https://r2.whiteboardfox.com/284744-0024-5649

نظریه اطلاعات: جلسه سوم 99/7/7

مرور احتمالات:
- نامساوی‌های احتمالاتی (مارکف و چبی‌شف)
- قانون اعداد بزرگ
- زنجیره مارکف

- معرفی ۳ آنتروپی (شانون) ← و خواص آن
- آنتروپی مشترک ← قاعده زنجیره‌ای
- مثال

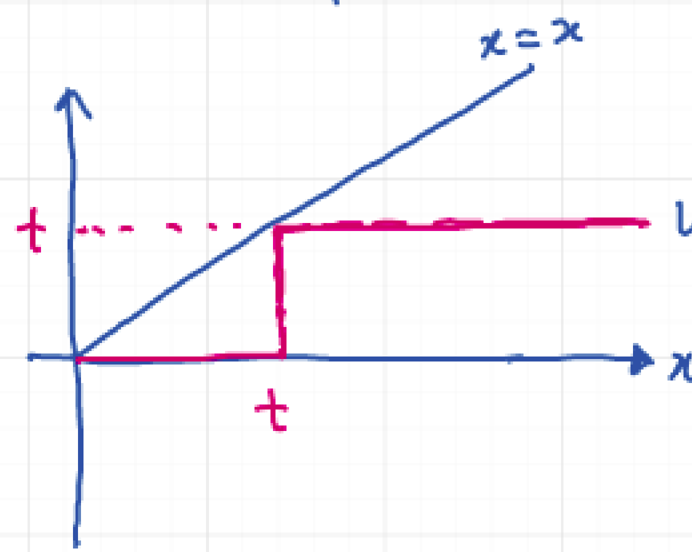
کران مارکف: برای یک متغیر تصادفی نامنفی X کران زیر را داریم:

$$\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$$

یک تابع شاخص

$$x \geq t \cdot u(x)$$

$$\mathbb{E}[X] \geq t \cdot \underbrace{\mathbb{E}[u(x)]}_{\mathbb{P}[X \geq t]}$$



$$u(x) = \begin{cases} 1 & : x \geq t \\ 0 & : x < t \end{cases}$$

کران چبی‌شف : Chebyshev :

متغیر تصادفی دلخواه Z را با امید ریاضی و واریانس متناهی μ_Z و σ_Z^2

یک متغیر نامنفی : $X \triangleq (Z-\mu_Z)^2$

مارکف برای X $\longrightarrow$

$$\mathbb{P}[X \geq \alpha] \leq \frac{\mathbb{E}[X]}{\alpha} = \frac{\sigma_Z^2}{\alpha}$$

$$\mathbb{P}[(Z-\mu_Z)^2 \geq \underbrace{\alpha}_{=\varepsilon^2}] \leq \frac{\sigma_Z^2}{\underbrace{\alpha}_{=\varepsilon^2}}$$

$$\boxed{\mathbb{P}[\,|Z-\mu_Z| \geq \varepsilon] \leq \frac{\sigma_Z^2}{\varepsilon^2}}$$

قانون (ضعیف) اعداد بزرگ : Weak Law of Large Number (WLLN)

n متغیر تصادفی $X_1 \cdots X_n$ داریم . فرض : iid هستند $\leftarrow$ μ_X

$S_n = X_1 + \cdots + X_n$ $\xrightarrow{\frac{S_n}{n}}$ Sample average / Arithmetic "

$$\mathbb{E}\left[\frac{S_n}{n}\right] = \mu_X$$

$$\mathbb{E}\left[\frac{S_n}{n}\right] = \mu_X$$

$$\mathrm{Var}\left[\frac{S_n}{n}\right] = \frac{1}{n^2}\left[\mathrm{Var}[X_1]+\dots+\mathrm{Var}[X_n]\right] = \frac{\sigma_X^2}{n} \xrightarrow{n\to\infty} 0$$

$$= \boxed{\mathbb{E}\left[\left(\frac{S_n}{n}-\mu_X\right)^2\right] \xrightarrow{n\to\infty} 0}$$

همگرایی با

معیار Mean Square

داخل پرانتز

چبیشف

$$\mathbb{P}\left[\left|\frac{S_n}{n}-\mu_X\right| > \varepsilon\right] \le \frac{\sigma_X^2}{n\cdot\varepsilon^2} \xrightarrow{n\to\infty} 0$$

قانون ضعیف اعداد بزرگ :

زنجیره مارکوف: فرض کنید که یک دنباله از متغیرهای تصادفی $X_1\ X_2\ X_3 \dots$ داریم. (یک فرایند تصادفی گسسته زمان است).

می‌گوییم یک زنجیره مارکوف داریم اگر که

$$\mathbb{P}\left[X_n=\alpha_n \mid X_{n-1}=\alpha_{n-1}, \dots, X_1=\alpha_1\right] = \mathbb{P}\left[X_n=\alpha_n \mid X_{n-1}=\alpha_{n-1}\right]$$

به صورت خاص: متغیرهای تصادفی X و Y و Z یک زنجیره مارکوف $X-Y-Z$ ساخته‌اند اگر که

$$P(z|y,x) = P(z|y) \iff P(x,z|y) = P(x|y)\cdot P(z|y)$$

* آنتروپی Entropy

متغیر تصادفی گسسته X را از الفبای $\mathcal{X}$ در نظر بگیرید.

$$H_b(X) = -\sum_{x\in\mathcal{X}} P_X(x)\cdot \log_2 P_X(x)$$

$2 \to$ bit

$e \to$ nat

$$= \mathbb{E}\left[\log_2 \frac{1}{P_X(X)}\right]$$

قرارداد : $0 = 0\log_2 0$

مثال :

$$X = \begin{cases} 0 & \text{w.p. } 0.05 \quad : \text{بارانی} \\ 1 & \text{w.p. } 0.94 \quad : \text{آفتابی} \\ 2 & \text{w.p. } 0.01 \quad : \text{برفی} \end{cases}$$

$$H(X) = -\left(0.05 \log_2 0.05 + 0.94 \log_2 0.94 + 0.01 \log_2 0.01\right)$$

$$\simeq \left(0.05 \times 4.322 + 0.94 \times \underbrace{0.089}_{\log \frac{1}{P_X(1)}} + 0.01 \times 6.644\right) \simeq 0.37 \text{ bit}$$

اگر

$$Y = \begin{cases} 0 & : 1/3 \\ 1 & : 1/3 \\ 2 & : 1/3 \end{cases} \longrightarrow H(Y) \simeq 1.585 \text{ bit} = \log_2 3$$

این کمیت جواب برخی از سوال های مهم است.

- تعداد (کمترین) سوال های لازم (به طور متوسط) برای پیدا کردن مقدار X

- کمترین طول توصیف متغیر X به طور متوسط

---

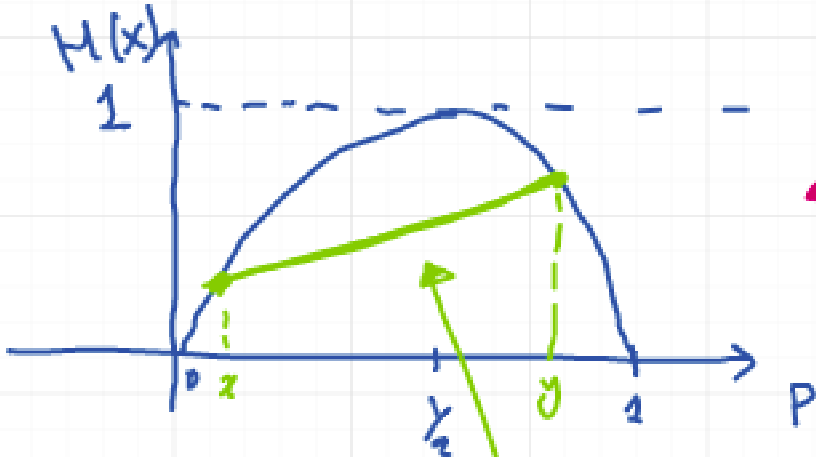
* یک خاصیت آنتروپی:

$$0 \leq P_x \leq 1 \rightarrow \log \frac{1}{P_x} \geq 0 \rightarrow H(X) \geq 0$$

* مثال: متغیر تصادفی برنولی:

$$X = \begin{cases} 1 & : p \\ 0 & : 1-p \end{cases}$$

$$H(X) = -p\log_2 p - (1-p)\log_2(1-p) = \overbrace{h_2(p)}^{\text{binary entropy}}$$



یک تابع مقعر (concave)

* تعریف: تابع $f(x)$ در بازه (a,b) مقعر است $(\cap)$ اگر که به ازای هر $0 \leq \lambda \leq 1$

$$f(\lambda x + (1-\lambda) y) \geq \lambda f(x) + (1-\lambda) f(y)$$

$0 \leq \lambda \leq 1$:

در حالت کلی یه روی الفبای $\mathcal{X}$ دو توزیع P و Q داشته باشیم

$$H(\underbrace{\lambda P + (1-\lambda) Q}) \geq \lambda \cdot H(P) + (1-\lambda) H(Q)$$

خودش یک توزیع احتمال است

---

* آنتروپی مشترک و آنتروپی شرطی :

دو متغیر X و Y با توزیع مشترک $P_{\underset{Z}{XY}}$

$$H(X,Y) \longrightarrow -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} P_{XY}(x,y) \log_2 P_{XY}(x,y)$$

آنتروپی شرطی: $(X,Y) \sim P_{XY}$

$P(x \mid Y=y)$ توزیع مشروط متغیر X به شرط وقوع واقعه $Y=y$

آنتروپی

$$H(X \mid Y=y) = -\sum_{x\in\mathcal{X}} P(x \mid Y=y) \log_2 P(x \mid Y=y)$$

Conditional Ent: $H(X \mid Y) = \mathbb{E}_{P_Y}\left[H(X \mid Y=y)\right]$

$$H(X \mid Y=y) = -\sum_{x \in \mathcal{X}} P(x \mid Y=y) \log_2 P(x \mid Y=y)$$

Conditional Ent: $H(X \mid Y) = \mathbb{E}_{P_Y}\left[H(X \mid Y=y)\right]$

$$= \sum_{y \in \mathcal{Y}} P_Y(y)\, H(X \mid Y=y)$$

$$= -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{XY}(x,y) \log_2 P_{X|Y}(x \mid y)$$

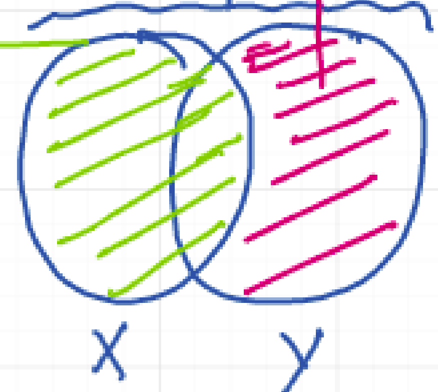
$$= \mathbb{E}_{P_{XY}}\left[-\log_2 P_{X|Y}(X \mid Y)\right]$$

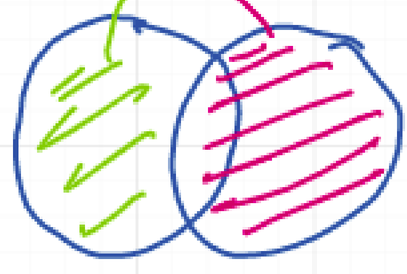
قاعده زنجیره‌ای (chain rule) :

$P(x,y,z) = P(x) \cdot P(y \mid x) \cdot P(z \mid x,y)$: قاعده زنجیره‌ای برای احتمال‌ها

یادآوری

$$H(X,Y) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y)$$



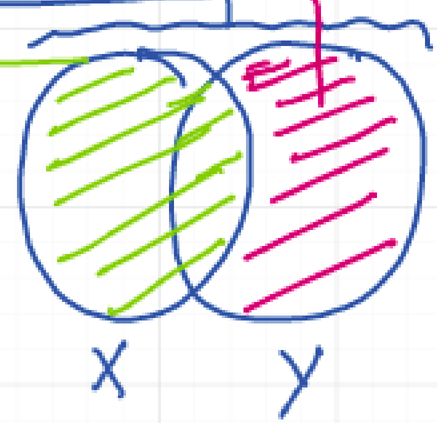


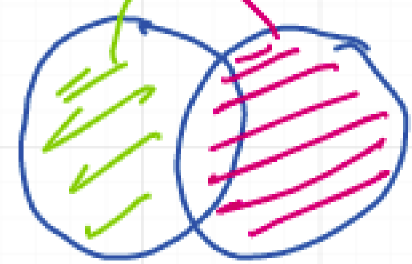
قاعده زنجیره‌ای (chain rule) :

قاعده زنجیره‌ای کلی (احتمال) : $P(x,y,z)=P(x)\cdot P(y|x)\cdot P(z|x,y)$

یادآوری

$$H(X,Y)=H(X)+H(Y|X) = H(Y)+H(X|Y)$$





قاعده زنجیره‌ای برای n متغیر:

$$H(X_1,\dots,X_n)=\sum_{i=1}^{n} H(X_i|X_{i-1},\dots,X_1)$$

$$= H(X_1)+H(X_2|X_1)+\dots+H(X_n|X_{n-1},\dots,X_1)$$

قاعده زنجیره‌ای برای آنتروپی شرطی:

$$H(X,Y|Z)=H(X|Z)+H(Y|X,Z)$$