to 1} \frac{\lambda^i}{i!} \frac{\overbrace{(1-\lambda/n)^n}^{\to e^{-\lambda}}}{\underbrace{(1-\lambda/n)^i}_{\to 1}}$$

$$\approx e^{-\lambda} \frac{\lambda^i}{i!}$$, for n is large, p is small

$E[X] = \lambda = np$ $\qquad$ $Var(X) = \lambda$

3.3 Geometric r.v.

The first successer number.

$$P\{X=n\} = (1-p)^{n-1}p \quad , n \geq 1.$$

Geometric r.v. with a parameter p.

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} (1-p)^{n-1} pn = \frac{1}{p}$$

$$\sum_{n=1}^{\infty} n x^{n-1} = \frac{d}{dx} \sum_{n=0}^{\infty} x^n \underset{=}{\overset{\text{when } x \in (0,1)}{=}} \qquad \frac{d}{dx}\frac{1}{(1-x)} = \frac{1}{(1-x)^2}$$

$$\sum_{n=0}^{\infty} x^2 = 1 + x + x^2 + \cdots = S$$

$$x + x^2 + \cdots = Sx$$

$$S - Sx = 1 \rightarrow S = \frac{1}{1-x}$$

$$Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

$$= \frac{1-p}{p^2}$$

3.4 The Negative Binomial

Number of trials if we had r successes.

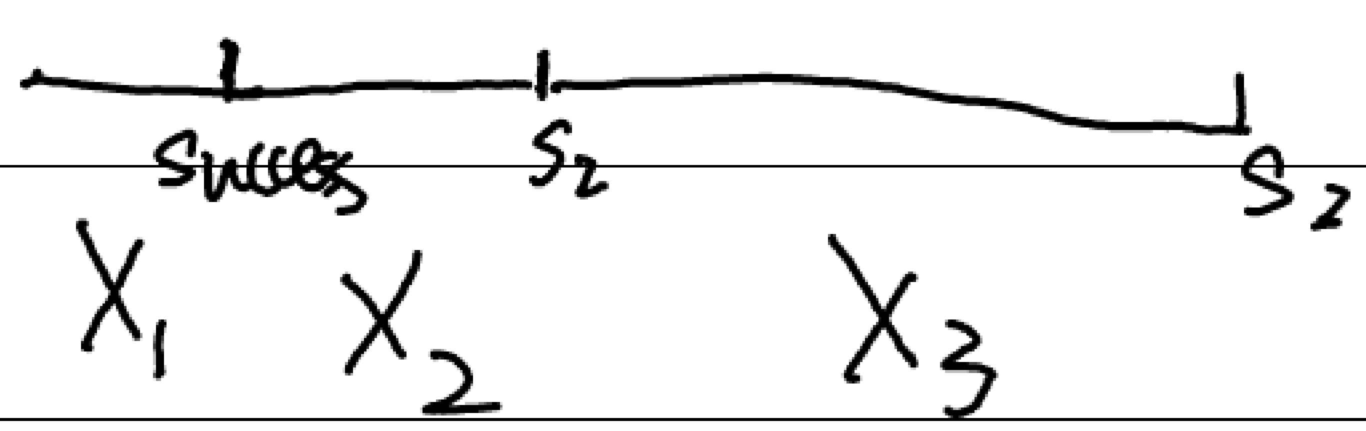
binomial: n trials, how many success

negative binomial: r success, how many trials

pmf: $P\{X=n\} = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad n \geq r$

Also called Pascal sometimes.

let X_i, $i=1,\cdots,r$ be the number of trials between $i-1$ to ith success.



$X_i \sim$ Geometric r.v.

$$X = \sum_{i=1}^{r} X_i \rightarrow \mathbb{E}[X] = r\,\mathbb{E}[X_i] = \frac{r}{p}$$

$$Var(X) = r\,Var(X_i) = \frac{(1-p)r}{p^2}$$

3.5 Hypergeometric r.v.

From $N+M$ balls, N - light color, M - dark color, randomly pick n balls, the light color number is X.

$$P\{X=i\} = \frac{\binom{N}{i}\binom{M}{n-i}}{\binom{N+M}{n}}$$

let $X_i = \begin{cases} 1 & \text{if ith is light} \\ 0 & \text{otherwise,} \end{cases}$ $\quad i=1,\cdots,n.$

$$X = \sum_{i=1}^{n} X_i.$$

$$E[X] = n\sum_{i=1}^{n} E[X_i] = \frac{nN}{N+M}$$

$$Var(X) = \frac{nNM}{(N+M)^2}\left(1 - \frac{n-1}{\underbrace{N+M-1}_{\text{covariance}}}\right)$$

4. Continuous r.v.

4.1 Uniform

pdf: $f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a<x<b \\ 0 & \text{otherwise} \end{cases}$

$$\mathbb{E}[X] = \int_a^b x f(x)\, dx = \frac{1}{b-a} \int_a^b x\, dx.$$

$$= \frac{1}{b-a} \left[\frac{1}{2} x^2 \Big|_a^b \right].$$

$$= \frac{(b+a)(b-a)}{2(b-a)} = \frac{b+a}{2}$$

$$\mathbb{E}[X^2] = \int_a^b x^2 f(x)\, dx = \frac{1}{b-a} \int_a^b x^2\, dx$$

$$= \frac{1}{b-a} \frac{1}{3} \left[b^3 - a^3 \right]$$

$$= \frac{(a^2+b^2+ab)(b-a)}{3(b-a)} = \frac{a^2+b^2+ab}{3}$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{a^2+b^2+ab}{3} - \frac{a^2+b^2+2ab}{4}$$

$$= \frac{a^2+b^2-2ab}{12} = \frac{(b-a)^2}{12}$$

$$F(x) = \int_a^x f(s)\, ds = \int_a^x \frac{1}{b-a}\, ds = \frac{x-a}{b-a}$$

4.2 Normal

$X \sim N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-(s-\mu)^2/2\sigma^2} ds.$$

$$= \Phi\left(\frac{x-\mu}{\sigma}\right)$$

standard. $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ $\quad X \sim Z(0,1)$.

$$F(x) = \int_{-\infty}^{x} f(s) ds = \Phi(x).$$

4.3 Exponential

$f(x) = \lambda e^{-\lambda x}$, $0 < x < \infty$, $\lambda > 0$.

$$F(x) = \int_0^x \lambda e^{-\lambda s} ds = -e^{-\lambda s}\Big|_0^x = 1 - e^{-\lambda x}$$

$\mathbb{E}[X] = \frac{1}{\lambda}$ $\quad Var(X) = \frac{1}{\lambda^2}$

* memoryless property $P\{X > s+t \mid X > s\} = P\{X > t\}$.

$$\Updownarrow$$

$$P\{X > s+t\} = P\{X > t\} P\{X > s\}$$

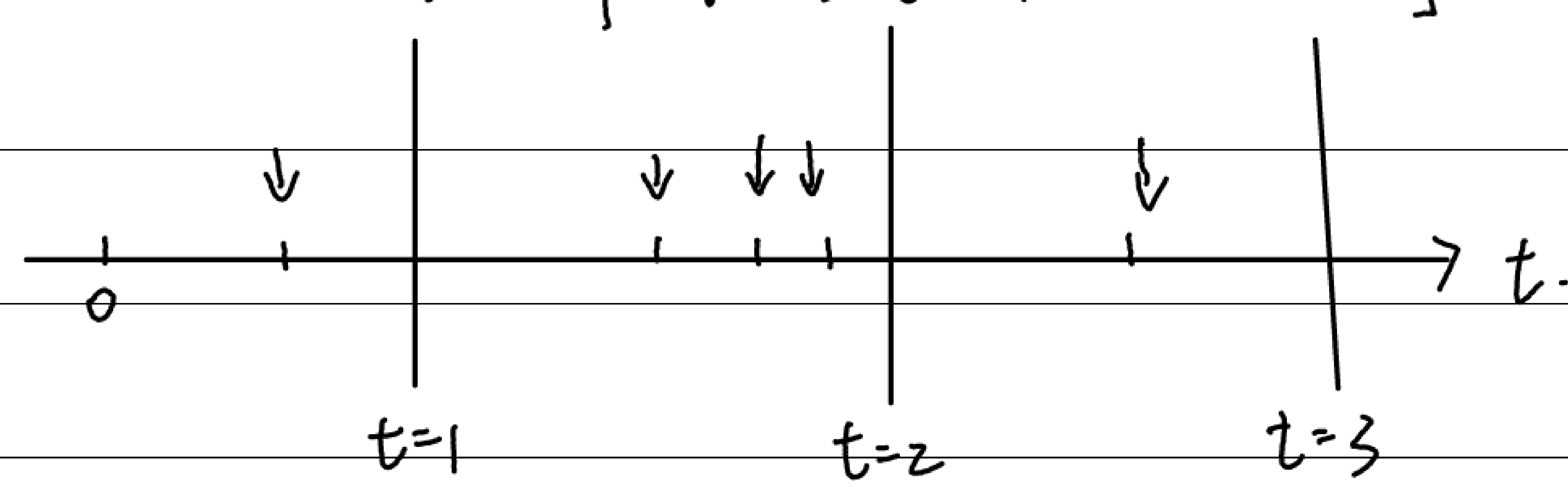
4.4 Compare geometry, poison, exponential

g	discrete		X = count of trials first success.
p	discrete	λ	X = how many events within time with rate λ
e	continuous	λ	X = time of a event happend.

4.5 Poisson process and Gamma r.v.

Suppose events are occuring at random time points.

$N(t)$ — the number of events occur in $[0, t]$



$N(0)=0$ $\quad N(1)=1$ $\quad N(2)=4$ $\quad N(3)=5$

$N(t)$ is a Poisson process having rate λ, $\lambda>0$

a) $N(0)=0$. b) Events are independent.

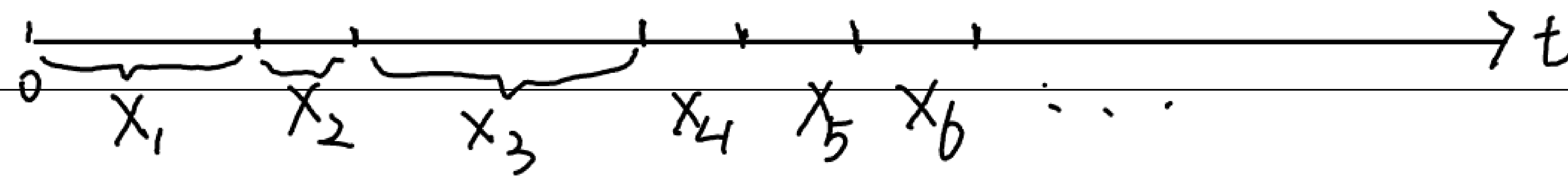
c) The distribution at $N(t)-N(s)$ depends on $t-s$

d) $\lim_{h\to 0} P\{N(h)=1\}/h=\lambda$.

e) $\lim_{h\to 0} P\{N(h)\geq 2\}/h=0$.

* $N(t) \sim \text{Poisson}(\lambda t)$.

<u>Propostion</u> The interarrival times $X_1, X_2, \ldots$ are independent and identically distributed exponential r.v. with λ.



$X \sim \text{Exponential}(\lambda)$ $\quad f(x) = \lambda e^{-\lambda x}$

$$F(x) = 1 - e^{-\lambda x}.$$

* $N(t) \sim \text{Poisson}(\lambda)$ $\quad p_i = e^{-\lambda} \dfrac{\lambda^i}{i!}$

<u>Gamma distribution</u>

def $\quad f(t) = \lambda e^{-\lambda t} \dfrac{(\lambda t)^{n-1}}{(n-1)!}, \quad t > 0,$

is a gamma r.v. with parameters (n, λ).

* Gamma distribution is the sum of n i.i.d. exponential λ.

$$S_n = \Sigma_{i=1}^{n} X_i.$$

$$F_n(t) = P\{S_n \le t\} = P\{N(t) \ge n\} = \sum_{j=n}^{\infty} e^{-\lambda} \frac{(\lambda t)^j}{j!}$$

4.6 Nonhomogeneous Poisson Process

From a modeling point of view the major weakness of Poisson process its assumption that events are just as likely to occur in all intervals of equal size.

nonhomogeneous Poisson Process.

a) $N(0) = 0$ b) Number of events in disjoint time are indep.

c) $\lim_{h \to 0} P\{\text{exactly } 1 \text{ event between } t \text{ and } t+h\}/h = \lambda(t)$.

d) $\lim_{h \to 0} P\{2 \text{ or more events } \ldots\}/h = 0$.

$$m(t) = \int_0^t \lambda(s)\, ds, \quad t \geq 0$$ is called mean-value

$N(t+s) - N(t) \sim$ Poisson r.v. $\lambda = m(t+s) - m(t)$

5. Conditional Expectation and Conditional Variance

If X and Y are jointly discrete r.v.s.

$$\mathbb{E}[X \mid Y=y] = \sum_x x P\{X=x \mid Y=y\}$$

$$= \frac{\sum_x x P\{X=x, Y=y\}}{P\{Y=y\}}$$

Continuous. $$\mathbb{E}[X \mid Y=y] = \frac{\int x f(x,y)\,dx}{\int f(x,y)\,dx}.$$

<u>Proposition</u> Tower properties.

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X]$$

$$\mathrm{Var}(X \mid Y) = \mathbb{E}[(X-\mathbb{E}[X \mid Y])^2 \mid Y]$$

$$= \mathbb{E}[X^2 \mid Y] - (\mathbb{E}[X \mid Y])^2$$

$$\Downarrow$$

$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(\mathbb{E}[X \mid Y]).$$