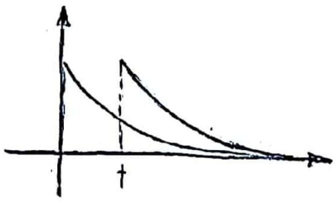
Exp is memoryless

$$\mathbb{P}(X \geq t+s) = e^{-\lambda(t+s)}$$

$$\mathbb{P}(X \geq t+s \mid X \geq t) = \mathbb{P}(X \geq s) = e^{-\lambda s}$$

$$\nparallel$$

$$\mathbb{P}(X \geq t+s)$$



⊛ موقعی می‌گیم دوتا R.V. از هم مستقل هستند که تمام event هایی که فقط با X به دست میان نسبت به تمام event هایی که از Y به دست میان مستقل باشن

Binomial

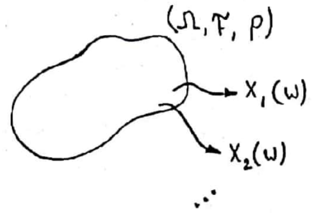
$X, Y \sim \text{Bern}(p), \text{Bern}(\tilde{p}) \qquad \tilde{p} = \frac{k}{n}$

$$p(\tilde{p}) \asymp \frac{e^{-n D_{KL}(\tilde{p} \| p)}}{\sqrt{2\pi n \tilde{p}(1-\tilde{p})}}$$

به ازای n بزرگ حول peak ش شبیه گاوسی میشه

$$D_{KL}(\tilde{p} \| p) \triangleq \tilde{p} \lg \frac{\tilde{p}}{p} + (1-\tilde{p}) \lg \frac{1-\tilde{p}}{1-p}$$

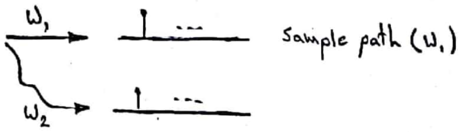
Geometric ≡ discrete version of Exp



$\{X_i(\omega) \mid i \in \mathbb{N}\}$ sample path : ω

Bernoulli process

$X_1, X_2 \ldots \overset{iid}{\sim} \text{Bern}(p)$



⊛ متغیر تصادفی رو به صورت $X(\omega)$ می‌نویسیم ولی فرایند تصادفی رو به صورت $X(\omega,t)$ می‌نویسیم و به ازای هر ω به جای یک عدد، یک sample path می‌گیریم

⊛ تمام randomness در ω عه و بعد از رخ دادن اون t_i ها به صورت deterministic خواهند بود. در حقیقت ω خیلی پیچیده است و هر بخشش یکی از t_i ها رو هندل می‌کنه ⟵ $\omega \equiv$ seed

⊛ فرایند تصادفی معادل R.V. های joint عه. فقط تعداد R.V. هاش بی‌نهایت عه

⊛ فرایند تصادفی یا باید طولش نامحدود باشه یا اگر محدوده باید طولش خودش تصادفی باشه. در غیر این صورت بهش می‌گن random vector

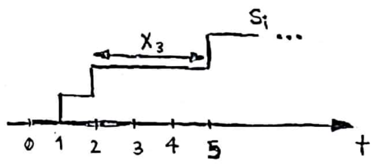
Session 4

1403.12.4

Bernoulli process

View 1 $\{Z_i \mid i \in \mathbb{N}\}$ $\quad Z_1 \cdots \overset{iid}{\sim} \text{Bern}(p)$

View 2 $\{S_i \mid i \in \mathbb{N}\}$ $\quad S_i = \sum_{j=1}^{i} Z_j \qquad S_1 \cdots \overset{iid}{\sim} \text{Binom}$



View 3 $\{X_i \mid i \in \mathbb{N}\}$ $\quad X_1 \cdots \overset{iid}{\sim} \text{Geom}(p)$

ویژگی ها

- stationary ← اگه از زمان t_1 به بعد رو نشون مون بدن، نمی‌فهمیم $(0, t_1)$ بریده شده بود
- time reversible ← اگر جهت زمان رو برعکس کنیم همچنان فرایند برنولی خواهد ماند
- memoryless ← منظور اینه که inter arrival (X_i) های فرایند memoryless هستند

Expectation

Moments

MGF

Concentration Inequalities

Review

Laws of Large Numbers

واریانس رو نیازی نداره

$X_1 \cdots X_n \overset{iid}{\sim} f_X(x) \qquad \mathbb{E}[X] = \mu$

$S_n \triangleq X_1 + \cdots + X_n$

weak LLN $\quad \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = 0$

strong LLN $\quad \mathbb{P}\left(\lim_{n\to\infty} \frac{S_n}{n} = \mu\right) = 1$

CLT

واریانس رو نیاز داره

$Z_n \triangleq \frac{S_n - n\mu}{\sigma\sqrt{n}}$

$\lim_{n\to\infty} Z_n \sim \mathcal{N}(0,1)$

I.D. convergence

$X \sim \mathcal{N}(0; \sigma^2)$

$\mathbb{E}[X^{2k}] = \frac{(2k)!}{k!\, 2^k} \sigma^{2k} \qquad k = 1, \dots$

memoryless بودن برای یک R.V. نه یک فرایند

X is memoryless if

$p(X \geq t+s) = p(X \geq t)\, p(X \geq s)$

$\equiv$

$p(X \geq t+s \mid X \geq s) = p(X \geq t)$

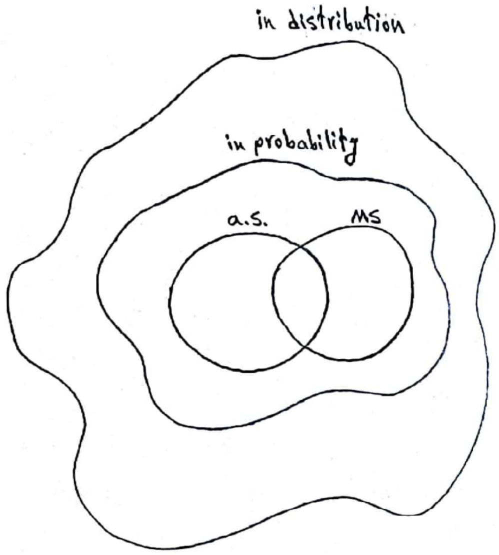
Convergence in Distribution

sequence $z_1 \dots$ converges in distribution to z^* if

$$\lim_{n\to\infty} F_{z_n}(z) = F_{z^*}(z) \quad \forall z \text{ for which } F_z(z) \text{ is continuous}$$

e.g. CLT $\longrightarrow \lim_{n\to\infty} \mathbb{P}(z_n < x) = \mathbb{P}(\mathcal{N}(0,1) \leq x)$

$$\forall x \in \mathbb{R}$$



Convergence in probability

Sequence $z_1 \dots$ converges in probability to z^* if

$$\lim_{n\to\infty} \mathbb{P}(|z_n - z^*| > \varepsilon) = 0 \qquad \forall \varepsilon > 0$$

e.g. weak LLN $\longrightarrow \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = 0 \qquad \forall \varepsilon > 0$

⊛ CLT اطلاعات بیشتری میده و میگه گاوسی هم هست ولی نوع convergence ش ضعیف تره

Convergence in mean square

sequence $z_1 \dots$ converges in mean square to z^* if

$$\lim_{n\to\infty} E\left[(z_n - z^*)^2\right] = 0$$

Convergence with probability 1 (a.s.)

sequence $z_1 \dots$ converges with probability 1 to z^* if

$$\mathbb{P}\left(\lim_{n\to\infty} z_n(\omega) = z^*(\omega) \,\middle|\, \omega \in \Omega\right) = 1$$

$$\equiv$$

$$\lim_{n\to\infty} z_n \overset{a.s.}{=\!=} z^*$$

e.g. strong LLN $\longrightarrow \mathbb{P}\left(\lim_{n\to\infty} \frac{S_n(\omega)}{n} = \mu\right) = 1$

$\boxed{\text{poisson process}} \subset$ point process , Arrival process

A renewal process over $\mathbb{R}$ where $X_1 \dots \underset{\text{iid}}{\sim} \lambda e^{-\lambda x} u(x)$ for some fixed $\lambda > 0$

λ → rate

پیوسته

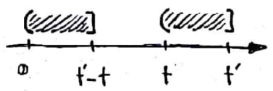
• Memoryless property

$$P(X_i > t+s \mid X_i > t) = P(X_i > s)$$

• stationary Increment property

A counting process $\{N(t) \mid t>0\}$ has this property if $\underbrace{N(t') - N(t)}$ has the same CDF as $N(t'-t)$ $\forall t' > t > 0$

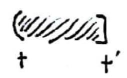
تعداد arrival های بازه $(t, t']$



⊛ هر دو تا بازه دلخواه، حتی اگر اشتراک داشته باشن، تعداد arrival هایی که توشون رخ میده، هم توزیع هست اگر طول بازه ها یکسان باشه

• Independent Increment property

$$\tilde{N}(t, t') \triangleq N(t') - N(t)$$



We have this property if for all $n \geq 2$, $0 < t_1 < \dots < t_n \in \mathbb{R}$

$N(t_1)$, $\tilde{N}(t_1, t_2), \dots, \tilde{N}(t_{n-1}, t_n)$ are independent

⊛ تعداد arrival های یک بازه، هیچ اطلاعاتی نسبت به تعداد arrival های بازه های دیگه نمیده
(بازه های بدون اشتراک)

راه های نشون دادن Arrival process ها :

• $\{X_i \mid i \in \mathbb{N}\}$ inter arrival periods

• $\{S_n \mid n \in \mathbb{N}\}$ epoch

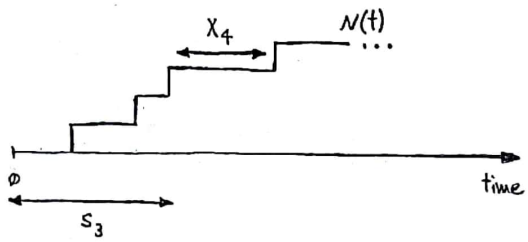
زمانی که n امین arrival اتفاق بیفتد

$$S_n = \sum_{i=1}^{n} X_i$$

• $\{N(t) \mid t > 0\}$ counting process

تعداد arrival ها تا زمان t

$$⊛ \ \{S_n \leq t\} \equiv \{N(t) \geq n\} \quad \forall n \in \mathbb{N},\ t > 0$$



Renewal process is a type of arrival process where inter arrival periods $X_1 \dots$ are iid

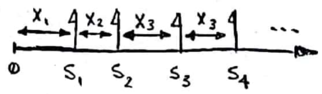
⊛ در فرایند های گسسته فقط برنولی memoryless ه، در فرایند های پیوسته فقط پواسون memoryless ه

⊛ فرایند پواسون هم time reversible ه

Session 6

1403.12.13

$\boxed{f_{s_n}(s_n)}$



$$S_n = \sum_{i=1}^{n} x_i \quad \nearrow \ Var(S_n) = \frac{n}{\lambda^2} \quad \searrow \ \mathbb{E}[S_n] = \frac{n}{\lambda}$$

$$S_1 = X_1 \sim \lambda e^{-\lambda x} u(x)$$

$$S_2 = X_1 + X_2 \sim (f_X * f_X)(x) = \int_{-\infty}^{\infty} \lambda e^{-\lambda t} u(t)\, \lambda e^{-\lambda(x-t)} u(x-t)\, dt$$

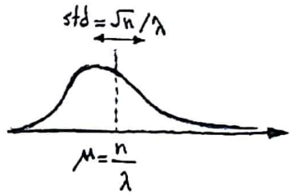
$$= \lambda^2 e^{-\lambda x} \underbrace{\int_{-\infty}^{\infty} u(t)\, u(x-t)\, dt}_{\int_0^x dt} = \lambda^2 x e^{-\lambda x} u(x)$$

⋮

$$S_n = X_1 + \cdots + X_n \sim \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x} u(x)$$

Erlang's distribution

جمع n تا نمایی مستقل با پارامتر λ



$\boxed{f_{s_1 \cdots s_n}(s_1, \cdots, s_n)}$ $\quad 0 < s_1 < \cdots < s_n$

$$f_{s_1}(s_1) = \lambda e^{-\lambda s_1} u(s_1)$$

$$f_{s_1 s_2}(s_1, s_2) = \underbrace{f_{s_1}(s_1)}_{\lambda e^{-\lambda s_1} u(s_1)} \underbrace{f_{s_2|s_1}(s_2|s_1)}_{\lambda e^{-\lambda(s_2-s_1)} u(s_2-s_1)} = \lambda^2 e^{-\lambda s_2} u(s_1) u(s_2-s_1)$$

memoryless property

⊛ اگر s_2 رو بدونیم توزیع s_1 توزیع یکنواخت بین $(0, s_2)$ میشه؟

⊛ اگر s_1 رو بدونیم توزیع s_2 نمایی میشه؟

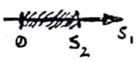
⋮

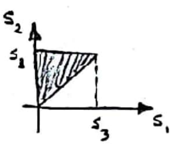
$$f_{s_1 \cdots s_n}(s_1 \cdots s_n) = \lambda^n e^{-\lambda s_n} \mathbb{1}(0 < s_1 < \cdots < s_n)$$

$$f_{s_n}(s_n) = \int_0^{s_n} \int_0^{s_{n-1}} \cdots \int_0^{s_2} \lambda^n e^{-\lambda s_n} \, ds_1 \cdots s_{n-1} = \lambda^n e^{-\lambda s_n} \frac{s_n^{n-1}}{(n-1)!}$$

s_{n-1} $\quad s_{n-2}$ $\quad s_1$

⋮

area $\quad \int_0^{s_2} ds_1 =$  $= s_2$

$\int_0^{s_3} \int_0^{s_2} ds_1 ds_2 =$  $= \frac{s_3^2}{2}$

$$f_{s_1 \cdots s_{n-1} | s_n}(s_1 \cdots s_{n-1} | s_n) = \frac{\lambda^n e^{-\lambda s_n} \mathbb{1}(0 < s_1 < \cdots < s_n)}{\frac{\lambda^n s_n^{n-1} e^{-\lambda s_n}}{(n-1)!}} = \frac{(n-1)!}{s_n^{n-1}} \mathbb{1}(0 < s_1 < \cdots < s_n)$$

uniform $(0, s_n)$

$$f_{s_n | s_1 \cdots s_{n-1}}(s_n | s_1 \cdots s_{n-1}) = \lambda e^{-\lambda(s_n - s_{n-1})} u(s_n - s_{n-1})$$

exponential

⊛ فرایند پواسون ⟵ در هر بازه (borel set بازه ها) به طول t ، تعداد arrival ها از توزیع پواسون با پارامتر λt پیروی می‌کند

یک راه دیگه برای ساختن فرایند پواسون اینه که $S_n \sim$ Erlang رو سمپل بگیریم
بعدش n-1 بار $S_{n-i} \sim$ uniform ها رو سمپل بگیریم و sort شون کنیم

تعریف 1 فرایند پواسون constructive

فرایند renewal ای که inter arrival period هاش $Exp(\lambda)$ باشند

تعریف 2 فرایند پواسون implicit

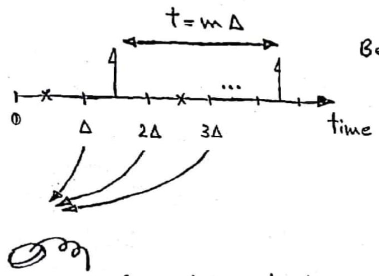
فرایندی که ویژگی های stationary increment و independent increment را داشته باشد
و تعداد arrival هاش در هر بازه به طول t ، از توزیع پواسون با پارامتر λt پیروی کند

تعریف 3 فرایند پواسون

limit of a shrinking Bernoulli process

session 7
1403.12.18

$\boxed{\text{Def 3} \longrightarrow \text{Def 1}}$



Bernoulli process

$$p = \lambda\Delta \quad \left(\Delta < \frac{1}{\lambda}\right)$$

$$M(m) \sim \text{Geo}(\lambda\Delta)$$

$$X(t) = \Delta \times M(m)$$

باید اثبات کنیم Geo به Exp میل می‌کند:

$$M(m) \sim (\lambda\Delta)(1-\lambda\Delta)^{m-1}$$

$$\mathbb{P}(t < X_i < t+\Delta) = (\lambda\Delta)(1-\lambda\Delta)^{\frac{t}{\Delta}-1}$$

$$\rightarrow \underbrace{\lim_{\Delta\to 0} \frac{\mathbb{P}(t<X<t+\Delta)}{\Delta}}_{f_{X_1}(t)} = \underbrace{\lim_{\Delta\to 0} \lambda(1-\lambda\Delta)^{\frac{t}{\Delta}-1}}_{\lambda e^{-\lim_{\Delta\to 0}(\lambda\Delta)(\frac{t}{\Delta}-1)}}$$

$$\rightarrow f_{X_i}(t) = \lambda e^{-\lambda t} u(t)$$

$\boxed{\text{Def 3} \longrightarrow \text{Def 2}}$

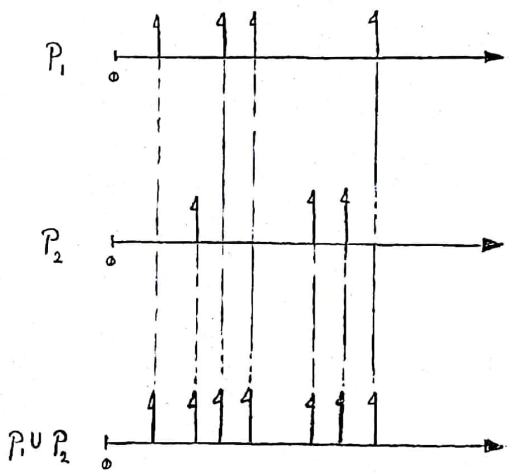
⊛ می‌دانیم که فرایند برنولی خواص stationary increment و independent increment را دارد

اگر $\lfloor \frac{t}{\Delta} \rfloor$ تا $\text{Bern}(\lambda\Delta)$ را باهم جمع کنیم به $\text{Binom}(\lfloor \frac{t}{\Delta} \rfloor, \lambda\Delta)$ می‌رسیم

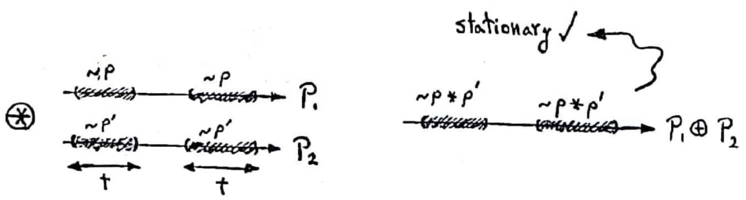
$$\lim_{\Delta\to 0} \text{Binom}\left(\lfloor \tfrac{t}{\Delta} \rfloor, \lambda\Delta\right) = \text{poisson}(\lambda t)$$

$\boxed{\text{poisson process Merging}}$

اگر فرایند های $P_1(\lambda_1)$ و $P_2(\lambda_2)$ پواسون و مستقل باشند، فرایند $P_1 \oplus P_2$ یک فرایند پواسون با نرخ $\lambda_1+\lambda_2$ خواهد بود



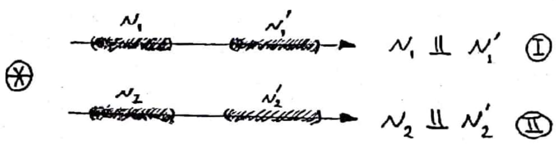
$\boxed{\text{اثبات با Def2}}$



$$⊛ \quad P(N_1(t)+N_2(t)=k) = \sum_{i=0}^{k} \frac{(\lambda_1 t)^i}{i!} e^{-\lambda_1 t} \frac{(\lambda_2 t)^{k-i}}{(k-i)!} e^{-\lambda_2 t}$$

$$= e^{-(\lambda_1+\lambda_2)t} \frac{t^k}{k!} \underbrace{\sum_{i=0}^{k} \underbrace{\frac{\lambda_1^i \lambda_2^{k-i}}{i!\,(k-i)!} k!}_{\binom{k}{i}\lambda_1^i \lambda_2^{k-i}}}_{(\lambda_1+\lambda_2)^k}$$

$$= \frac{((\lambda_1+\lambda_2)t)^k}{k!} e^{-(\lambda_1+\lambda_2)t} \rightsquigarrow \text{poisson} ✓$$



$P_1 \perp\!\!\!\perp P_2 \longrightarrow N_1 \perp\!\!\!\perp N_2 \,,\, N_1' \perp\!\!\!\perp N_2'$ Ⅲ

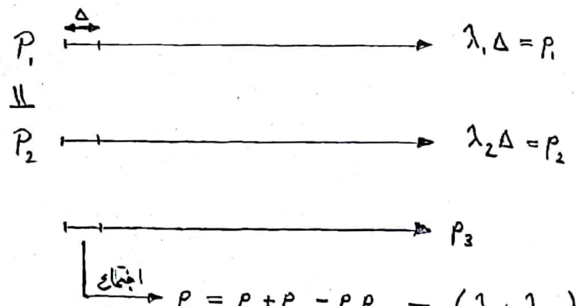
$\xrightarrow[\text{Ⅲ}]{\text{Ⅰ Ⅱ}} N_1 + N_2 \perp\!\!\!\perp N_1' + N_2' \rightsquigarrow$ independent ✓

Session 8

1403.12.20

poisson process merging

اثبات با Def 3



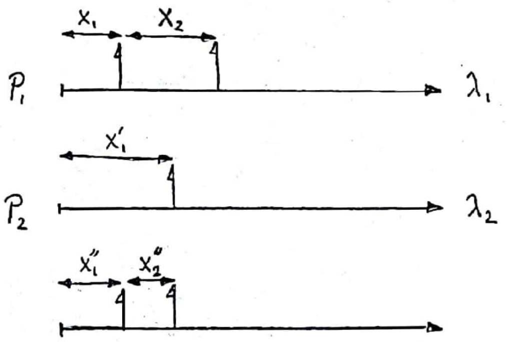
P_1 ، P_2 رو به صورت فرایند های برنولی با پارامتر $\lambda_i\Delta$ در نظر می‌گیریم.
و $P_1 \oplus P_2$ رو به صورت اجتماع برنولی های هر bin در نظر می‌گیریم یعنی
اگر جفت P_1 ، P_2 رخ داده باشند ، یکی را در خروجی لحاظ می‌کنیم.

$$p_3 = p_1 + p_2 - p_1p_2 = (\lambda_1+\lambda_2)\Delta + \lambda_1\lambda_2\Delta^2 \xrightarrow{\Delta\to 0} Bern\left((\lambda_1+\lambda_2)\Delta\right) \checkmark$$

همچنین واضح است که $(\lambda_1+\lambda_2)\Delta + O(\Delta^2) \to 0$ ، $\lfloor \frac{t}{\Delta} \rfloor \to \infty$ بنابراین به پواسون با پارامتر $(\lambda_1+\lambda_2)t$ می‌رسیم

⊛ احتمال رخ دادن هم زمان P_1 و P_2 به 0 میل می‌کند و واضح است که برنولی های هر bin از هم مستقل اند

اثبات با Def 1



$$X_1'' = \min\{X_1, X_1'\} \qquad X_1 \perp\!\!\!\perp X_1' , \quad X_1 \sim \lambda_1 e^{-\lambda_1 x}, \quad X_1' \sim \lambda_2 e^{-\lambda_2 x}$$

$$\mathbb{P}(X_1'' > x) \overset{\perp\!\!\!\perp}{=} \mathbb{P}(X_1 > x)\,\mathbb{P}(X_1' > x) = e^{-(\lambda_1+\lambda_2)x}$$

$$\longrightarrow X_1'' \sim Exp(\lambda_1+\lambda_2)$$

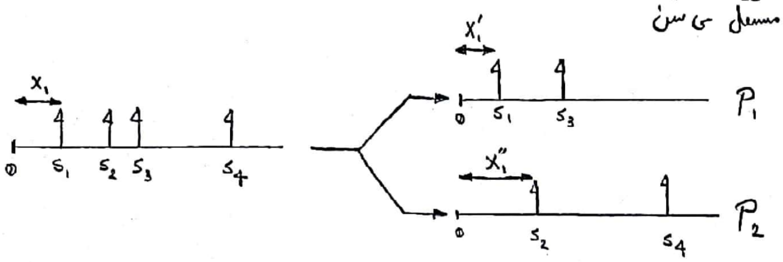
همچنین برای X_2'' نیز واضح است که به طور مشابه برابر min دو نمایی مستقل با پارامتر λ_1 ، λ_2 است زیرا فرایند P_1 از X_1 به بعد طبق تعریف یک نمایی با پارامتر λ_1 است و فرایند P_2 نیز به دلیل memoryless بودن از X_1 تا X_1' یک نمایی با پارامتر λ_2 است

$$X_2'' = \min\{X_2, X_1' \mid x > X_1\} \sim Exp(\lambda_1+\lambda_2)$$

به همین صورت تمام inter arrival period ها به صورت $X_i'' \overset{iid}{\sim} Exp(\lambda_1+\lambda_2)$ خواهند بود ✓

$\boxed{\text{poisson process splitting}}$ AKA branching
AKA forking

مثلاً اگر انتخاب یکی در میان باشد و s_1 ، s_3 از فرایند اول رو داشته باشیم ، می‌دونیم s_2 فرایند دوم بین این 2 تا رخ خواهد داد



• اگر split مون به صورت deterministic باشه ، نه شاخه ها مستقل می‌شن نه هر شاخه فرایند پواسون میشه

زیرا X_1' و X_1'' ها جمع تعدادی Exp می‌شن که دیگه Exp نخواهد بود (Erlang میشه)

• اگر split مون به صورت random ، با احتمال p هر پیچم بره به شاخه اول و به احتمال 1-p بره به شاخه دوم ، اون وقت هم شاخه ها مستقل می‌شن هم هر شاخه فرایند پواسون میشه (شاخه‌ی اول ، پواسون با پارامتر λp و شاخه‌ی دوم ، پواسون با پارامتر $\lambda(1-p)$)

اثبات با Def 1 ← اگر Geo تا Exp رو باهم جمع کنیم همچنان Exp می‌مونه

$$\left.\begin{aligned} & X_1' = \sum_{i=1}^{k} X_i \\ & X_i \overset{iid}{\sim} \text{Exp} \\ & k \sim \text{Geo} \end{aligned}\right\} \longrightarrow X_1' \sim \text{Exp}$$

اثبات با Def 3 ← هر bin شاخه‌ی اول یک $\text{Bern}(\lambda\Delta p)$ و هر bin شاخه‌ی دوم یک $\text{Bern}(\lambda\Delta(1-p))$ میشه

اثبات با Def 2 ← گویا اگر اثبات کنیم $P(N_1(t)=m, N_2(t)=k) = P(N_1(t)=m)\, P(N_2(t)=k)$ کافیه !

$$P(N_1=m, N_2=k) = \underbrace{P(N_1=m, N_2=k \mid N_{mother}=m+k)}_{\frac{(m+k)!}{m!\,k!} p^m (1-p)^k} \underbrace{P(N_{mother}=m+k)}_{\frac{(\lambda t)^{m+k}}{(m+k)!} \underbrace{e^{-\lambda t}}_{e^{-\lambda p t} - \lambda(1-p)t}}$$

$$= \underbrace{\frac{(\lambda p t)^m}{m!} e^{-\lambda p t}}_{\substack{P(N_1(t)=m) \\ \text{poisson}(\lambda p t)}} \times \underbrace{\frac{(\lambda(1-p)t)^k}{k!} e^{-\lambda(1-p)t}}_{\substack{P(N_2(t)=k) \\ \text{poisson}(\lambda(1-p)t)}}$$

Session 9

1403.12.25

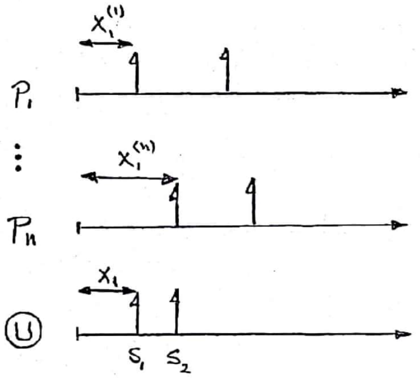
$\boxed{\text{Importance of poisson process}}$

if $P_1 \cdots P_n$ are independent ~~renewal~~ arrival processes, not necessarily from the same distribution,

then $\lim\limits_{\substack{n \to \infty \\ rate_i \to 0}} P_1 \oplus \cdots \oplus P_n$ is a poisson process $\equiv$ CLT

$(n \times rate_i = \lambda)$

$\boxed{\text{proof}}$



$$S_1 = \min\{X_1^{(1)} \cdots X_1^{(n)}\}$$

$$\xrightarrow{\cdots} F_{S_1}(s_1) = 1 - \prod_{i=1}^{n}\left(1 - F_i(s_1)\right)$$

$$\xrightarrow{\frac{d}{ds_1}} f_{S_1}(s_1) = \sum_{i=1}^{n} f_i(s_1) \prod_{\substack{j=1 \\ j \neq i}}^{n} \left(1 - F_j(s_1)\right)$$

اثبات دقیق نیست و شهودی محسوب می‌شه!

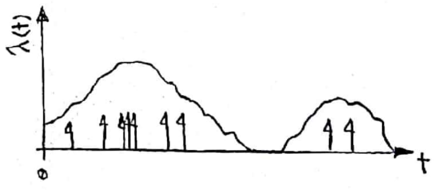
حرف نادقیق عه ولی چون s_1 کوچک هست با سری تیلور f_{S_1} رو تقریب می‌زنیم

$$f_{S_1}(s_1) \simeq \sum_{i=1}^{n} f_i(0) \prod_{\substack{j=1 \\ j \neq i}}^{n} \underbrace{(1 - \overbrace{f_j(0) s_1}^{\text{کوچک}})}_{e^{-f_j(0) s_1}} = \sum_{i=1}^{n} f_i(0)\, e^{-s_1 \sum_{\substack{j=1 \\ j \neq i}}^{n} f_j(0)}$$

$$= e^{-s_1 \sum_{j=1}^{n} f_j(0)} \sum_{i=1}^{n} f_i(0)\, \underbrace{e^{\overbrace{f_i(0) s_1}^{0}}}_{1} \quad \overset{\lambda \triangleq \sum_{i=1}^{n} f_i(0)}{=\!=\!=} \quad \lambda e^{-\lambda s_1} u(s_1)$$

نمایی شد ✓

$\boxed{\text{Non-homogeneous poisson process}}$



چیزی که تا الآن دیدیم پواسون همگن بود

قبلاً طبق Def3 می‌گفتیم احتمال هر شیر اومدن $\lambda\Delta$ عه. الآن می‌گیم $\lambda(t)\Delta$ عه.

⊛ چون λ ثابت نیست X_i ها هم نمایی نیستن و بنابراین memoryless نیست همچنین خاصیت stationary increment رو هم نداره

⊛ تعداد arrival ها در هر borel set همچنان از توزیع پواسون میاد و خاصیت independent increment رو هم داره و $N(t)$ ها هم مستقل اند

$$\mathbb{P}\left(N(t) = k\right) = \frac{\Lambda^k}{k!} e^{-\Lambda}$$

$$\Lambda = \int_0^t \lambda(t')\, dt'$$

$\boxed{f_{X_1}}$ (extra)

$X_1 \sim$ Geo with variable ps

$$p_m \prod_{i=1}^{m-1} (1 - p_i)$$

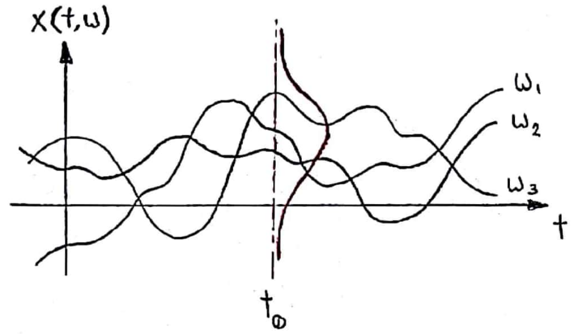
$$f_{X_1}(t) = \lambda(t)\Delta \prod_{i=1}^{t/\Delta} \left(1 - \lambda(i\Delta)\Delta\right)$$

$$= \lambda(t)\Delta\, e^{-\sum_{i=1}^{t/\Delta} \lambda(i\Delta)\Delta} \overset{\text{نمایی}}{=\!=} \lambda(t) e^{-\int_0^t \lambda(t')\, dt'}$$

$\boxed{\text{Gaussian process}} \not\subset$ arrival process

Session 10

1404.1.16



$$\{X(\underbrace{t}_{\text{time}},\underbrace{\omega}_{\text{ensemble}}) \mid t \geq 0\}$$

⊛ به ازای هر برش زمانی مثل t_0 یک متغیر تصادفی گاوسی داریم

⊛ فرایند های پواسون با rate شون مشخص می‌شدن
فرایند های گاوسی با kernel شون مشخص می‌شن

⊛ کاربرد ⟵ سیستم های LTI , control theory , regression , NTK

review

MGF

$$M_X(r) = \mathbb{E}[e^{rx}] = \int_{-\infty}^{\infty} e^{rx} f_X(x)\,dx \equiv \begin{bmatrix} e^{r_i x_j} \end{bmatrix}_{r\times x} @ \begin{bmatrix} f_X(x) \end{bmatrix}_x$$

⊛ اگر $r=-i\omega$ به تبدیل فوریه می‌رسیم

inverse: $f_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} M_X(r)\, e^{-rx}\, dr$

$$X \sim \mathcal{N}(\mu,\sigma^2) \longrightarrow M_X(r) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} \underbrace{e^{\frac{-(x-\mu)^2}{2\sigma^2}}\; e^{rx}}_{\exp\left(\frac{-x^2-\mu^2+2x\mu+2r\sigma^2 x}{2\sigma^2}\right)} dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(\frac{-(x-\mu-r\sigma^2)^2}{2\sigma^2}\right) \exp\left(\frac{r^2\sigma^4 + 2\mu r\sigma^2}{2\sigma^2}\right) dx$$

$$= e^{r\mu + \frac{1}{2}r^2\sigma^2}$$

review

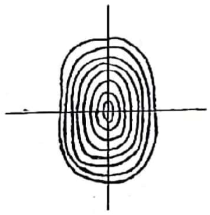
Jointly Gaussian

$$X, Y \sim J.G(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$$

$$f_{XY}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}\,\sigma_x\sigma_y} \exp\left[\frac{-1}{2(1-\rho^2)}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right)\right]$$

$$-1 \leq \rho \leq 1$$

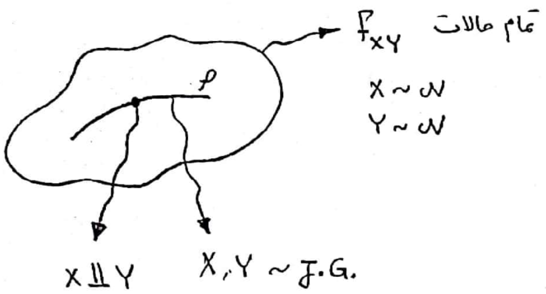
⊛ اگر $\rho \to 1$ اون وقت $f_{J.G} \simeq \frac{c}{\sqrt{1-\rho^2}} e^{\frac{-(x-y)^2}{2(1-\rho^2)}}$ که فقط جاهایی که $x=y$ باشه f_{XY} زیاد میشه و بقیه جاها تقریبا ≈ 0



⊛ ضریب همبستگی فقط ارتباط خطی بین X , Y رو capture می‌کنه و از استقلال خیلی ضعیف تره. ولی با توجه به اینکه J.G. ها با μ , Σ کامل مشخص می‌شن همین ارتباط خطی کافیه و به طور خاص برای J.G. داریم: $\rho=0 \longleftrightarrow$ استقلال

⊛ برای برقراری استقلال باید داشته باشیم:

$$\forall k,l \in \mathbb{N} \quad \mathbb{E}\left[(X-\mu_X)^k (Y-\mu_Y)^l\right] = \mathbb{E}\left[(X-\mu_X)^k\right] \mathbb{E}\left[(Y-\mu_Y)^l\right]$$



یک مثال که X , Y گاوسی باشند اما f_{XY} J.G. نباشد:

$X \sim N(0,\sigma^2)$ و بعد به طور مستقل سکه می‌اندازیم. اگر شیر آمد $Y=X$ و در غیر این صورت $Y=-X$.

$$f_{Y|X}(y|x) = \frac{1}{2}\delta(y-x) + \frac{1}{2}\delta(y+x)$$

$$\text{cov}(X,Y) = \mathbb{E}[XY] = \frac{1}{2}\mathbb{E}[XX] + \frac{1}{2}\mathbb{E}[X(-X)]$$
$$= 0$$

با اینکه $\rho=0$ ولی مستقل نیستن. چون اصلا J.G نبودن

$$\rho = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\ \text{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

دلیل اینکه $-1 \leq \rho \leq 1$ نامساوی کوشی شوارتزه:

$$\langle u,v\rangle^2 \leq \langle u,u\rangle \cdot \langle v,v\rangle$$
$$\Downarrow$$
$$\text{cov}(X,Y)^2 \leq \text{cov}(X,X) \cdot \text{cov}(Y,Y)$$

session 11

1404.1.18

$\boxed{\text{Covariance matrix}}$

$$\Sigma = \begin{bmatrix} Var(X) & cov(X,Y) \\ cov(X,Y) & Var(Y) \end{bmatrix} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix} \in \mathbb{R}^2$$

$$\det(\Sigma) = \sigma_X^2\sigma_Y^2(1-\rho^2)$$

$$\Sigma^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix} \frac{1}{\sigma_X^2} & \frac{-\rho}{\sigma_X\sigma_Y} \\ \frac{-\rho}{\sigma_X\sigma_Y} & \frac{1}{\sigma_Y^2} \end{bmatrix}$$

⊛ اگر $-1 < \rho < 1$ آنگاه $\det(\Sigma) \neq 0$ و فرم ماتریسی رو بدون مشکل داریم ولی اگر ± 1 بشه هم باز با رفع ابهام درست می‌شه

$$\xrightarrow[\text{zero mean دو بعدی 2}]{\text{فرم ماتریسی J.G}} f_{\underline{X}}(x) = f_{XY}(x,y) = \frac{1}{2\pi\sqrt{\det(\Sigma)}}\exp\left(-\frac{1}{2}\underline{X}^T\Sigma^{-1}\underline{X}\right)$$

$$\underline{X} = \begin{bmatrix} X \\ Y \end{bmatrix}$$

⊛ بردار های J.G معادل برداری است که المان هاش به صورت مستقل گاوسی باشند و بعد روی این بردار تبدیل خطی اعمال شه:

$$\underline{u} = \begin{bmatrix} u_1 \sim \mathcal{N}(0,1) \\ u_2 \sim \mathcal{N}(0,1) \end{bmatrix} \quad u_1 \perp\!\!\!\perp u_2 \quad , \quad {\Sigma^{\frac{1}{2}}}^T\Sigma^{\frac{1}{2}} = \Sigma$$

- Σ همواره PSD عه پس طبق eig val decomp به طور یکتا $\Sigma^{\frac{1}{2}}$ رو داریم
- $Var(\text{vector}) \equiv \det(\text{cov matrix})$

$$f_{\underline{u}}(\underline{u}) = f_{u_1}(u_1) \times f_{u_2}(u_2) = \frac{1}{2\pi}e^{\frac{-u_1^2-u_2^2}{2}} = \frac{1}{2\pi}e^{\frac{-\|u\|_2^2}{2}}$$

$$\underline{X} \triangleq \Sigma^{\frac{1}{2}}\underline{u} + \underline{\mu}$$

$$\xrightarrow{\text{change of variable}} f_{\underline{X}}(x) = \frac{1}{|\det(\Sigma^{\frac{1}{2}})|}f_{\underline{u}}(\Sigma^{-\frac{1}{2}}(x-\mu)) = \frac{1}{\sqrt{\det(\Sigma)}}\frac{1}{2\pi}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

$\boxed{\Sigma \text{ is always PSD}}$

proof $\quad \forall\alpha \in \mathbb{R}^n$

$$\alpha^T\Sigma\alpha = \sum_{i,j=1}^{n}\alpha_i\alpha_j\Sigma_{ij} = \mathbb{E}\left[\sum_{i,j=1}^{n}\alpha_i\alpha_j(X_i-\mu_i)(X_j-\mu_j)\right]$$

$$\Sigma_{ij} = \mathbb{E}[(X_i-\mu_i)(X_j-\mu_j)]$$

$$= \mathbb{E}\left[\left(\sum_{i=1}^{n}\alpha_i(X_i-\mu_i)\right)^2\right] \geq 0$$

⊛ اگر هیچ کدوم از R.V. ها از ترکیب خطی بقیه deterministically به دست نیاد $\underbrace{\text{(مستقل جبرخطی)}}_{\sum_{i=1}^{n}\alpha_i(X_i-\mu_i) \neq 0}$ اون وقت Σ PD میشه.

$$\Sigma^{\frac{1}{2}} = \Lambda^{\frac{1}{2}}Q^T \xleftarrow{\text{PSD}} \Sigma = Q\Lambda Q^T$$

positive definite matrix $\longrightarrow$ full rank matrix

symmetric matrix $\longrightarrow$ eig val decomp $\quad \Sigma = Q\Lambda Q^T$

Jointly Gaussian

AKA Multivariate Gaussian

$$\underline{X} = [X_1 \dots X_n] \in \mathbb{R}^n$$

$$f_{\underline{X}}(\underline{x}) = \frac{1}{\sqrt{(2\pi)^n \, |\det(\Sigma)|}} \exp\left(-\frac{1}{2}(\underline{x}-\underline{\mu})^T \Sigma^{-1} (\underline{x}-\underline{\mu})\right) \quad \forall \underline{x} \in \mathbb{R}^n$$

$\underline{\mu} \in \mathbb{R}^n \longrightarrow$ mean vector

$\Sigma \in \mathbb{R}^{n \times n} \ (\succeq 0) \longrightarrow$ covariance matrix

$X_1 \dots X_n$ are independent $\xleftrightarrow{J.G}$ Σ is diagonal ⊛

$\underline{X_1} \perp\!\!\!\perp \underline{X_2} \xleftrightarrow{J.G} \Sigma = \begin{array}{c} \begin{bmatrix} \blacksquare & 0 \\ 0 & \blacksquare \end{bmatrix} \begin{matrix} [\underline{X_1}] \\ [\underline{X_2}] \end{matrix} \end{array}$ ⊛

⊛ اگر $\underline{X} \sim J.G$ و روی $\underline{X'} \subset \underline{X}$ marginalize کنیم μ, Σ جدید همان μ, Σ قدیم می‌شوند که فقط سطر و ستون های متناظر با اعضای $\underline{X'}$ از آن ها حذف شده است

⊛ می توان $\underline{X} \sim \mathcal{N}(\underline{\mu}, \Sigma) \in \mathbb{R}^n$ را با $u_1 \dots u_m \overset{iid}{\sim} \mathcal{N}(0,1)$ ساخت به طوری که $m < n$ باشد. به این صورت Σ دیگه full rank نخواهد بود

$$\Sigma = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{XY}^T & \Sigma_{YY} \end{bmatrix} \begin{matrix} [X] \\ [Y] \end{matrix} \longrightarrow X|Y \sim \mathcal{N}\left(\Sigma_{XY}\Sigma_{YY}^{-1}(Y-\mu_Y)+\mu_X ,\ (\quad)\right)$$ ⊛

⊛ اگر $\underline{X} = [X_1 \dots X_n] \sim J.G$ به ازای هر ماتریسی مانند A، به دلیل فرمول change of variable داریم $A\underline{X} \sim J.G$

Gaussian process

$\{X(t) \mid t \in T\}$ is a Gp iff $\forall n \in \mathbb{N}, \ \forall t_1 \dots t_n \in T \quad [X(t_1) \dots X(t_n)] \sim J.G$

- $\xrightarrow{\text{discrete } T}$ DT-Gp $\quad (\mathbb{Z}, \mathbb{Z}_{\geq 0}, \dots)$
- $\xrightarrow{\text{continuous } T}$ CT-Gp $\quad (\mathbb{R}, \mathbb{R}_{\geq 0}, \dots)$

این تعریف constructive نیست. مگه اگه یک جوری فرایندی به دست آوردی که این خاصیت رو داشت، فرایندت GP عه

IID Gp

$\subset$ DT-Gp

$$W = \{w(n) \mid n \in \mathbb{Z}\} \quad \dots w(-1), w(0), w(1), \dots \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

$$\mu_W(n) = 0$$

$$\Sigma_{WW}(n, n+k) = \sigma^2 \delta_{n, n+k} = \sigma^2 \delta(k=0)$$

Cmo stationary

⊛ فرایند گاوسی همون J.G عه که ∞ تایی‌ش کردیم پس $\mu \in \mathbb{R}^n$ تبدیل میشه به $\mu(t): T \to \mathbb{R}$

و $\Sigma \in \mathbb{R}^{n \times n}$ تبدیل میشه به $\Sigma(t_1,t_2): T \times T \to \mathbb{R}$

⊛ فرایند پواسون با λ مشخص می‌شد ، فرایند گاوسی هم با $\mu(t)$, $\Sigma(t_1,t_2)$ به طور یکتا مشخص می‌شه ؟

⊛ $\mu(n)$, $\Sigma(n,m)$ belong to a Gp iff $\forall n \in \mathbb{N}$, $\forall t_1 \cdots t_n \in T$ $\quad \left[\Sigma(t_i,t_j)\right]_{i,j=1}^{n}$

is symmetric and PSD (no conditions for $\mu(n)$)

$\xrightarrow{\text{e.g.}}$ $\mu(n)=0$, $\Sigma(n,m)=\delta_{n-m}$ $\quad$ Valid

$\xrightarrow{\text{e.g.}}$ $\mu(n)=\sin(2\pi \omega n)$, $\Sigma(n,m)=\rho\, e^{-|n-m|}$ $\quad$ Valid

AKA Gaussian Random walk

$\boxed{\text{Gaussian sum process}}$ $\subset$ DT-Gp

$$S=\{s(n) \mid n \in \mathbb{N}\}$$

$$s(n)=\sum_{i=1}^{n} w(i) \quad , \; w(i) \text{ is IID Gp}$$

می دانیم که IID Gp یک فرایند گاوسی است . بنابراین $\forall n$

$$\begin{bmatrix} s(1) \\ \vdots \\ s(n) \end{bmatrix} = \begin{bmatrix} 1 & & 0 \\ \vdots & \ddots & \\ 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} w(1) \\ \vdots \\ w(n) \end{bmatrix}$$

نیز J.G است پس S نیز یک فرایند گاوسی Valid است

change of variable

$$\mu_s(n)=\mathbb{E}\left[\sum_{i=1}^{n} w(i)\right]=0$$

$$\Sigma_{ss}(n,m)=\mathbb{E}\left[s(n)\,s(m)\right] \overset{\text{اگر } n \leq m}{=\!=\!=} \mathbb{E}\left[s(n)\left(s(n)+w(n+1)+\cdots+w(m)\right)\right]$$

$$=\underbrace{\mathbb{E}\left[s(n)^2\right]}_{n\sigma^2}+\underbrace{\mathbb{E}\left[\left(w(1)+\cdots+w(n)\right)\left(w(n+1)+\cdots+w(m)\right)\right]}_{w(i)\text{s are iid} \,\to\, 0}$$

$$\to \Sigma_s(n,m)=\min(n,m)\,\sigma^2 \quad \text{Valid}$$

stationary نیست

زیرا $\sum_{i=1}^{n} w(i)$ و $\sum_{i=1}^{n+\tau} w(i)$ هم توزیع نیستن

$\boxed{\text{Gauss-Markov process}} \subset \text{DT-Gp}$

فرم ۱

$X = \{X(n) \mid n \in \mathbb{Z}_{\geq 0}\}$

$X(0) = 0$

$X(n+1) = \alpha X(n) + w(n)$

$|\alpha| < 1$, $w(n)$ is IID Gp

$\longrightarrow X(n) = w(n-1) + \alpha w(n-2) + \dots + \alpha^{n-1} w(0)$

$\longrightarrow \mu_X(n) = 0$

$\longrightarrow \Sigma_{XX}(n, n+k) = \mathbb{E}[X(n)\, X(n+k)]$

$= \sigma^2 \dfrac{1-\alpha^{2n}}{1-\alpha^2} \alpha^k$

استقرا روی k

برای n های بزرگ چون $|\alpha|<1$ ، Σ_{XX} به n بستگی نداره

پس این فرایند asymptotically stationary است

فرم ۲

$X = \{X(n) \mid n \in \mathbb{Z}\}$

$X(n+1) = \alpha X(n) + w(n)$

$|\alpha| < 1$, $w(n)$ is IID Gp

$\longrightarrow X(n) =$ مساله؟

$\longrightarrow \mu_X(n) = 0$

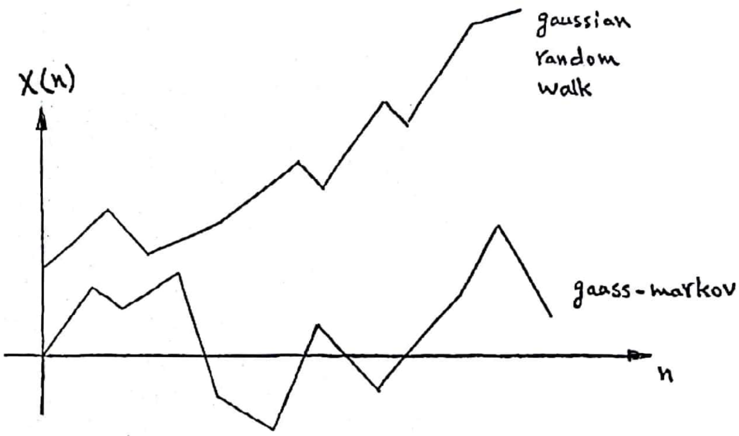
$\longrightarrow \Sigma_{XX}(n, n+k) = \dfrac{\sigma^2}{1-\alpha^2} \alpha^k$

stationary هست

* تعریفی داریم که میگه برای Gp ها، فرایند stationary است اگر و تنها اگر

$\forall t,u \in T \quad \mathbb{E}[X(t)] = \mathbb{E}[X(0)]$

and $\Sigma_{XX}(t, t+u) = \Sigma_{XX}(0, u)$



* stationary بودن فرم ۱ و فرم ۲ رو با این تعریف چک کنیم

$\boxed{\text{Stationarity}}$ AKA strictly stationary

$\forall n \in \mathbb{N} \quad \forall t_1 \cdots t_n \in T \quad \forall \Delta \mid t_i + \Delta \in T$

X is stationary if

$$\big(X(t_1) \cdots X(t_n)\big) \overset{\text{I.D.}}{=\!=} \big(X(t_1+\Delta) \cdots X(t_n+\Delta)\big)$$

$$\equiv$$

$$F_{X(t_1)\cdots X(t_n)}(x_1 \cdots x_n) = F_{X(t_1+\Delta)\cdots X(t_n+\Delta)}(x_1 \cdots x_n)$$

⊛ ایستایی برای فرایند تعریف میشه نه متغیر تصادفی

⊛ $X(t_i)$ ها باید هم توزیع باشند نه مستقل

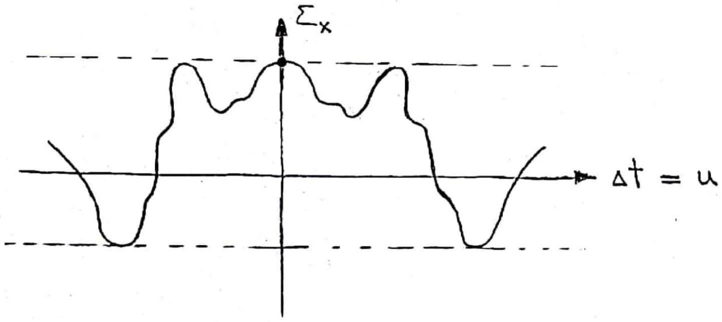
$\boxed{\text{Wide sense stationary (WSS)}}$ AKA weakly stationary
AKA second-order stationary

⊛ WSS فقط به moment های مرتبه 1 و 2 نگاه می کنه

X is WSS if $\forall t, u \in T$

- $\mathbb{E}[X(t)] = \mathbb{E}[X(0)] = \text{const}$
- $\Sigma_{XX}(t, t+u) = \Sigma_{XX}(0, u) \triangleq \Sigma_X(u)$

چون فقط به فاصله ربط داره به صورت تک متغیره می نویسیمش



⊛ تابع Σ_X باید زوج باشه و max ش در $\Delta t = 0$ اتفاق افتاده باشه (Var ها). جاهای دیگه هم می تونه دوباره max بشه. min ش هم نهایتا max- خواهد بود.

stationary $\xleftarrow{\text{GP}}$ WSS ⊛

stationary $\longrightarrow$ WSS ⊛

$\boxed{\text{Ergodicity}}$

اگر تنها با یک sample path بشه راجع به ویژگی های statistical یک تک برش زمانی مثل $X(t_0,\omega)$ حرف زد می گیم اون فرایند ergodic عه.
مثلاً اگه بتونیم $\mathbb{E}[X(t_0)]$ رو تنها با یک sample path به دست بیاریم.

$$\text{if} \quad \underbrace{\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} X(t_0, \omega_i)}_{\text{ensemble average}} = \underbrace{\lim_{T\to\infty} \frac{1}{T} \int_{t_0-\frac{T}{2}}^{t_0+\frac{T}{2}} X(t,\omega)\, dt}_{\text{time average}} \quad \text{then } X(t,\omega) \text{ is min-ergodic}$$

ensemble average $\equiv$ ground truth

⊛ اگر علاوه بر ممان اول، همه ی ممان ها این تساوی رو داشته باشند، فرایند ergodic خواهد بود.

$$\boxed{DT-Gp \longrightarrow CT-Gp}$$

AKA orthonormal expansion

ساختن فرایند های CT کار سختی است اما می‌توان
با کمک پایه های orthonormal از روی فرایند های DT
فرایند های CT ساخت.
بسته به قدرت پایه‌ی انتخاب شده، طیف های مختلفی
از فرایند های CT را می‌توان ساخت

$$\{\phi_i(t) \mid i \in \mathbb{N}\} \quad \phi_i : \mathbb{R} \longrightarrow \mathbb{C}$$

پایه های orthonormal
برای فضای توابع

$$\int_{-\infty}^{\infty} |\phi_i(t)|^2 \, dt = 1 \quad \forall i \in \mathbb{N}$$

شرط نرمال بودن

$$\int_{-\infty}^{\infty} \phi_i(t) \, \phi_j^H(t) \, dt = 0 \quad i \neq j$$

شرط تعامد

$$\left(\equiv \int_{-\infty}^{\infty} \phi_i(t) \, \phi_j(t) \, dt = \delta_{ij} \right)$$

$$\longrightarrow X(t) = \sum_{i=1}^{\infty} X_i \, \phi_i(t)$$

و

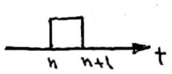
$$X_i = \int_{-\infty}^{\infty} X(t) \, \phi_i^H(t) \, dt$$

⊛ در ادامه تعاریف constructive برای زیرکلاس های مختلفی
از CT-Gp ارائه می‌شود

Taylor series Basis

$$\phi_n(t) = \frac{t^n}{n!} \longrightarrow \left\{ X(t) \,\middle|\, X(t) = \sum_{i=0}^{\infty} \frac{\overbrace{W(i)}^{\text{IID Gp}}}{i!} t^i \right\}$$

pulse Basis

$\phi_n(t) =$ 

فرایند هایی که توسط این پایه ساخته شوند stationary نیستند
زیرا اگر غیر صحیح شیفت بخورند معلوم می‌شود.
⊛ اگر فقط صحیح شیفت بخورند معلوم نمی‌شود

$$\mathbb{E}\left[X(t)\, X(t+\Delta)\right] = \mathbb{E}\left[\sum_{i=0}^{\infty} X_i \, \phi_i(t) \cdot \sum_{j=0}^{\infty} X_j \, \phi_j(t+\Delta)\right]$$

$$= \mathbb{E}\left[\sum_{i,j=0}^{\infty} X_i X_j \, \phi_i(t) \, \phi_j(t+\Delta)\right]$$

$$= \sum_{i,j=0}^{\infty} \phi_i(t) \, \phi_j(t+\Delta) \, \mathbb{E}\left[X_i X_j\right]$$

session 14
1404.1.30

if $\{\phi_n(t)\}_{n=1}^{\infty}$ are an orthonormal basis for L^2 over $\mathbb{R}$

and $X_i \sim \mathcal{N}(0, \sigma_i^2)$, $X_i \perp\!\!\!\perp X_j$ $i \neq j$ (not iid) and $\sum_{i=1}^{\infty} \phi_i^2(t)\, \sigma_i^2 < \infty$ $\forall t \in T$

then $\{X(t) \mid X(t) = \sum_{i=1}^{\infty} X_i\, \phi_i(t)\}$ is ct-GP

⊛ یعنی با هر پایه‌ای orthonormal expansion انجام بدیم، نتیجه‌ش GP عه

intuative proof

$$\forall N \in \mathbb{N} \quad \begin{bmatrix} X(t_1) \\ \vdots \\ X(t_N) \end{bmatrix} = \begin{bmatrix} \phi_1(t_1) & \cdots \\ \vdots & \\ \phi_1(t_N) & \cdots \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \end{bmatrix}$$

برای GP بودن باید هر زیرمجموعه N تایی از فرایند J.G می‌شد. حال داریم
و چون $\begin{bmatrix} X_1 \\ \vdots \end{bmatrix} \sim$ J.G پس طبق change of variable داریم $\begin{bmatrix} X(t_1) \\ \vdots \end{bmatrix} \sim$ J.G

همچنین چون واریانس گاوسی نمی‌تونه ∞ باشه باید داشته باشیم $Var(X(t)) < \infty$

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

$$= \mathbb{E}\left[\sum_{i,j=1}^{\infty} X_i X_j\, \phi_i(t)\, \phi_j(t)\right] \xrightarrow{X_i \perp\!\!\!\perp X_j} \sum_{i=1}^{\infty} \phi_i^2(t)\, \underbrace{\mathbb{E}\left[X_i^2\right]}_{\sigma_i^2}$$

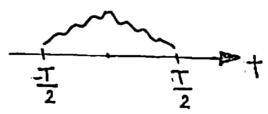
که طبق فرض $< \infty$ بود ✓

$\boxed{\text{Fourier Basis}}$ ⟶ continuous fourier series

$$\phi_n(t) = e^{\frac{i 2\pi n t}{T}} \qquad |t| < \frac{T}{2}$$

⊛ تمام CT-GP های square integrable بین $\frac{-T}{2}$ تا $\frac{T_2}{2}$ رو میشه با پایه‌ی فوریه ساخت

سیگنال و سیستم :: C_n ⟷ $x(t)$
discrete infinite — continuous periodic



⊛ می‌بره به $e^{-\cdots}$ حوزه‌ی فوریه

فرایند :: X_n ⟷ $X(t)$

⊛ می‌شد به جای CFS از DTFT استفاده کرد
به این صورت: $\phi_n(t) = e^{-\frac{i2\pi n t}{T}}$

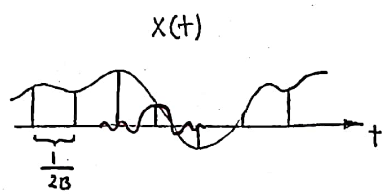
$$\frac{2\pi}{T} t = \omega = 2\pi f$$

$\boxed{\text{Gaussian sinc process}}$ $\subset$ CT-GP

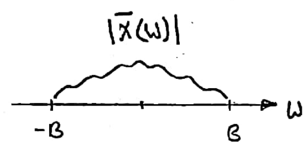
$$\phi_n(t) = \text{sinc}(2Bt - n)$$

طبق قضیه‌ی Nyquist اگر $X(t)$ تابعی band limited باشد، یعنی $\overbrace{\left|FT\{X(t)\}\right|}^{\tilde{X}(\omega)}$ بین $-B, B$ باشد، با نرخ نمونه برداری $2B$ می‌توان $X(t)$ اصلی را کامل بازیابی کرد و داریم:

$$X(t) = \sum_{n=-\infty}^{\infty} X\left(\frac{n}{2B}\right) \text{sinc}(2Bt - n)$$



FT ⟷



$$\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$$

$$\int_{-\infty}^{\infty} \text{sinc}^2(x)\, dx = 1$$

$$\int \text{sinc}(t-n)\, \text{sinc}(t-k)\, dt = \delta_{nk}$$

⊛ توابع band limited نمی‌تونن innovation زیادی داشته باشن

⊛ این جا نیازی نیست قیدی روی σ_i داشته باشیم چون خود پایه ها
وقتی i زیاد بشه (در نقطه‌ی t) به صفر میل می‌کنند و $\phi_i^2(t)$
شرط $Var(X(t))<\infty$ رو ارضا می‌کنه

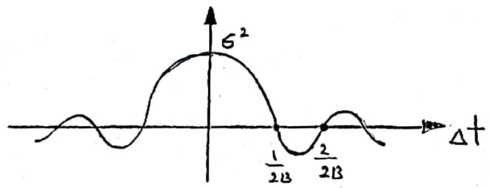
$$\Sigma_{xx}(t_1, t_2) = \mathbb{E}\left[X(t_1)\, X(t_2)\right]$$

$$= \mathbb{E}\left[\sum_{n,k=-\infty}^{\infty} X\left(\frac{n}{2B}\right) X\left(\frac{k}{2B}\right) \operatorname{sinc}(2Bt_1 - n)\operatorname{sinc}(2Bt_2 - k)\right]$$

$$\overset{\perp}{=} \mathbb{E}\left[\sum_{n=-\infty}^{\infty} X^2\left(\frac{n}{2B}\right) \operatorname{sinc}(2Bt_1 - n)\operatorname{sinc}(2Bt_2 - n)\right]$$

$$= \sigma^2 \sum_{n=-\infty}^{\infty} \operatorname{sinc}(2Bt_1 - n)\operatorname{sinc}(2Bt_2 - n)$$

$$= \sigma^2 \operatorname{sinc}\left(2B\overbrace{(t_2 - t_1)}^{\Delta t}\right)$$



مشاهده می‌شود که cov. func. تنها به Δt بستگی دارد و
$\Sigma_{xx}(t,t) = \sigma^2$ که const است ⟵ فرایند WSS است
و چون GP است ⟵ stationary هست

⊛ همان طور که مشاهده می‌شود در $\Delta t \to \infty$ به سمت استقلال می‌رویم و وقتی $\Delta t \to 0$ به سمت وابستگی بیشتر می‌رویم

⊛ محدودیت Gaussian sinc process نسبت به پایه‌ی سری فوریه کمتره

⊛ CFS basis $= \phi_n(t) = \begin{cases} \frac{1}{\sqrt{T}} e^{\frac{i2\pi n t}{T}} & |t| \leq \frac{T}{2} \\ 0 & \text{otherwise} \end{cases}$ exponential form

$$= \begin{cases} \sqrt{\frac{2}{T}} \cos\left(\frac{2\pi n t}{T}\right) & n>0 \;,\; |t| \leq \frac{T}{2} \\ \sqrt{\frac{1}{T}} & n=0 \;,\; |t| \leq \frac{T}{2} \\ \sqrt{\frac{2}{T}} \sin\left(\frac{-2\pi n t}{T}\right) & n<0 \;,\; |t| \leq \frac{T}{2} \\ 0 & \text{otherwise} \end{cases}$$ sine – cosine form

⊛ $X_n \in \mathbb{R} \longleftrightarrow x(t)$ is even

فرایند هایی که توسط پایه های فوریه ساخته می‌شوند stationary نیستند
زیرا periodic اند
$$\Sigma_{xx}(t, t+T) = \mathbb{E}\left[x(t)^2\right]$$
به t بستگی دارد نه Δt

$\boxed{\text{Filtered cT processes}}$

linear system $\quad \xrightarrow[X_1(t)+X_2(t)]{\alpha X(t)} \boxed{} \xrightarrow[Y_1(t)+Y_2(t)]{\alpha Y(t)}$

$X(t) \longrightarrow \boxed{h(t)} \xrightarrow{Y(t)}$ ⊛
deterministic
LTI

time invariant system $\quad \xrightarrow{X(t-\tau)} \boxed{} \xrightarrow{Y(t-\tau)} \quad \equiv$ stationarity

⊛ سیستم های LTI با پاسخ ضربه شون به صورت یکتا مشخص می‌شوند. یعنی اگر خروجی به ازای تک ورودی $\delta(t)$ رو بدونیم، همه چیز رو می‌دونیم
چون هر ورودی ای رو میشه به صورت $x[t] = \sum_{n=-\infty}^{\infty} x[n]\,\delta[t-n]$ نوشت

$$⊛\ Y(t) = X(t) * h(t) = \int_{-\infty}^{\infty} X(t-\tau)\, h(\tau)\, d\tau$$

R.V.	LTI system	R.V.
sp		sp
⊛ zero mean sp		zero mean sp
GP		GP
wss sp		wss sp
stationary sp		stationary sp

$$\mu_Y(t) = \mathbb{E}\left[\int_{-\infty}^{\infty} X(t-\tau)\, h(\tau)\, d\tau\right]$$
$$= \int_{-\infty}^{\infty} \underbrace{\mathbb{E}\left[X(t-\tau)\right]}_{\mu_X(t-\tau)} h(\tau)\, d\tau = \mu_X(t) * h(t)$$

با فرض zero mean Y

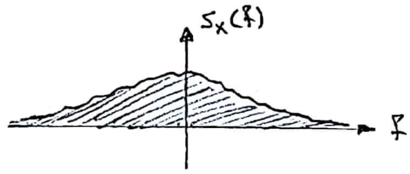
$$\Sigma_{YY}(t, t+\tau) = \mathbb{E}\left[\int_{-\infty}^{\infty} X(t-u)\, h(u)\, du \cdot \int_{-\infty}^{\infty} X(t+\tau-v)\, h(v)\, dv\right]$$
$$= \iint_{-\infty}^{\infty} \underbrace{\mathbb{E}\left[X(t-u)\, X(t+\tau-v)\right]}_{\Sigma_{XX}(t-u,\, t+\tau-v) \ \underset{\text{if wss}}{=} \ \Sigma_X(\tau - v + u)} h(u)\, h(v)\, du\, dv$$
$$\overset{\dots}{=} h(-\tau) * \Sigma_X(\tau) * h(\tau)$$

چگالی طیفی
$\boxed{\text{power spectral Density}}$ ⇝ defined for wss sp

$$S_X(f) = FT\{\Sigma_X(\tau)\} = \int_{-\infty}^{\infty} \Sigma_X(\tau)\, e^{-i2\pi f\tau}\, d\tau$$

$$\Sigma_X(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(f)\, e^{i2\pi f\tau}\, df$$

$$\rightarrow \underbrace{\Sigma_X(0) = \mathbb{E}\left[|X(t)|^2\right]}_{\underset{X(t)}{\overset{\text{for wss}}{=}} \ \text{power}(X(t))} = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(f)\, df$$



$$\text{Energy}(X(t)) = \mathbb{E}\left[\int_0^T |X(t)|^2\, dt\right]$$
$$\text{power}(X(t)) = \lim_{T\to\infty} \frac{1}{T}\, \text{Energy}(X(t))$$

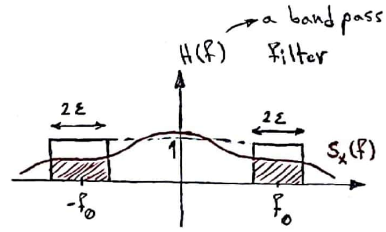
$$\xrightarrow[\mathbb{E}[|X(t)|^2] \text{ is cnst}]{\text{for wss } X(t)} \text{power}(X(t)) = \mathbb{E}\left[|X(t)|^2\right]$$

$$\mu_x(t)\ ,\ \Sigma_x(\tau) \xrightarrow{\ h\ } \mu_x(t) * h(t)\ ,\ h(-\tau) * \Sigma_x(\tau) * h(\tau)$$

$$\bar{\mu}_x(f)\ ,\ S_x(f) \xrightarrow{\ H\ } \bar{\mu}_x(f)\, H(f)\ ,\ S_x(f)\, |H(f)|^2$$

$$2\pi \Sigma_Y(0) = \int_{-\infty}^{\infty} S_Y(f)\, df = \int_{-\infty}^{\infty} S_x(f)\, |H(f)|^2\, df$$

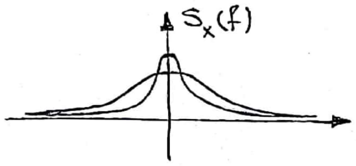
$$= \int_{\pm f_0 - \varepsilon}^{\pm f_0 + \varepsilon} S_x(f)\, df \simeq 2\varepsilon \left(S_x(f_0) + S_x(-f_0) \right)$$



⊛ با $S_x(f)$ می‌توان یک zero mean GP را به صورت یکتا مشخص کرد زیرا توسط آن می‌توان Σ_x را داشت.

⊛ با توجه به اینکه $S_Y(f) = S_x(f)\, |H(f)|^2$ برای کامل مشخص شدن فرایند Y فقط به $|H(f)|$ نیاز داریم و فاز $H(f)$ بی اهمیت است.

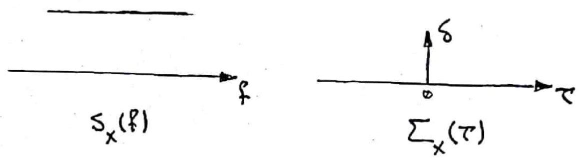
$\boxed{\text{White Noise}}$



$$\frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(f)\, df = 1$$

⊛ spectral density مربوط به noise ها Flat تر است (اما مساحت زیر آن ثابت می‌ماند)

⊛ در حالت حدی اگر $S_x(f)$ را ثابت فرض کنیم نویز خاصی به نام white noise به دست می‌آید که در عمل وجود ندارد اما ویژگی های جالبی دارد



مانند نور سفید همه فرکانس‌ها را دارد

$$\text{if } X(t) \text{ is white noise} \longrightarrow \frac{1}{2\pi} \int S_x(f)\, df = \Sigma_x(0) = \infty$$

ولی گاوسی با واریانس ∞ وجود ندارد

⊛ سیستم های LTI واقعی همگی low pass هستند

$\boxed{\text{wiener process}}$ AKA Brownian motion process
$\subset$ CT-GP

البته $X(t)$ برخلاف $N(t)$ لزوما افزایشی نیست

a zero mean GP that has stationary inc. and independent inc. properties $\left(X(t) \triangleq N(t) \right)$
and $X(0)=0$ and has continuous sample functions wp1 $\{X(t)\,;\, t \geq 0\}$

stationary نیست

theorem if $\{X(t)\,;\, t \geq 0\}$ has stationary inc. and independent inc. and $X(0)=0$

then $\forall \tau > t$ $\quad \mathbb{E}[X(t)] = t\,\mathbb{E}[X(1)]$ and $\Sigma_{xx}(t,\tau) = \operatorname{var}(X(t))$
$= t \operatorname{var}(X(1))$

⊛ فرایند هایی که stationary inc. , independent inc. را دارند نمی‌توانند stationary باشند زیرا $\Sigma_{xx}(t,\tau)$ آنها به t بستگی دارد

⊛ اگر فرایندی stationary باشد آنگاه خاصیت stationary inc. را نیز دارد

proof

$$X(n\Delta) \xleftarrow{t=n\Delta} \quad X(n\Delta) = \left(X(n\Delta) - X((n-1)\Delta) \right) + \cdots + \left(X(\Delta) - X(0) \right)$$

طبق stationary inc. تمام پارامترها با $X(\Delta)$ هم توزیع اند
و طبق independent inc. پارامتر ها مستقل اند

$$\mathbb{E}[X(n\Delta)] = n\,\mathbb{E}[X(\Delta)] \checkmark$$

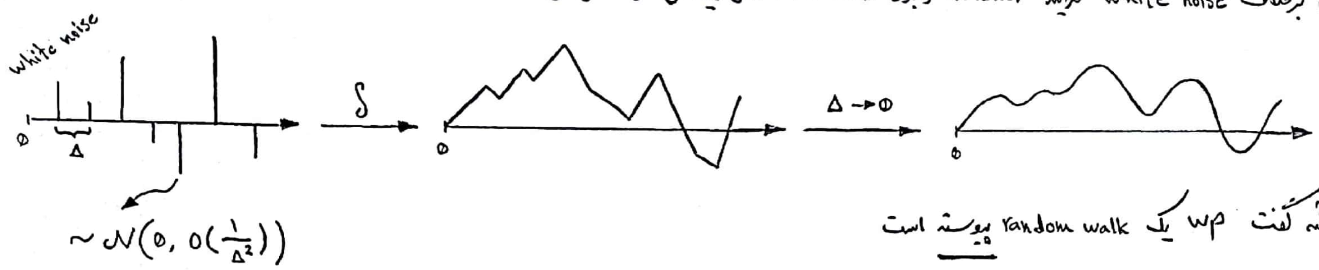
$$\operatorname{Var}(X(n\Delta)) = n \operatorname{Var}(X(\Delta))$$

$$\Sigma_{xx}(t,\tau) \overset{\tau>t}{=\!=} \mathbb{E}\left[X(t)\left(X(t) + X(\tau-t)\right)\right] = \underbrace{\mathbb{E}[X^2(t)]}_{\operatorname{Var}(X(t))} + \underbrace{\mathbb{E}[X(t)]}_{0}\,\underbrace{\mathbb{E}[X(\tau-t)]}_{0 \text{ zero mean}}$$

$$\longrightarrow \Sigma_{xx}(t,\tau) = t \operatorname{var}(X(1)) \checkmark$$

$\boxed{\text{نحوه‌ی ساخت wp}}$

⊛ برخلاف white noise فرایند wiener وجود داره. از راه های دیگه ای می سازنش



میشه گفت wp یک random walk پیوسته است

ویژگی های دیگه wp

- symmetry ← منفی sample path هم همچنان یک sample path معتبره

- reflection ← $\tilde{w}(t)$ هم یک sample path معتبره

$$T_a \triangleq \inf\{t \mid w(t) \geq a\} \qquad \tilde{w}(t) \triangleq \begin{cases} w(t) & t \leq T_a \\ 2a - w(t) & t \geq T_a \end{cases}$$

$$\xrightarrow[\mathbb{E}[X(1)]=0\ ,\ \operatorname{var}(X(1))=1]{\text{standard wp}} \mathbb{E}[X(t)] = 0 \ ,\ \Sigma_{xx}(t,\tau) = \operatorname{Var}(X(t)) = \min(t,\tau)$$

session 16
1404.2.6

⊛ CT Markov chain هم وجود داره ولی ما باهاش کاری نداریم

$\boxed{\text{Markov chain}} \subset$ Markov process

$\{X_n \mid n \in \mathbb{Z}_{\geq 0}\}$ where $X_0 \cdots \in S$ (state space) and S is countable

- $S = \{1, \dots, M\} \longrightarrow$ finite state markov chain
- $S = \{1, \dots\} \longrightarrow$ infinite state markov chain

and $\mathbb{P}(X_n = j \mid X_{n-1} = i, \dots, X_0 = l) = \mathbb{P}(X_n = j \mid X_{n-1} = i)$

⊛ در حالت کلی markov process ها می‌تونن S پیوسته مثل IR داشته باشن.
به زیرمجموعه ای که S‌ش شمارا باشه
گفته میشه markov chain

⊛ در ادامه هرجا گفتیم MC منظورمون DT finite space MC عه!

$$P(X_0 \cdots X_n) \xlongequal{\text{chain rule}} P(X_0) \prod_{i=1}^{n} P(X_i \mid X_{i-1} \cdots X_0)$$

$$\xlongequal{\text{markov chain}} P(X_0) \prod_{i=1}^{n} P(X_i \mid X_{i-1})$$

⊛ تنها اخیر ترین state اهمیت داره یعنی:
$X_n \mid X_{n-2} \xlongequal{\text{I.O.}} X_n \mid X_{n-2}, X_{n-5}$

initial state distribution $\left[\mathbb{P}(X_0 = i)\right]_{i=1}^{M} = \pi_0 \xrightarrow{\text{sum}} 1$

transition probability matrix $\left[\mathbb{P}(X_{n+1} = j \mid X_n = i)\right]_{i,j=1}^{M} = \begin{bmatrix} P_{11} & \cdots & P_{1M} \\ \vdots & & \vdots \\ P_{M1} & \cdots & P_{MM} \end{bmatrix} = P$, rows $\xrightarrow{\text{sum}} 1$

⊛ random walk حالت خاصی از markov chain عه

⊛ کار ما با homogenous markov chain عه که $\mathbb{P}(X_{n+1} = j \mid X_n = i) = P_{ij}$ به n بستگی نداره.
یعنی همواره احتمال رفتن از استیت i به j یکسان عه.

⊛ یک markov chain با یک initial state (M تایی) و transition با $M^2 - M$
به صورت یکتا مشخص میشه

زیرا ستون آخر ماتریس رو میشه به صورت $\begin{bmatrix} 1 - \Sigma P_{1i} \\ \vdots \\ 1 - \Sigma P_{Mi} \end{bmatrix}$ نوشت

Definitions

n step walk $\longrightarrow$ an ordered string of nodes $(i_0, \dots, i_n)$ $\underline{n \geq 1}$ where $p_{X_0 \cdots X_n}(i_0 \cdots i_n) > 0$

path $\longrightarrow$ a walk in which no nodes are repeated

cycle $\longrightarrow$ a walk in which the first and the last nodes are the same and no other node is repeated

$(i \to j) \longrightarrow$ state j is <u>accessible</u> from i if there is a walk from i to j $\xleftrightarrow{\text{iff}}$ $\exists n \; P_{ij}^{n} > 0$

$(i \leftrightarrow j) \longrightarrow$ distinct states i, j <u>communicate</u> if $i \to j$ and $j \to i$

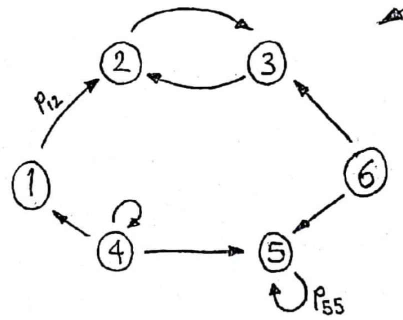
$P_{ij}^{n} \triangleq \mathbb{P}(X_n = j \mid X_0 = i)$

از قضا این عبارت برابر اینه که ماتریس P رو به توان n برسونی
و بعد المان [i,j] ش رو برداری!

$$\left.\begin{array}{l} (i \to j) \quad P_{ij}^{n} > 0 \\ (j \to k) \quad P_{jk}^{m} > 0 \end{array}\right\} \longrightarrow P_{ik}^{n+m} \geq P_{ij}^{n} P_{jk}^{m} > 0$$

$(i \to k)$

Session 17
1404.2.8



⊛ به دلیل گسسته بودن در زمان و state ها و محدود بودن تعداد state ها، می توانیم برای Finite state markov chain ها نمایش گرافی داشته باشیم

⊛ هر markov chain مرتبه k رو میشه به صورت markov chain مرتبه 1 درست.

مثلا برای $k=2$ داریم :

$$S=\{1,\dots,M\} \longrightarrow \tilde{S}=\{(1,1)\cdots(M,M)\}$$

$$|S|=M \qquad |\tilde{S}|=M^2$$

$$P(X_{n+1}=j \mid X_n=i, X_{n-1}=k) = P(\underbrace{X_{n+1}=j, X_n=i}_{\tilde{X}_{n+1}=(j,i)} \mid \underbrace{X_n=i, X_{n-1}=k}_{\tilde{X}_n=(i,k)})$$

به این صورت ماتریس P دارای M^4 درایه است که M^3 تای آن که به فرم $P_{(j,i),(i,k)}$ هستند حاوی اطلاعات اند و سایر المان ها صفر اند.

Definitions

class of states ⟶ a non empty set of states $C \subseteq \{1,\dots,M\}$ such that if $i\in C$ and $j\neq i$ then if $i \xleftrightarrow{\text{com.}} j \rightarrow j\in C$ and if $i \not\xleftrightarrow{\text{com.}} j \rightarrow j\notin C$

گراف افراز میشه به تعدادی کلاس و هر عضو یک کلاس با تمام اعضا communicate می کنه و با هیچ کلاس دیگه ای communicate نمی کنه

recurrent state ⟶ state i is recurrent if $i\rightarrow j$ implies $j\rightarrow i$

transient state ⟶ a state which is not recurrent

استیت هایی که اگر یک بار دیده بشن با احتمال 1 بی نهایت بار دیگه هم دیده میشن

⊛ هرچقدر هم احتمال دیده شدن یک state کم باشه اگر صفر نباشه میگیم با احتمال 1 بالاخره دیده میشه

periodic state ⟶ state i which its period

$$d(i) \triangleq \underset{\text{ب.م.م}}{GCD}(\{n\geq 1 \mid P_{ii}^n > 0\}) > 1$$

recurrent class ⟶ a class where all its states are recurrent

transient class ⟶ 〃 〃 〃 are transient

periodic class ⟶ 〃 〃 〃 are periodic

اثبات میشود که state های یک کلاس یا تماما rec. اند یا تماما trans.

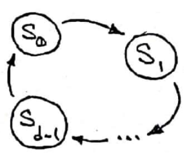
اثبات میشود که تمام state های یک کلاس period های یکسان دارند

ergodic class ⟶ a class where all its states are recurrent and aperiodic

ergodic chain ⟶ a markov chain consisting only of one ergodic class

⊛ هر markov chain حداقل یک recurrent class داره وگرنه halt میشیم و دیگه طول فرایند، بی نهایت نخواهد بود.

⊛ recurrent class ها فقط یال بهشون وارد میشه



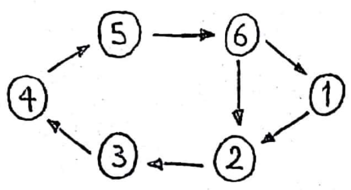
⊛ اگر period یک کلاس برابر d باشه، میشه state های اون کلاس رو به d بخش افراز کرد به طوری که :

⊛ periodic بودن نشون دهنده‌ی وجود یک جور نظم عه که مطلوب ما نیست !

⊛ اگر یک state طوقه داشته باشه ⟵ aperiodic

$\boxed{\text{Theorem}}$ For an ergodic M state markov chain, for all i,j, $m \geq (M-1)^2+1$ we have $P_{ij}^{m} > 0$

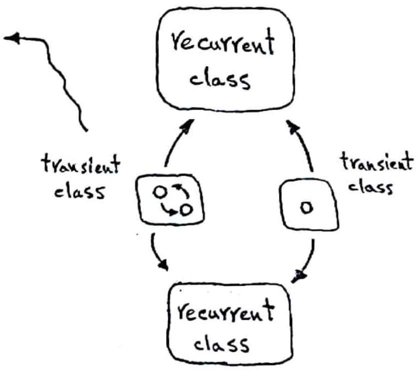
این کران میگه بعد از m، به ازای هر k، می توان با $m+k$ گام از هر state ای به هر state دیگری رفت.
ممکنه با m' ای که کمتر از این کران باشه هم بشه از هر state به هر state دیگه رفت اما با $m'+1$ نشه!



در این مثال به ازای نود 1 به حد پایین این کران می رسیم. یعنی

$$P_{11}^{(M-1)^2} = 0$$

periodic یا transient
با پریود 2



Matrix Representation

$$\underset{1\times M}{[\pi_0]} @ \underset{M\times 1}{[\mathbb{1}]} = 1$$

$$\underset{M\times M}{[P]} @ \underset{M\times 1}{[\mathbb{1}]} = \underset{M\times 1}{[\mathbb{1}]}$$

[ماتریس P حداقل یک مقدار ویژه 1 دارد.
جمع سطرهاش 1 می شود و]
[قدر مطلق سایر مقدار ویژه هاش کمتر مساوی 1
است]

بزرگ ترین المان بردار u را u_k می نامیم $\rightarrow$ $|u_k| \geq |u_j| \;\forall j$

$$(Pu)_k = \lambda u_k$$

$$\rightarrow |\lambda||u_k| = |(Pu)_k| = \left|\sum_{j=1}^{M} P_{kj} u_j\right|$$

$$\leq \sum_{j=1}^{M} P_{kj} \underbrace{|u_j|}_{\leq |u_k|} \leq |u_k| \underbrace{\sum_{j=1}^{M} P_{kj}}_{1} \rightarrow |\lambda| \leq 1$$

$$X_0 \sim \pi_0$$

$$X_1 \sim \pi_0 P = \pi_1 \xrightarrow{\text{proof}} \text{احتمال کل}$$

$$\pi_1(i) = \sum_{j=1}^{M} \pi_0(j) P_{ji}$$

$$\vdots$$

$$X_n \sim \pi_0 P^n = \pi_n$$

Steady state of a Markov chain

steady state vector π^{ss} for an M-state markov chain is a row vector that satisfies $\pi^{ss} = \pi^{ss} P$

⊛ تمایل داریم ببینیم آیا $\lim_{n\to\infty} \pi_0 P^n$ می تونه مستقل از π_0 بشه یا نه. به این منظور باید $\underline{P}$ ویژگی هایی داشته باشه.

⊛ به جای فرم lim از فرم ss استفاده می کنیم

if $\lambda_0 = 1$ and $\forall i \in \{1,\dots,M-1\}$ $|\lambda_i| < 1$

$$P \overset{\text{SVD}}{=\!=} \mathbb{1}\pi^{ss} + \sum_{i=2}^{M} \lambda_i u_i v_i^T \xrightarrow[\text{متعامد کردن } u_i \text{ ها}]{|\lambda_i|<1} \lim_{n\to\infty} P^n = \mathbb{1}\pi^{ss}$$

$$= \begin{bmatrix} -\pi^{ss}- \\ \vdots \\ -\pi^{ss}- \end{bmatrix}$$

$$P\mathbb{1} = \mathbb{1} \rightarrow u_1 = \mathbb{1}$$

$$\pi^{ss} P = \pi^{ss} \rightarrow v_1^T = \pi^{ss}$$

مشاهده شد که $\lim_{n\to\infty} P_{ij}^{n}$ از i و n مستقل شد. مستقل شدن از n به معنی همگرایی است و مستقل شدن از i به معنی مستقل شدن از π_0 است

$$\rightarrow \lim_{n\to\infty} \pi_0 P^n = \underbrace{\pi_0 \mathbb{1}}_{1} \pi^{ss} = \pi^{ss} \checkmark$$

⊛ نوشتن $\underset{M\times M}{[P]}$ به صورت جمع M ماتریس rank1، فرم دیگری از SVD است

⊛ معادله‌ی $\pi^{ss}=\pi^{ss}P$ همیشه، به ازای هر P، حداقل یک جواب را دارد اما ما دوست داریم تنها یک جواب یکتا داشته باشد تا همواره در $n\to\infty$ به صورت unconditionally به یک π^{ss} یکتا همگرا شویم.

⊛ معادله‌ی $\pi^{ss}=\pi^{ss}P$ $\overbrace{\text{یک جواب یکتا}}^{\equiv \text{ استقلال از } \pi_0}$ دارد و $\overbrace{P^n \text{ همگرا}}^{\equiv \text{ استقلال از } n}$ می‌شود اگر:

- شرط لازم و کافی: $1=|\lambda_1|>|\lambda_2|\geq\cdots\geq|\lambda_M|$
- شرط کافی: $\forall i,j \quad P_{ij}>0$ یعنی از هر استیتی به هر استیتی به طور مستقیم بشه رفت

lemma: if $\forall i,j \; P_{ij}>0$ and $\alpha \triangleq \min_{i,j} P_{ij}$

then for all states j and all $n\geq 1$:

$$\max_i P_{ij}^n - \min_i P_{ij}^n \leq (1-2\alpha)^n$$

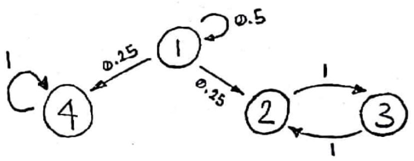
$$\xrightarrow{n\to\infty} P^n = \begin{bmatrix} -\ \pi^{ss}\ - \\ \vdots \\ -\ \pi^{ss}\ - \end{bmatrix}$$

- شرط کافی: markov chain being ergodic

$\xrightarrow{\text{ergodic}}$ for $n\geq (M-1)^2+1 \quad \forall i,j \; P_{ij}^n \geq \beta > 0$

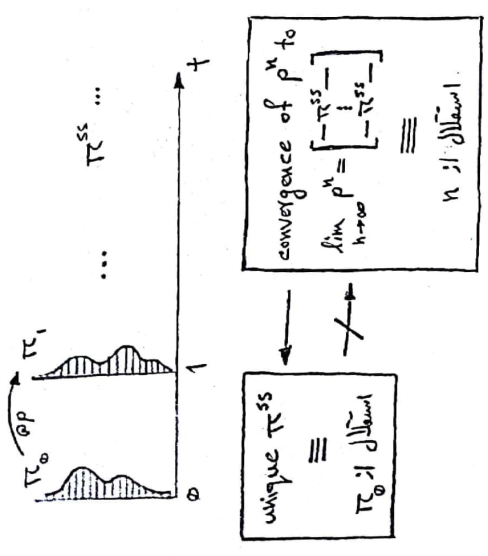
$$\longrightarrow \max_i P_{ij}^n - \min_i P_{ij}^n \leq (1-2\beta)^{\left\lfloor \frac{n}{(M-1)^2+1} \right\rfloor}$$

یک مثال که steady state یکتا ندارد:



$$\begin{bmatrix} 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.5 & 0.25 & 0 & 0.25 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

⊛ periodic بودن همگرایی P^n رو خراب می‌کنه
مثلاً اگر $P=\begin{bmatrix}0&1\\1&0\end{bmatrix}$ آنگاه $\lim_{n\to\infty} P^n$
بین $\begin{bmatrix}0&1\\1&0\end{bmatrix}$ و $\begin{bmatrix}1&0\\0&1\end{bmatrix}$ نوسان می‌کنه.
این در حالی است که π^{ss} یکتا داریم



⊛ به زمانی که طول میکشه تا به steady state برسیم و اثر initialization از بین بره می‌گن mixing time. هرچی spectral gap $=1-|\lambda_2|$ بزرگتر باشه mixing time کمتره.
اولین بزرگترین
دومین بزرگترین

⊛ irreducible MC زنجیره‌ای است که تمام استیت هاش با هم communicate می‌کنند

⊛ در finite state MC: تمام زنجیره یک rec. class است $\longleftrightarrow$ irreducible MC

irreducible MC $\longrightarrow$ جواب π^{ss} یکتا

irreducible + aperiodic $\longrightarrow$ π^{ss} یکتا و همگرایی P^n

$\boxed{\text{Reversibility}}$

a markov chain that has a unique $\pi^{ss}=\lim_{n\to\infty}\pi_0 P^n$ is reversible in time if $\forall i,j$

$$\pi^{ss}(i)\, P_{ij} = \pi^{ss}(j)\, P_{ji}$$

Session 19

1404. 2 .15

$\boxed{\text{Estimation}}$

data $\longrightarrow$ model

$\boxed{\text{parametric Estimation}}$

$$\{X_1, \cdots X_n\} \overset{iid}{\sim} \underset{\mathcal{N}(\mu,\sigma^2)}{P_\theta} \in \underset{\mathcal{N}}{\mathcal{P}} \qquad \underset{\begin{bmatrix}\mu \\ \sigma\end{bmatrix}}{\theta} \in \underset{\mathbb{R}\times\mathbb{R}_+}{\Theta}$$

$\theta^* \in \Theta \qquad X_1, \cdots X_n \overset{iid}{\sim} P_{\theta^*}$

$$\underbrace{\mathcal{P}}_{\text{distribution family}} = \{P_\theta \mid \underbrace{\theta}_{\text{parameter vector}} \in \underbrace{\Theta}_{\text{parameter space}}\}$$

Goal : create an estimator $\underset{\text{R.V.}}{\hat\theta(X_1 \cdots X_n)}$

$$\hat\theta : \underbrace{\bigcup_{i=1}^{\infty} \mathcal{X}^i}_{\text{فضای دیتاست‌ها}} \longrightarrow \underbrace{\Theta}_{\text{فضای پارامترها}}$$

$\boxed{\text{Non-parametric Estimation}}$

⊛ در تخمین پارامتریک خانواده‌ی توزیع ثابت و اندازه‌ی بردار θ محدوده اما در تخمین غیر پارامتریک میشه اندازه‌ی θ رو بی‌نهایت در نظر گرفت

⊛ تخمین پارامتریک مستعد underfitting ـه

⊛ تخمین غیر پارامتریک مستعد overfitting ـه

⊛ فرض این درس اینه که approx. error $= 0$ یعنی:

$X_1, \cdots X_n \overset{iid}{\sim} P_{\theta^*}$

$\theta^* \in \Theta$

⊛ estimation error $\longrightarrow$ ارور $\hat\theta$ نسبت به θ^*

⊛ approximation error $\longrightarrow$ ارور حاصل از انتخاب خانواده بد برای $\mathcal{P}$

⊛ $\hat\theta$ می‌تونه یک تابع deterministic باشه و یا اینکه randomness هم داشته باشه اما در این درس ما با det. est. کار داریم

⊛ $\mathbb{E}_{X_1,\cdots X_n \overset{iid}{\sim} P_{\theta^*}}\left[\hat\theta(X_1 \cdots X_n)\right]$ تابعی است از $n, \mathcal{P}, \Theta, \ldots, \theta^*$

- $\text{Bias}(n,\theta^*) \triangleq \mathbb{E}_X\left[\hat\theta(X_1 \cdots X_n)\right] - \theta^*$

- $\text{Var}(n,\theta^*) \triangleq \text{Var}(\hat\theta) = \mathbb{E}_X\left[(\hat\theta - \mathbb{E}\hat\theta)^2\right]$

- Consistency $\lim_{n\to\infty} \hat\theta(X_1 \cdots X_n) \xrightarrow{\text{I.P. or a.s.}} \theta^*$

⊛ اگر $\theta^*, \hat\theta$ برداری باشند، بایاس و واریانس هم بردار میشن

⊛ بایاس می‌تونه منفی هم بشه. ما بیشتر با $|\text{bias}|$ کار داریم

⊛ بایاس به ازای یک الگوریتم تخمین تعریف میشه نه تک experiment

چند مثال arbitrary از تخمینگر برای میانگین گاوسی با واریانس ثابت:

$$\mathcal{P} = \{\mathcal{N}(\theta, \underset{\text{known}}{\sigma^2}) \mid \theta \in \Theta\}$$

$X_1, \cdots X_n \overset{iid}{\sim} \mathcal{N}(\mu^*, \sigma^2)$

$\hat\mu_1 = \dfrac{X_1 + \cdots + X_n}{n}$ sample mean / empirical mean

$\text{Bias}(n,\mu^*) = \mu^* - \mu^* = 0$

$$\text{var}(n,\mu^*) = \mathbb{E}_X\left[\left(\frac{X_1+\cdots+X_n}{n} - \mu^*\right)^2\right] = \mathbb{E}_X\left[\left(\frac{1}{n}\sum_{i=1}^n (X_i - \mu^*)\right)^2\right] = \frac{1}{n^2}\sum_{i,j=1}^n \mathbb{E}_X\left[(X_i-\mu^*)(X_j-\mu^*)\right]$$

$$= \frac{1}{n^2}\sum_{i=1}^n \mathbb{E}_X\left[(X_i - \mu^*)^2\right] = \frac{\sigma^2}{n}$$

$$\hat{\mu}_2 = X_1$$

$$\text{Bias}(n, \mu^*) = 0$$

$$\text{Var}(n, \mu^*) = \mathbb{E}\left[(X_1 - \mu^*)^2\right] = \sigma^2$$

$$\hat{\mu}_3 = \frac{nX_n + (n-1)X_{n-1} + \cdots + X_1}{\frac{n(n+1)}{2}}$$

$$\text{Bias}(n, \mu^*) = 0$$

$$\text{Var}(n, \mu^*) = \left(\frac{2}{n(n+1)}\right)^2 \overbrace{\left[n^2 \overbrace{\text{Var}(X_n)}^{\sigma^2} + \cdots + \overbrace{\text{Var}(X_1)}^{\sigma^2} \right]}^{\frac{n(n+1)(2n+1)}{6}\sigma^2}$$

$$= \frac{2}{3} \frac{2n+1}{n(n+1)} \sigma^2 \simeq \frac{4}{3} \frac{\sigma^2}{n}$$

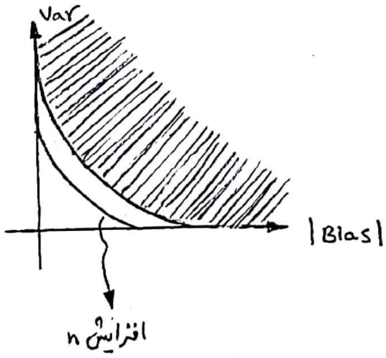
همون empirical mean بهتر بود!

$$\hat{\mu}_4 = 5$$

$$\text{Bias}(n, \mu^*) = 5 - \mu^*$$

$$\text{Var}(n, \mu^*) = 0$$

$\boxed{\text{Bias Variance Dillema}}$



for $\underline{\text{unbiased } \hat{\theta}}$ we have Cramer-Rao lower bound:

$$\text{Var}_{\hat{\theta}}(n, \theta) \geq \frac{1}{n \underbrace{I(\theta)}_{\text{Fisher information}}}$$

$$I(\theta) \triangleq \mathbb{E}_X\left[\left(\frac{\partial \lg f_\theta(X)}{\partial \theta}\right)^2\right]$$

$$= \mathbb{E}_X\left[-\frac{\partial^2}{\partial \theta^2} \lg f_\theta(X)\right] \geq 0$$

bias $\equiv$ estimation error

Var $\equiv$ approximation error

⊛ ضعفِ تخمین پارامتریک اینه که با افزایش بُعد دچار curse of dimensionality نمی‌شیم. مثلاً اگر برای گاوسی یک بُعدی با $O(1)$ سمپل به نتیجه مناسب برسیم، برای گاوسی d بُعدی هم با $\underbrace{O(d^2)}_{\text{تعداد پارامترهای گاوسی d بُعدی}}$ سمپل به نتیجه‌ی مشابه می‌رسیم.

⊛ واریانس بالا $\equiv$ overfitting

Session 20

1404.2.20

Maximum Likelihood Estimation

$$\hat{\theta}_{MLE} \triangleq \underset{\theta \in \Theta}{\operatorname{argmax}} \underbrace{\prod_{i=1}^{n} P_\theta(x_i)}_{\text{likelihood}}$$

⊛ ممکنه LH چندجا peak بزنه که در این صورت argmax هم یک مجموعه شامل همه نقاط رو برمی‌گردونه ولی ما با این حالت کاری نداریم.

⊛ MLE فرض می‌کنه x_i ها مستقل اند ولی اگر نباشند هم میشه ازش استفاده کرد. فقط یکسری از ویژگی هاش رو از دست میده.

- $x_1 \cdots x_n \overset{iid}{\sim} \mathcal{N}(\overset{\text{fix}}{\theta}, \sigma^2)$

$$\hat{\theta}_{MLE} = \underset{\theta \in \mathbb{R}}{\operatorname{argmax}} \prod_{i=1}^{n} f_\theta(x_i) = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} \lg f_\theta(x_i) = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{n} -(x_i-\theta)^2 \overset{\frac{d}{d\theta}=0}{=\!=\!=} \bar{X}_n$$

بایاس و واریانس $\bar{X}_n$ رو حساب کرده بودیم

حالا اگر میانگین fix باشه و واریانس رو با MLE حساب کنیم میشه $\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$

نکته‌ی جالب اینه که

$Bias(\hat{\sigma}^2_{MLE}) = 0$

$Bias(\hat{\sigma}_{MLE}) \neq 0$

- $x_1 \cdots x_n \overset{iid}{\sim} \mathcal{N}(\mu, \overset{\text{unknown}}{\sigma^2})$

$$(\hat{\mu}, \hat{\sigma}) = \underset{\substack{\mu \in \mathbb{R} \\ \sigma > 0}}{\operatorname{argmax}} \sum_{i=1}^{n} \lg \frac{1}{\sigma} - \frac{1}{2\sigma^2}(x_i-\mu)^2 + \text{const}$$

$$\xrightarrow{\frac{d}{d\mu}=0} \hat{\mu}_{MLE} = \bar{X}_n \quad \begin{cases} \text{bias} = 0 \\ \text{var} = \frac{\sigma^2}{n} \end{cases} \quad \text{consistent} \checkmark$$

$$\xrightarrow{\frac{d}{d\sigma}=0} \hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2$$

بررسی شهودی

همیشه بایاس داره و کمتر از σ^2 عه چون اگه $x_1 \cdots x_n$ در دست، مقدار μ متفاوت نباشن $\bar{X}_n$ خودش رو طوری تنظیم می‌کنه که وسط بشه و بنابراین $(x_i-\bar{x}_n)^2$ همواره از $(x_i-\mu)^2$ کمتر مساوی خواهد بود

بررسی ریاضی

با فرض $\mu=0$ برای راحتی

$$\mathbb{E}\left[\hat{\sigma}^2_{MLE}\right] = \underbrace{\mathbb{E}\left[x_i^2\right]}_{\sigma^2} + \underbrace{\mathbb{E}\left[\bar{x}_n^2\right]}_{\frac{\sigma^2}{n}} - 2\underbrace{\mathbb{E}\left[x_i\bar{x}_n\right]}_{\frac{\sigma^2}{n}} = \frac{n-1}{n}\sigma^2$$

$$\longrightarrow Bias(\hat{\sigma}^2_{MLE}) = \frac{-\sigma^2}{n} \neq 0$$

اگر $\hat{\sigma}^2_{MLE}$ رو در $\frac{n}{n-1}$ ضرب کنیم unbiased میشه یعنی:

$\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x}_n)^2$ is unbiased

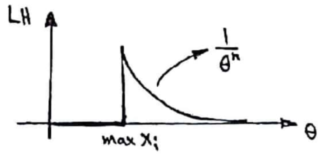
محاسبه‌ی واریانس $\hat{\sigma}^2_{MLE}$ سخت است و در آن moment چهارم ظاهر می‌شود اما اردر آن $O(\frac{1}{n})$ می‌شود

⊛ قضیه ای داریم که می‌گه $g(\hat{\theta}_{MLE}) = \widehat{(g(\theta))}_{MLE}$ اگر g یک به یک باشد.

مثلاً میشه $\hat{\sigma}_{MLE}$ رو حساب کرد بعد با به توان رسوندن به $(\hat{\sigma}_{MLE})^2$ رسید که همون $\widehat{(\sigma^2)}_{MLE}$ میشه!

ولی خب ممکنه بایاس مربوط به تخمین θ و $g(\theta)$ یکی نشن چون g لزوما خطی نیست و $\mathbb{E}$ ازش رد نمیشه!

⊛ برای گاوسی $\hat{\mu}_{MLE} = \bar{X}_n$ بین $\hat{\mu}$ های unbiased کمترین واریانس رو داره



- $X_1 \dots X_n \overset{iid}{\sim} Unif(0,\theta) \quad \theta > 0$

$$\hat{\theta}_{MLE} = \underset{\theta>0}{\operatorname{argmax}} \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{1}(X_i \leq \theta) \overset{*}{=} \max\{X_1 \dots X_n\}$$

⊛ کوچکترین θ ای که هیچ $f_\theta(X_i)$ ای صفر نشه

$$\longrightarrow F_{\hat{\theta}_{MLE}}(x) = F_X(x)^n \longrightarrow f_{\hat{\theta}_{MLE}}(x) = n \underbrace{F_X(x)^{n-1}}_{\left(\frac{x}{\theta}\mathbb{1}(0\leq x\leq\theta)\right)^{n-1}} f_X(x) = \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1} \quad 0 \leq x \leq \theta$$

$$\longrightarrow Bias(n,\theta) = \int_0^\theta \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1} x \, dx - \theta = \frac{n\theta}{n+1} - \theta = \frac{-\theta}{n+1}$$

یعنی همیشه under estimate می‌کنه

⊛ اگه تعریف کنیم $\hat{\theta}_{custom} \triangleq \frac{n+1}{n}\hat{\theta}_{MLE}$ اون وقت unbiased می‌شه

- $X_1 \dots X_n \overset{iid}{\sim} Exp(\lambda) \quad \lambda > 0$

$$\hat{\lambda}_{MLE} = \underset{\lambda>0}{\operatorname{argmax}} \prod_{i=1}^{n} \lambda e^{-\lambda X_i} = \underset{\lambda>0}{\operatorname{argmax}} \; n \lg \lambda - \lambda \sum_{i=1}^{n} X_i$$

$$\xrightarrow{\frac{d}{d\lambda}=0} \hat{\lambda}_{MLE} = \frac{n}{X_1 + \dots + X_n} \longrightarrow \hat{\mu}_{MLE} = \frac{1}{\hat{\lambda}_{MLE}} = \bar{X}_n$$

session 21
1404.2.22

$\boxed{\text{Cramer - Rao Lower Bound and Fisher Information}}$

if $x_1 \dots x_n \overset{iid}{\sim} f_\theta(x)$ and for $\forall \hat\theta$ which is unbiased then $\text{Var}_{\hat\theta}(n,\theta) \geq \frac{1}{nI(\theta)}$ if $0 < I(\theta) < \infty$

$\boxed{\text{proof for } n=1}$

$$\mathbb{E}\hat\theta - \theta = 0 \longrightarrow \int(\hat\theta-\theta)f_\theta(x)\,dx = 0 \xrightarrow{\frac{d}{d\theta}} \int(\hat\theta-\theta)\frac{df_\theta(x)}{d\theta}dx - \underbrace{\int f_\theta(x)\,dx}_{1} = 0 \quad \text{I}$$

$$\frac{d}{d\theta}\lg f_\theta(x) = \frac{f'_\theta(x)}{f_\theta(x)} \longrightarrow \frac{d f_\theta(x)}{d\theta} = f_\theta(x)\frac{d}{d\theta}\lg f_\theta(x) \quad \text{II}$$

$$I(\theta) \triangleq \mathbb{E}_x\left[\left(\frac{\partial \lg f_\theta(x)}{\partial\theta}\right)^2\right] = \mathbb{E}_x\left[-\frac{\partial^2}{\partial\theta^2}\lg f_\theta(x)\right] \geq 0$$

$$\xrightarrow[\text{in I}]{\text{sub II}} \int(\hat\theta-\theta)f_\theta(x)\frac{d}{d\theta}\lg f_\theta(x)\,dx = 1$$

$$\xrightarrow[\text{schwartz}]{\text{cauchy}} \underbrace{\left(\int(\hat\theta-\theta)\sqrt{f_\theta(x)}\sqrt{f_\theta(x)}\frac{d}{d\theta}\lg f_\theta(x)\,dx\right)^2}_{1} \leq \underbrace{\int(\hat\theta-\theta)^2 f_\theta(x)\,dx}_{\text{Var}_{\hat\theta}(1,\theta)} \cdot \overbrace{\int f_\theta(x)\left(\frac{d}{d\theta}\lg f_\theta(x)\right)^2 dx}$$

⊛ Fisher Information برای یک خانواده از توزیع ها مثل f_θ تعریف میشه و داریم $I(\theta)\in[0,\infty)$. اگر $I(\theta)=0$ or ∞ کران CR معنی نداره !

⊛ معنی $\frac{\partial}{\partial\theta}$ در $I(\theta)$ ⟵ اگر $f_\theta(x)$ نسبت به θ حساس باشه $\frac{\partial \lg f_\theta(x)}{\partial\theta}$ هم بیشتر میشه و $I(\theta)$ هم بیشتر میشه.
در حقیقت $I(\theta)$ حساسیت توزیع نسبت به پارامتر هاش رو نشون میده.

⊛ if $\theta\in\mathbb{R} \longrightarrow I(\theta)\in\mathbb{R}_{\geq 0}$
if $\theta\in\mathbb{R}^k \longrightarrow I(\theta)\in\mathbb{R}^{k\times k} \rightsquigarrow$ کار ما نیست

$\boxed{\text{MLE properties}}$

- always consistent
- if $x_1 \dots x_n$ are iid from f_θ it has asymptotic normality (gaussianity)

i.e. $\sqrt{n}\left(\hat\theta_{MLE}(x_1\dots x_n) - \theta\right) \xrightarrow[n\to\infty]{\text{I.D.}} \mathcal{N}\left(0, \frac{1}{I(\theta)}\right)$

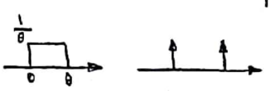
⊛ $\hat\theta_{MLE}$ لزوما unbiased نیست اما همواره در حد unbiased میشه

$\xrightarrow{\text{corollary}} \lim_{n\to\infty} \text{Var}_{\hat\theta_{MLE}}(n,\theta) = \frac{1}{nI(\theta)}$

یعنی $\hat\theta_{MLE}$ همواره در حد کمترین واریانسی که CR میگه رو میگیره
و به اصطلاح asymptotically CR efficient هست

⊛ برقراری asymp. normality خاصیت consistency رو نتیجه میده چون داره میگه $\hat\theta_{MLE}-\theta$ با نرخ حداقل $\frac{1}{\sqrt{n}}$ داره کم میشه.

حساس به پارامتر

e.g. 

e.g. 

⊛ اگر $I(\theta)=0$ or ∞ خاصیت asymp. normality رو نداریم

$I(\theta)=\infty \overset{\text{always}}{\longleftrightarrow} \exists\hat\theta \ \ \text{Var}(\hat\theta) < O(\frac{1}{n})$ e.g. $O(\frac{1}{n^2})$

$I(\theta)=0 \underset{\text{usualy}}{\longleftrightarrow} \exists\hat\theta \ \ \text{Var}(\hat\theta) > O(\frac{1}{n})$ e.g. $O(\frac{1}{\sqrt{n}})$

$\downarrow$ unblased

⊛ (MVUE) minimum variance unbiased estimator تخمینگری است که بین unbiased ها کمترین واریانس رو داره

⊛ برای بعضی از مسائل MVUE شون همون minimum ای که CR میگه اما برای تمام مسائل این طور نیست

⊛ به یک تخمینگر میگن CR efficient اگر به ازای یک n خاص unbiased باشه و واریانسش هم کمترینی که CR میگه باشه

Fisher Information of $P = \{\mathcal{N}(\mu, \overset{\text{fix}}{\sigma^2}) \mid \mu \in \mathbb{R}\}$

هرچی واریانس گاوسی بیشتر بشه حساسیت به μ کمتر میشه

$$I(\mu) = \mathbb{E}_x\left[\left(\frac{\partial}{\partial \mu} \lg\left(\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right)\right)^2 \,\middle|\, \mu, \sigma^2\right] = \mathbb{E}_x\left[\left(\frac{\mu - x}{\sigma^2}\right)^2\right] = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}$$

$$\longrightarrow \text{Var}_{\hat{\mu}}(n, \mu) \geq \frac{\sigma^2}{n}$$

Fisher Information of $P = \{\text{Unif}(0, \theta) \mid \theta > 0\}$

چون ساپورت توزیع به پارامتر توزیع (θ) وابسته است
اصلا Fisher inf. برای توزیع Unif تعریف نمیشه!

assume unbiased estimator $\hat{\theta}_{custom} = \frac{n+1}{n} \max_i x_i$

$$\text{Var}(\hat{\theta}_{custom}) = \mathbb{E}[\hat{\theta}^2] - \theta^2$$

$$= \underbrace{\left(\frac{n+1}{n}\right)^2 \int_0^\theta \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1} x^2 \, dx}_{\frac{(n+1)^2 \theta^2}{n^2(n+2)}} - \theta^2 = \frac{\theta^2}{n(n+2)} = O\left(\frac{1}{n^2}\right)$$

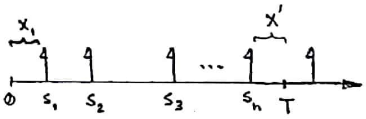
چون تخمینگر unbiased ای وجود دارد که
واریانس آن سریع تر از $O(\frac{1}{n})$ می افتد
پس می توان گفت $I(\theta) = \infty$

⊛ در حقیقت برای $\text{Unif}(0,\theta)$ تخمینگر CR efficient نداریم چون $I(\theta) = \infty$.
همین $\hat{\theta}_{custom}$ تخمینگر MVUE هه

Estimating the parameter of a poisson process

Session 22
1404.2.27

given a snapshot of a poisson process with length T, estimate its λ



$\boxed{\text{راه 1}}$ ‏$n+1$ تا سمپل داریم ($X_1, \dots X_n$ های inter arrival و X' که غایب هستند)

$$\hat{\lambda}_{MLE} = \underset{\lambda>0}{\operatorname{argmax}}\ f_{snapshot(T)} = \underset{\lambda>0}{\operatorname{argmax}} \prod_{i=1}^{n} \lambda e^{-\lambda X_i}\ e^{-\lambda X'}$$

$$= \underset{\lambda>0}{\operatorname{argmax}}\ \lambda^n e^{-\lambda(\sum_{i=1}^{n} X_i + X')} = \underset{\lambda>0}{\operatorname{argmax}}\ \lambda^n e^{-\lambda T} = \frac{n}{T}$$

- $\text{Bias}(\hat{\lambda}_{MLE}) = \frac{\mathbb{E}[n]}{T} - \lambda$

$$= \frac{\lambda T}{T} - \lambda = 0$$

- $\text{Var}(\hat{\lambda}_{MLE}) = \frac{\text{Var}(n)}{T^2}$

$$= \frac{\lambda T}{T^2} = \frac{\lambda}{T}$$

$\boxed{\text{راه 2}}$ یک سمپل داریم (n که از توزیع $\text{poisson}(\lambda T)$ میاد)

$$\hat{\lambda}_{MLE} = \underset{\lambda>0}{\operatorname{argmax}}\ f_{snapshot(T)} = \underset{\lambda>0}{\operatorname{argmax}}\ \frac{(\lambda T)^n}{n!} e^{-\lambda T}$$

$$= \underset{\lambda>0}{\operatorname{argmax}}\ \text{const} \times \lambda^n e^{-\lambda T} = \frac{n}{T}$$

$$\text{fisher inf.}\Big(\overbrace{f_{snapshot(T)}}^{\lambda^n e^{-\lambda T}}\Big) = \mathbb{E}_n\left[-\frac{\partial^2}{\partial\lambda^2}\left(n\lg\lambda - \lambda T\right)\right]$$

$$= \mathbb{E}_n\left[\frac{n}{\lambda^2}\right] = \frac{\lambda T}{\lambda^2} = \frac{T}{\lambda} \rightsquigarrow$$

⊛ در این مسئله n , X_i ها تصادفی هستند

مشاهده می‌شود که $\text{var}(\hat{\lambda}_{MLE}) = \frac{1}{1 \times I(\lambda)}$
بنابراین تخمینگر ارائه شده CR efficient است

⊛ پارامتر λ فقط در به دست آمدن n اثر داره و بعدش condition شده روی n، X_i ها از $\text{Unif}(0,T)$ میان

⊛ n اینجا با n ای که قبلاً برای تعداد سمپل ها استفاده می‌کردیم فرق داره.
اینجا تعداد سمپل ها ($\equiv$ تعداد فرایند) برابر 1 عه.
برای همین n در $\text{Var}(\hat{\lambda}_{MLE})$ ظاهر نشد.

Sufficient statistic

For parametric estimation

a statistic is a function of random samples $X_1, \dots X_n$ e.g. $\bar{X}_n$, $\max_i X_i$, n

- T is a sufficient statistic if by knowing T we can do as good a job of estimating θ as by knowing the entire random sample $X_1 \dots X_n$

- a statistic $T = r(X_1 \dots X_n)$ is a sufficient statistic if

$$\forall t \quad P(\underline{X} \mid T=t, \theta) = P(\underline{X} \mid T=t)$$

$$\theta \longrightarrow T(\underline{X}) \longrightarrow \underline{X}$$

$$P = \{\mathcal{N}(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma > 0\} \longrightarrow T(\underline{X}) = \left(\sum X_i , \sum X_i^2\right) \text{ or } \left(\text{sample mean}, \text{sample var}\right)$$

$$P = \{\text{Unif}(a,b) \mid a < b \in \mathbb{R}\} \longrightarrow T(\underline{X}) = \left(\min_i X_i , \max_i X_i\right)$$

$$P = \{\text{poisson process}(\lambda) \mid \lambda > 0\} \longrightarrow T(\underline{X}) = n$$

⊛ suff. stat. یکتا نیست

⊛ suff. stat. ویژگی خانواده‌ی توزیع است و به روش تخمین ربطی ندارد!

Fisher - Neyman Factorization Theorem

let $\underline{X}$ be a random sample with joint density $f_\theta(\underline{X})$, then T is sufficient for θ iff

$$\exists\, g, h \geq 0 \quad f_\theta(\underline{X}) = h(\underline{X})\, g_\theta(T(\underline{X}))$$

$\rightsquigarrow$ h must not depend on θ

$\rightsquigarrow$ g can depend on $\underline{X}$ only through $T(\underline{X})$

$$\underline{X} \overset{iid}{\sim} \mathcal{N}(\underset{\text{fix}}{\mu}, \sigma^2) \longrightarrow f_\mu(x) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \prod_{i=1}^{n} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$$

$$= \underbrace{\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\right]}_{h(\underline{x})} \cdot \underbrace{\exp\left[\frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_i\right)\mu - \frac{n\mu^2}{2\sigma^2}\right]}_{g_\mu(\underbrace{\sum x_i}_{T(\underline{X})})}$$

snapshot

$$\text{poisson process snapshot } (T) \longrightarrow f_\lambda(s) = \underbrace{f(\underline{X}|n)}_{h(\underline{X})}\ \underbrace{P_{\lambda T}(n)}_{g_\lambda(\underbrace{n}_{T(\underline{X})})}$$

poisson(λT) Unif$(0,T)$

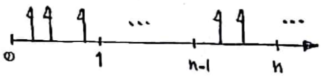
Rao - Blackwell's Theorem

let $\hat\theta$ be an estimator of θ with $\mathbb{E}_x[\hat\theta^2] < \infty$ for all θ. suppose T is sufficient for θ and let $\tilde\theta \triangleq \mathbb{E}_x[\hat\theta(\underline{X}) \mid T(\underline{X})]$ then

$$\forall\theta \quad MSE(\tilde\theta, \theta) = \mathbb{E}_x[(\tilde\theta-\theta)^2] \leq \mathbb{E}_x[(\hat\theta-\theta)^2] = MSE(\hat\theta, \theta)$$

$$\xrightarrow{\text{for unbiased } \hat\theta} Var(\tilde\theta) \leq Var(\hat\theta)$$

برای مثال فرض کنیم یک فرایند پواسون به صورت



داده شده و سمپل های $X_1 \dots X_n$ را داریم که X_i تعداد arrival های بازه‌ی $(i-1, i)$ را مشخص می‌کند.

یک تخمینگر بی‌خود برای $\theta = e^{-\lambda}$ را به صورت $\hat\theta \triangleq \mathbb{1}(X_1=0)$ تعریف می‌کنیم.

$$\mathbb{E}\hat\theta = \mathbb{P}(X_1=0) = \frac{(\lambda)^0}{0!}e^{-\lambda} \longrightarrow \text{unbiased}$$

می‌دانیم که suff. stat. برای λ و همچنین $e^{-\lambda}$ برابر تعداد arrival ها (یعنی $\sum X_i$) است.

$$\tilde\theta = \mathbb{E}_x\left[\hat\theta \,\middle|\, \sum_{i=1}^{n} X_i = S_n\right] = \mathbb{P}\left(X_1=0 \,\middle|\, \sum_{i=1}^{n} X_i = S_n\right)$$

$$= \frac{\mathbb{P}(X_1=0)\ \mathbb{P}\left(\sum_{i=2}^{n} X_i = S_n\right)}{\mathbb{P}\left(\sum_{i=1}^{n} X_i = S_n\right)} = \frac{e^{-\lambda}\ \frac{((n-1)\lambda)^{S_n}}{S_n!}\ e^{-(n-1)\lambda}}{\frac{(n\lambda)^{S_n}}{S_n!}\ e^{-n\lambda}} = \left(1-\frac{1}{n}\right)^{S_n}$$

به تخمینگر بهتری رسیدیم

$$\lim_{n\to\infty} \tilde\theta = \lim_{n\to\infty}\left(1-\frac{1}{n}\right)^{S_n} \simeq \lim_{n\to\infty}\left(1-\frac{1}{n}\right)^{n\lambda} = e^{-\lambda} \longrightarrow \tilde\theta \text{ is consistent}$$

$$\mathbb{E}\tilde\theta = \mathbb{E}_x\left[\left(1-\frac{1}{n}\right)^{S_n}\right] \overset{S_n \sim \text{poisson}(n\lambda)}{=\!=\!=} \sum_{k=0}^{\infty}\left(1-\frac{1}{n}\right)^k \frac{(\lambda n)^k}{k!} e^{-\lambda n} = e^{-\lambda n} \underbrace{\sum_{k=0}^{\infty} \frac{[\lambda(n-1)]^k}{k!}}_{e^{\lambda(n-1)}} = e^{-\lambda}$$

$\longrightarrow$ $\tilde\theta$ is unbiased

اگر از اول به جای $\hat{\theta} = \mathbb{1}[X_1 = 0]$ با MLE مسئله رو حل می‌کردیم:

$$\left(\widehat{e^{-\lambda}}\right)_{MLE} = e^{-\hat{\lambda}_{MLE}} = e^{-\frac{S_n}{n}}$$

$$\lim_{n\to\infty} \hat{\theta}_{MLE} = \lim_{n\to\infty} e^{-\frac{S_n}{n}} = e^{-\frac{\lambda n}{n}} = e^{-\lambda} \longrightarrow \hat{\theta}_{MLE} \text{ is consistent}$$

$$\mathbb{E}\,\hat{\theta}_{MLE} = \mathbb{E}_x\left[e^{-\frac{S_n}{n}}\right] = \sum_{k=0}^{\infty} e^{-\frac{k}{n}} \frac{(\lambda n)^k}{k!} e^{-\lambda n} = e^{-\lambda n} \underbrace{\sum_{k=0}^{\infty} \frac{\left[e^{-\frac{1}{n}} \lambda n\right]^k}{k!}}_{e^{e^{-\frac{1}{n}}\lambda n}} = e^{\lambda \overbrace{(n e^{-\frac{1}{n}} - n)}^{=-1+o(\frac{1}{n})}} = e^{-\lambda} e^{\lambda o(\frac{1}{n})}$$

$\longrightarrow \hat{\theta}_{MLE}$ is biased

$\boxed{\text{Bayesian Estimation}}$

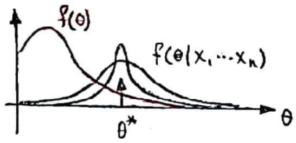
Session 23
1404.2.29

$$\begin{cases} \theta^* \sim f_\theta(\theta) \rightsquigarrow \text{prior distribution} \\ x_1 \cdots x_n \mid \theta^* \overset{iid}{\sim} f_{\theta^*}(x) \rightsquigarrow \text{likelihood} \end{cases}$$

$\longrightarrow f(\theta \mid x_1 \cdots x_n) \rightsquigarrow$ posterior distribution

$\boxed{n \simeq 1}$ $f(\theta \mid x_1 \cdots x_n) \simeq f(\theta)$

$\boxed{n \to \infty}$ $f(\theta \mid x_1 \cdots x_n) \longrightarrow \delta(\theta - \theta^*)$



⊛ در $n \to \infty$ فقط evidence مهمه و prior بی اثر میشه

$$p \sim \text{Unif}(0,1) = \mathbb{1}(0 \leq p \leq 1)$$

$$x_1 \cdots x_n \mid p \overset{iid}{\sim} \text{Bern}(p) = \prod_{i=1}^{n} (1-p)^{1-x_i} p^{x_i}$$

$$\longrightarrow f_{p \mid x_1 \cdots x_n} = \frac{\prod (1-p)^{1-x_i} p^{x_i} \ \mathbb{1}(0 \leq p \leq 1)}{\int_0^1 \underbrace{\prod (1-p)^{1-x_i} p^{x_i}}_{(1-p)^{n-\sum x_i} p^{\sum x_i}} dp} = \frac{(1-p)^{n-\sum x_i} p^{\sum x_i}}{g(n, \sum x_i)}$$

اینجا به طور خاص مخرج قابل محاسبه بود چون داریم:

$$\int_0^1 p^\alpha (1-p)^\beta dp = \frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)}$$

ولی در کل محاسبه‌ی مخرج برای کامل به دست آوردن posterior کار سختی است.

برای به دست آوردن $\hat{\theta}$ از توزیع $f(\theta \mid x_1 \cdots x_n)$ کارهای مختلفی میشه کرد:

- $\hat{\theta}_{MAP} \triangleq \underset{\theta}{\text{argmax}}$ posterior $\xrightarrow{\text{خوبی}}$ نیاز به محاسبه‌ی مخرج نیست چون ترمِ θ نداره و در مشتق گیری حذف میشه
  - بدی: بعضی مواقع فرمول توزیع رو نداریم که max ش رو حساب کنیم
  - توزیعی مثل نمایی argmax ش 0 عه و پارامتر توزیع کلا ignore میشه

- $\hat{\theta}_{MP} \triangleq \mathbb{E}[\theta \mid x_1 \cdots x_n]$ $\xrightarrow{\text{خوبی}}$ مواقعی که توزیع رو نداریم ولی ازش sample داریم می‌تونیم $\bar{X}_n$ رو به دست بیاریم و پشت‌مون به LLN گرم باشه
  - بدی: اگر توزیع θ گسسته باشه ممکنه چیزی خروجی بده که در ساپورت θ نیست

⊛ در روش های تخمین bayesian خود θ تصادفی فرض میشه بنابراین variance , bias هم R.V. هستند و بنابراین $E_\theta[\text{bias}]$, $E_\theta[\text{Var}]$ رو حساب می‌کنیم.

⊛ consistency هم در روش های bayesian تعریف متفاوتی دارد زیرا باید حد یک R.V. گرفته شود.

⊛ دقت شود که سمپل ها given θ از هم مستقل اند بنابراین: $f(x_1 \cdots x_n) \neq \prod_{i=1}^{n} f(x_i)$
اصلاً برای همین مخرج رو باید از طریق marginalization به دست بیاریم.

Maximum A posteriori

$$\hat{\theta}_{MAP} \triangleq \underset{\theta}{\operatorname{argmax}}\ f(\theta|x_1 \cdots x_n) = \underset{\theta}{\operatorname{argmax}} \frac{f(x_1 \cdots x_n|\theta)\ f(\theta)}{f(x_1 \cdots x_n)} = \underset{\theta}{\operatorname{argmax}}\ \lg f(\theta) + \sum_{i=1}^{n} \lg f(x_i|\theta)$$

$$\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$$

$$x_1 \cdots x_n | \mu \overset{iid}{\sim} \mathcal{N}(\mu, \overset{\text{fix}}{\sigma^2})$$

$$\rightarrow \hat{\mu}_{MAP} = \underset{\mu \in \mathbb{R}}{\operatorname{argmax}}\ \lg\left(\frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}}\right) + \sum \lg\left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\right)$$

$$= \underset{\mu}{\operatorname{argmin}} \left(\frac{\mu-\mu_0}{\sigma_0}\right)^2 + \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2 \xrightarrow{\frac{d}{d\mu}=0} \hat{\mu}_{MAP} = \frac{\frac{1}{\sigma_0^2}\mu_0 + \frac{n}{\sigma^2}\left(\frac{1}{n}\sum x_i\right)}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}$$

به میانگین وزن دار μ_0 و $\bar{X}_n$ رسیدیم که با n بیشتر وزن $\bar{X}_n$ بیشتر می‌شود

⊛ در initialization وزن های شبکه عصبی اگر توزیع پیشین وزن ها $(f(\theta))$ را توزیع لاپلاس فرض کنیم، وزن ها sparse تر می‌شوند:

$$\theta_i \overset{iid}{\sim} \text{Laplace}(\lambda) = \frac{\lambda}{2} e^{-\lambda|\theta_i|}$$

$$\hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}}\ \underbrace{\lg\left(\prod_{j=1}^{Lw^2} \frac{\lambda}{2} e^{-\lambda|\theta_j|}\right)}_{-\lambda\sum|\theta_j| \ \approx \ \text{regularization term}} + \sum_{i=1}^{n} \lg f(x_i|\underline{\theta}) \equiv \text{Lasso Regression}$$

$$\theta_i \overset{iid}{\sim} \mathcal{N}(0,1)$$

$$\hat{\theta}_{MAP} = \underset{\theta}{\operatorname{argmax}}\ \underbrace{\lg\left(\prod_{j=1}^{Lw^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{(\theta_j)^2}{2}}\right)}_{-\frac{1}{2}\sum\theta_j^2 \ \approx \ \text{regularization term}} + \sum_{i=1}^{n} \lg f(x_i|\underline{\theta}) \equiv \text{Ridge Regression}$$

Mean posterior

$$\hat{\theta}_{MP} \triangleq \mathbb{E}[\theta|x_1 \cdots x_n] = \int \theta f(\theta|x_1 \cdots x_n)\, d\theta = \frac{\int \theta \left(\prod_{i=1}^{n} f(x_i|\theta)\right) f(\theta)\, d\theta}{\int \left(\prod_{i=1}^{n} f(x_i|\theta)\right) f(\theta)\, d\theta}$$

⊛ دقت شود که استفاده از توزیع prior تنها در صورتی باعث بهبود می‌شود که source اطلاعاتی آن از data جدا باشد.

⊛ تنها MLE بود که برایش گارانتی هایی وجود داشت.

Conjugate prior

if given a likelihood function $f(X|\theta)$, the posterior distribution $f(\theta|X)$ is in the same probability distribution family as the prior $f(\theta)$, the prior is called a conjugate prior w.r.t that likelihood function.

⊛ خوبی conjugate prior اینه که posterior فرم بسته و از پیش دانسته ای پیدا می‌کنه و محاسبات بسیار ساده‌تر می‌شوند. مثلاً MAP , MP رو میشه یک ضرب از روی mode , mean توزیع به دست آورد.

⊛ اگر به جای توزیع پیشین $f(\theta)$ تنها آماره هایی از آن داشتیم، بهتره خانواده‌ی توزیع $f(\theta)$ رو conjugate prior متناسب با likelihood در نظر بگیریم که محاسبات مون راحت بشه!

⊛ گویا به ازای هر likelihood ای یک جفت conjugate distribution وجود داره!

$$p \sim \text{Beta}(\alpha, \beta)$$

$$X_1 \cdots X_n \mid p \overset{iid}{\sim} \text{Bern}(p)$$

$$\rightarrow p \mid X_1 \cdots X_n \sim \text{Beta}\left(\alpha + \sum X_i - 1 ,\ \beta + n - 1 - \sum X_i\right)$$

$\boxed{\text{proof}}$
$$f(p|\underline{X}) = \frac{f(\underline{X}|p)\, f(p)}{f(\underline{X})} = \frac{\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)} \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}}{\int_0^1 \text{numerator}\ dp}$$

$$= \frac{p^{\alpha+\sum X_i - 1}\,(1-p)^{\beta+n+1-\sum X_i}}{\quad} = \text{Beta}(\alpha', \beta')$$

$$\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$$

$$X_1 \cdots X_n \mid \theta \overset{iid}{\sim} \mathcal{N}(\theta, \overset{\text{fix}}{\sigma^2})$$

$$\rightarrow \theta \mid X_1 \cdots X_n \sim \mathcal{N}\left(\frac{\mu_0/\sigma_0^2 + n/\sigma^2\, \bar{X}_n}{1/\sigma_0^2 + n/\sigma^2} ,\ \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}\right)^{-1}\right)$$

$$\theta \sim \text{Inverse Gamma}(\alpha, \beta)$$

$$X_1 \cdots X_n \mid \theta \overset{iid}{\sim} \mathcal{N}(\overset{\text{fix}}{\mu}, \theta)$$

$$\rightarrow \theta \mid X_1 \cdots X_n \sim \text{Inverse Gamma}\left(\alpha + \frac{n}{2} ,\ \beta + \frac{\sum (x_i - \mu)^2}{n}\right)$$

Hidden Markov Model (HMM)

extra

Session 26

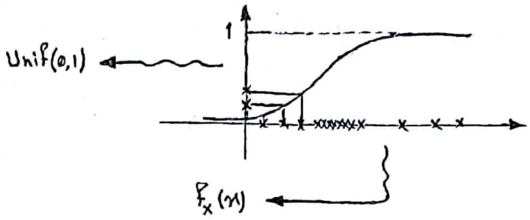
1403.3.10

Sampling

⊛ یک راه تقریبی برای سمپل گرفتن از توزیع $f_X(x)$ اینه که ساپورت رو به N تا bin تقسیم کنیم و احتمال رخ دادن هر bin رو معادل مساحت زیر f_X در اون بازه در نظر بگیریم

Inverse Transform Sampling

$$Z \sim \text{Unif}(0,1) \xrightarrow[\text{change of variable}]{} F_X^{-1}(Z) \sim f_X(x)$$



⊛ این روش دقیق است اما چالش های دارد

این روش برای سمپل گیری تک بُعدی است و تعمیم آن به توزیع های چند بُعدی سرراست نمی باشد. به بیان دقیق تر در ابعاد بالا وارون CDF فرم استانداردی ندارد و همچنین correlation متغیرها در این روش capture نمی‌شود.

برای بسیاری از توزیع ها (مانند گاوسی) فرم بسته ای برای تابع CDF و یا وارون آن وجود ندارد و بنابراین باید از روش های عددی برای تقریب آن ها استفاده شود. این کار به طور خاص در ابعاد بالا بسیار سخت و غیر بهینه است.

⊛ برای سمپل گرفتن از توزیع J.G. با d بُعد میشه d بار با روش inverse transform sampling از $\mathcal{N}(0,1)$ سمپل گرفت و $\Sigma^{\frac{1}{2}}\begin{bmatrix}\circ\\ \vdots\\ \circ\end{bmatrix}_{d\times 1} \sim \mathcal{N}(0,\Sigma)$
همچنین میشه d بار به صورت sequential ، $x_j \mid x_{j-1\dots1} \sim \mathcal{N}$ رو سمپل گرفت و دونه دونه در بردار چید.

Markov chain Monte Carlo (MCMC)

روش های MCMC ساپورت رو می‌گیرن و طوری p رو مهندسی می‌کنن که $P_{target} = \pi^{ss}$
سپس چون MC داریم راحت مرحله به مرحله سمپل می‌گیریم تا به سمپل های π^{ss} برسیم.

- نودهای MC $\equiv$ ساپورت توزیع
- p (یال ها) و π_0 $\equiv$ توزیع چیده شده روی این ساپورت
- به دلیل خاصیت مارکو سمپل گرفتن مرحله به مرحله از MC رو بلدیم
- میشه یال ها (p) رو طوری مهندسی کرد که:
  - π^{ss} یکتا و مستقل از π_0 داشته باشیم
  - π^{ss} زنجیره همان توزیع target شود

support = $\{1, \dots, m\}$ ① ... ⓜ

support = $\{0,1\}^d$ $\quad \begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix}_d \dots \begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix}_d$ → sample X_i

⋮

$$P_{target} = P_{target}\, T \longleftarrow \underbrace{P_{target}(a)\, T(a\to b) = P_{target}(b)\, T(b\to a)}_{\text{detailed balance condition AKA reversibility}}$$

⊛ عنوان MCMC چتری است که روش های متنوعی ذیل آن تعریف می‌شود

⊛ روش های MCMC به درد مسائلی می‌خورن که ساختار (ذات) ساده دارن ولی فرم explicit توزیع اجازه‌ی استفاده از این سادگی رو نمیده

⊛ mixing time برای رسیدن به توزیع target و waiting time برای شبه مستقل کردن نمونه هاست

$$\underbrace{x_0 \cdots x_{n-1}}_{\text{mixing time}},\ x_n \underbrace{\dots}_{\text{waiting time}} x_{n+k} \underbrace{\dots}_{\text{waiting time}} x_{n+2k} \dots$$

session 27

1404.3.12

Ising Model

$$P(x_1, \dots x_d) = \frac{e^{-\sum_{i,j=1\,(i\neq j)}^{d} \theta_{ij}\, x_i x_j}}{\gamma(\theta)} \qquad x_i \in \{\pm 1\}$$

$$\underbrace{\gamma(\theta) = \sum_{2^d \text{ حالت}} e^{-\frac{1}{2}\sum_{i,j} x_i x_j \theta_{ij}}}_{\text{partition function}} \quad , \quad \theta \equiv \Sigma^{-1} \ \text{precision matrix}$$

⊛ روش های MCMC برای حالتی که ماهیت توزیع پیوسته باشد نیز جواب می دهند و داریم:

$$\pi^{ss} = \pi^{ss} P$$

$$\longrightarrow \pi^{ss}(x) = \int \pi^{ss}(x')\, Tp(x|x')\, dx'$$

⊛ وقتی می‌گیم توزیع رو به صورت explicit نداریم یعنی بخشی‌ش رو داریم. مثلاً partition function ش رو نداریم.

Gibbs Sampling

منظور از x_{-i} تمام بردار x به غیر از المان i ام آن است

$x = (x_1, \dots x_d)$ choose i at random or sequentially, then

$$x'_{-i} = x_{-i} \quad , \quad x'_i \sim \underbrace{\pi(x_i | x_{-i})}_{\text{target distribution}}$$

⊛ این روش به درد توزیع های نمایی مانند Ising می‌خوره که محاسبه‌ی $\underbrace{p(\underline{x})}_{d \text{ بُعدی}}$ سخته اما محاسبه‌ی $\underbrace{p(x_i|x_{-i})}_{1 \text{ بُعدی}}$ راحته

if $\pi = P_{Ising}$

$$\longrightarrow \pi(x_i|x_{-i}) = \frac{P(x_i, x_{-i})}{P(x_{-i})} = \frac{P(x)}{P(x_{-i})} = \frac{\frac{1}{\gamma(\theta)} e^{-\sum_{r,t} x_r x_t \theta_{rt}}}{\frac{1}{\gamma(\theta)} \underset{(x_i=1)}{e^{-\sum_{r,t} x_r x_t \theta_{rt}}} + \frac{1}{\gamma(\theta)} \underset{(x_i=-1)}{e^{-\sum_{r,t} x_r x_t \theta_{rt}}}}$$

$$= \frac{e^{-\sum_j x_i x_j (\theta_{ij} + \theta_{ji})}}{e^{\sum_j x_j(\theta_{ij}+\theta_{ji})} + e^{-\sum_j x_j (\theta_{ij}+\theta_{ji})}} = \frac{e^{-\alpha x_i}}{e^{\alpha} + e^{-\alpha}} \rightsquigarrow \text{partition function حذف شد}$$

Metropolis - Hastings

$\pi(x) = A f(x) \longrightarrow$ target distribution

assume we have a good proposal function $q(x'|x)$ (d بُعدی)
such that $q(x'|x)/q(x|x')$ can be easily computed. then

مثلاً در حالت یک بُعدی میشه گفت:

$$q(x_{t+1}|x_t) \triangleq \mathcal{N}(x_t, \sigma^2)$$

$$x' \sim q(\cdot|x_t) \longrightarrow \begin{cases} \text{accept } x' \text{ with probability } \min\left(1, \frac{\pi(x')\, q(x|x')}{\pi(x)\, q(x'|x)}\right) \text{ and } x_{t+1} = x' \\ \text{reject } x' \text{ otherwise and } x_{t+1} = x_t \end{cases}$$

Hamiltonian Monte-Carlo (HMC)

$\pi(x) \propto e^{U(x)}$

$p \sim \mathcal{N}(0, \Sigma) \longrightarrow$ momentum

$k(p) = \frac{1}{2} p^T \Sigma^{-1} p \longrightarrow$ kinetic energy

$-\nabla_x U(x) =$ score function

$U(x) + k(p) = H(x,p) =$ const in time

$$\longrightarrow \frac{d}{dt} H(x,p) = \frac{dp}{dt}^T \nabla_p H + \frac{dx}{dt}^T \nabla_x H = 0$$

$$\begin{cases} \frac{dx}{dt} = \nabla_p H = \Sigma^{-1} p \\ \frac{dp}{dt} = -\nabla_x H = -\nabla U(x) \end{cases} \qquad \text{simulate from } 0 \longrightarrow T$$

⊛ langevin dynamics بر اساس HMC ~