Recap

KL divergence: $$D(P \| Q) = \sum_x p(x) \log \frac{p(x)}{q(x)}$$

$$= \sum p(x) \log \frac{1}{q(x)} - \sum p(x) \log \frac{1}{p(x)}$$

Properties:

- $D(P\|Q) \geq 0$
- Joint convexity in (P, Q)
- $D(P(X,Y) \| Q(X,Y)) = D(P(X) \| Q(X)) + \underbrace{D(P(Y|X) \| Q(Y|X))}_{\mathbb{E}_{x \sim P}\left[D(P(Y|X=x) \| Q(Y|X=x))\right]}$

Poll on Canvas

Discussion Times

Other distance measures

$$\delta_{TV}(P, Q) = \frac{1}{2}\|P - Q\|_1 = \frac{1}{2}\sum_x |p(x) - q(x)|$$

- Let $f : \mathcal{X} \to \{0, 1\}$ be any labeling function distinguishing P and Q. Then,

$$\left| \mathbb{E}_{x \sim P} f(x) - \mathbb{E}_{x \sim Q} f(x) \right| \le \frac{1}{2}\|P - Q\|_1 = \delta_{TV}(P, Q)$$

<u>Proof</u>:

$$\left| \sum_x (p(x) - q(x)) \cdot \left(f(x) - \frac{1}{2}\right) + \cancelto{0}{\sum_x (p(x) - q(x)) \cdot \frac{1}{2}} \right|$$

$$\le \sum_x |p(x) - q(x)| \cdot \underbrace{\left|f(x) - \frac{1}{2}\right|}_{= 1/2}$$

<u>Ex</u>: Prove that the above bound is tight i.e. $\exists f : \mathcal{X} \to \{0, 1\}$ s.t. $\left| \mathbb{E}_{x \sim P} f(x) - \mathbb{E}_{x \sim Q} f(x) \right| = \frac{1}{2}\|P - Q\|_1$

Pinsker's inequality

$$D(P \| Q) \geq \frac{1}{2\ln 2} \cdot \| P - Q \|_1^2$$

Step 1: For any r.v. X, P, Q $\exists$ binary Z s.t.

$$D(P(X) \| Q(X)) \geq D(P(Z) \| Q(Z)) \qquad (1)$$

$$\| P(X) - Q(X) \|_1 = \| P(Z) - Q(Z) \|_1 \qquad (2)$$

$$Z = f(x) = \begin{cases} 1 & p(x) > q(x) \\ 0 & p(x) \leq q(x) \end{cases}$$

$$D(P(x, f(x)) \| Q(x, f(x))) \geq D(P(\underbrace{f(x)}_{Z}) \| Q(\underbrace{f(x)}_{Z}))$$

$$\|$$

$$D(P(x) \| Q(x))$$

$$+ \underbrace{D(P(x \mid f(x)) \| Q(x \mid f(x)))}_{\geq 0}$$

Step 2: Pinsker for binary variables

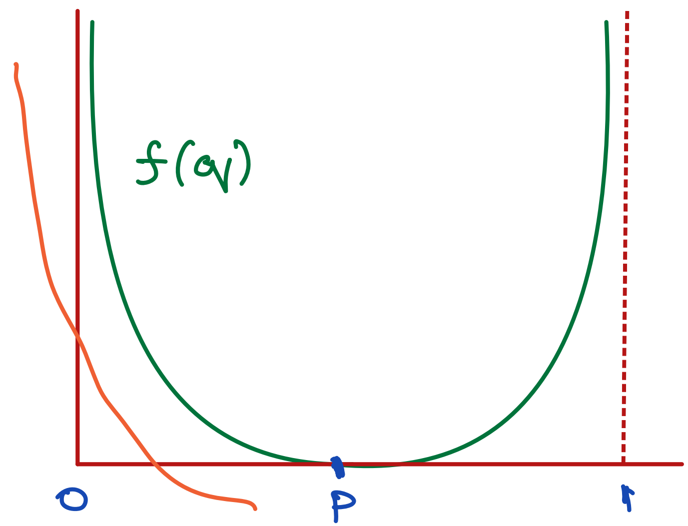
$$P = \begin{cases} 1 & \text{wp } p \\ 0 & \text{wp } 1-p \end{cases} \qquad Q = \begin{cases} 1 & \text{wp } q \\ 0 & \text{wp } 1-q \end{cases}$$

To prove: $$p \ln \frac{p}{q} + (1-p)\cdot \ln \frac{1-p}{1-q} \geq 2(p-q)^2$$

<u>Proof</u>: $$f(q) = p \ln \frac{p}{q} + (1-p) \ln \frac{1-p}{1-q} - 2(p-q)^2$$

$$\frac{\partial f}{\partial q} = -\frac{p}{q} + \frac{(1-p)}{1-q} + 4(p-q)$$

$$= \underbrace{(p-q)}_{\geq 0}\Big(\underbrace{\frac{-1}{q(1-q)}}_{\leq -4} + 4\Big)$$



Distinguishing coins

	H	T
P:	$\frac{1}{2}$	$\frac{1}{2}$
Q:	$\frac{1}{2}+\epsilon$	$\frac{1}{2}-\epsilon$

$$x_1, \ldots, x_n \longrightarrow \boxed{T} \longrightarrow \begin{cases} 0 & (P) \\ 1 & (Q) \end{cases}$$

Smallest n so that T is 90% accurate?

$$\underset{x_1 \ldots x_n \sim P^n}{\mathbb{P}} \left[T(x_1 \ldots x_n) = 0 \right] \geq \frac{9}{10}$$

$$\underset{x_1 \ldots x_n \sim Q^n}{\mathbb{P}} \left[T(x_1 \ldots x_n) = 1 \right] \geq \frac{9}{10}$$

Relating to distance

$$\left| \underbrace{\mathbb{E}_{x_1 \dots x_n \sim P^n}\left[T(x_1,\dots,x_n)\right]}_{\le 1/10} - \underbrace{\mathbb{E}_{x_1\dots x_n \sim Q^n}\left[T(x_1\dots x_n)\right]}_{\geq \frac{9}{10}} \right| \geq \frac{4}{5}$$

$$\frac{1}{2} \cdot \| P^n - Q^n \|_1 \geq \frac{4}{5}$$

$$D(P^n \| Q^n) \geq \frac{1}{2\ln 2} \| P^n - Q^n \|_1^2$$

$$\geq \frac{1}{2\ln 2}\left(\frac{8}{5}\right)^2 = \underbrace{\frac{32}{25 \cdot (\ln 2)^2}}_{C_0}$$

Lower Bound on n

$$D(P^n \| Q^n) = D(P(x_1 \ldots x_n) \| Q(x_1 \ldots x_n))$$

$$= \sum_i D(P(x_i \mid x_1 \ldots x_{i-1}) \| Q(x_i \mid x_1 \ldots x_{i-1}))$$

$$= n \cdot D(P \| Q) \geq c_0 \, ?$$

$$D(P \| Q) = \frac{1}{2} \log\left(\frac{1/2}{1/2+\epsilon}\right) + \frac{1}{2} \log\left(\frac{1/2}{1/2-\epsilon}\right)$$

$$= \frac{1}{2} \log\left(\frac{1}{1-4\epsilon^2}\right) \leq c_1 \cdot \epsilon^2$$

Need $n \geq \frac{c_0}{c_1} \cdot \frac{1}{\epsilon^2}$

Ex: Prove using Chernoff bounds that $n = O(1/\epsilon^2)$ suffices.

Infinite universes : the countable case

Say $\mathcal{X} = \mathbb{N}$. Summations now limits of series

Check convergence

$$1 = \sum_{x \in \mathbb{N}} p(x)$$

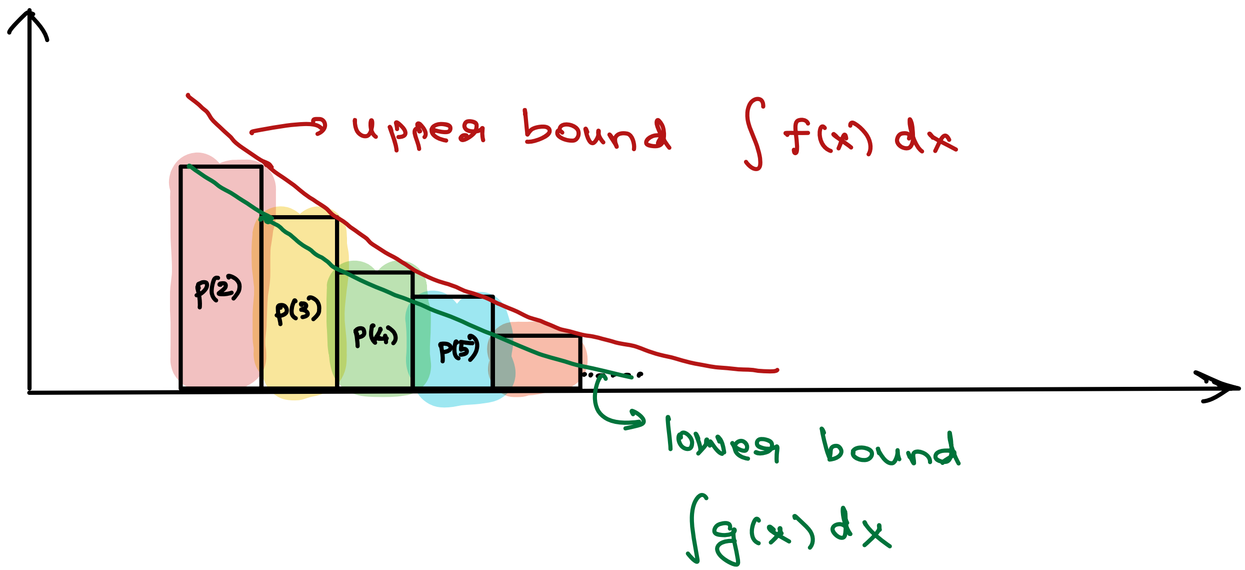
$$H(X) = \sum_{x \in \mathbb{N}} p(x) \log \frac{1}{p(x)}$$

$$I(X;Y) = \sum_{x,y \in \mathbb{N}} p(x,y) \log \frac{p(x|y)}{p(x)} = H(X) - H(X|Y)$$

$$D(P||Q) = \sum_{x \in \mathbb{N}} p(x) \log \frac{p(x)}{q(x)}$$

Infinite entropies

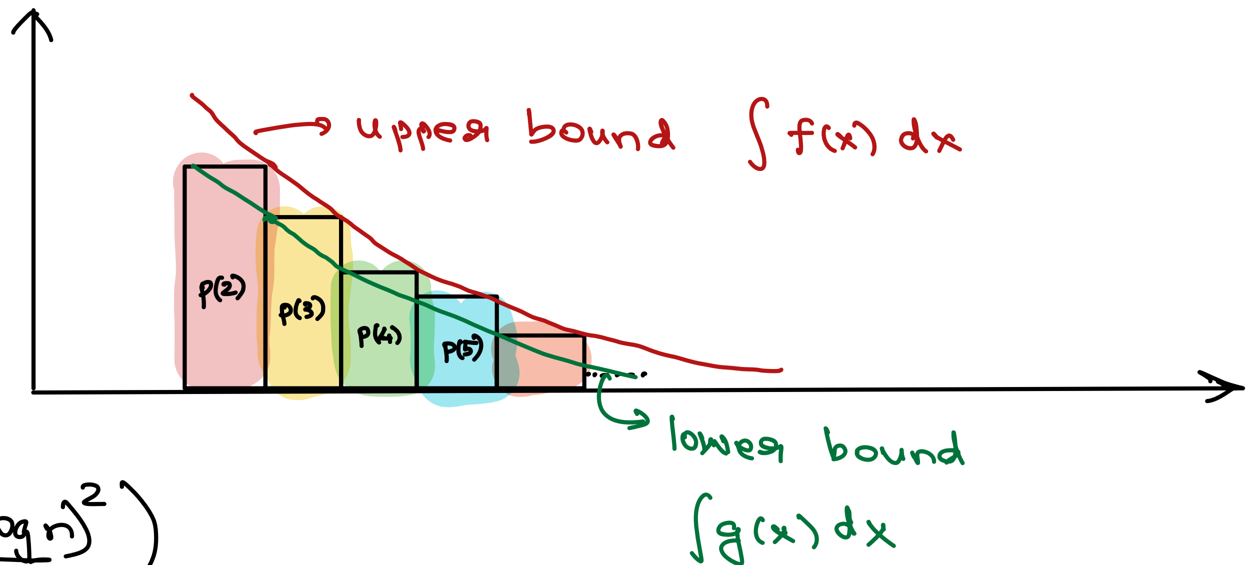
$$P[X = n] = p(n) = \frac{C}{n \cdot (\log n)^2} \quad \forall n \geq 2$$



$$\underbrace{\sum_{n \geq 2} \frac{1}{n(\log n)^2}}_{= C} \leq \frac{1}{2(\log 2)^2} + \int_3^\infty \frac{1}{(x-1)\log(x-1)^2}\,dx = \frac{1}{2} + \ln 2$$

Infinite entropies

$$P[X = n] = p(n) = \frac{c}{n\cdot(\log n)^2} \quad \forall n \geq 2$$

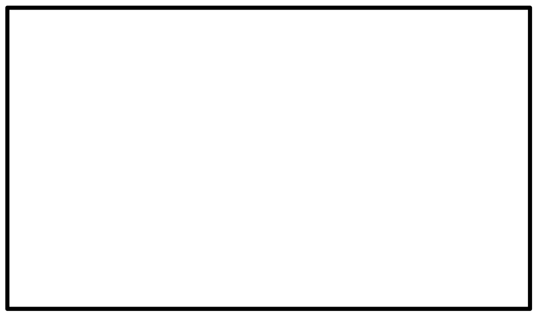


$$H(X) = \sum_{n\geq 2} \frac{c}{n(\log n)^2} \log\left(\frac{n\cdot(\log n)^2}{c}\right)$$

$$= \sum_{n\geq 2} \frac{c}{n\cdot(\log n)^2} \left(\log n + 2\log\log n - \log c\right)$$

$$\sum_{n\geq 2} \frac{1}{n\log n} \geq \int_2^\infty \frac{1}{x\log x}\,dx \longrightarrow \infty$$

Uncountably infinite spaces



$\Omega, \underbrace{\mathcal{F}}_{\sigma\text{-algebra}}, \mu$

Random variables

$X : \Omega \longrightarrow \mathbb{R}^n$ (say) } with Borel σ-algebra

Continuous random variables: There exists $\underbrace{p : \mathbb{R}^n \longrightarrow \mathbb{R}_{\geq 0}}_{\text{density function}}$

s.t. for any box $B \subseteq \mathbb{R}^n$

$$\mathbb{P}[X \in B] = \int_B p(x)\, dx$$

Gaussian Random Variables

1-dimensional: $N(\mu, \sigma^2)$

$\mathbb{E}[X] = \mu$

$\mathbb{E}[(x-\mu)^2] = \sigma^2$

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

n-dimensional: $N(\mu, \Sigma)$

$\mathbb{E}[x] = \mu \in \mathbb{R}^n$

$\mathbb{E}[(x-\mu)(x-\mu)^T] = \Sigma \in \mathbb{R}^{n \times n}$

$$p(x) = \frac{\exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)}{(2\pi)^{n/2}\cdot(\det(\Sigma))^{1/2}}$$

Differential entropy

$$h(X) = \int_{x} p(x) \cdot \log \frac{1}{p(x)} \, dx$$

e.g.

X uniform on $[0,1]$: $h(X) = \int_0^1 1 \cdot \log 1 = 0$

$p(x) = 1$

$Y = \frac{X}{2}$: $h(Y) = \int_0^{1/2} 2 \cdot \log \frac{1}{2} = -1$

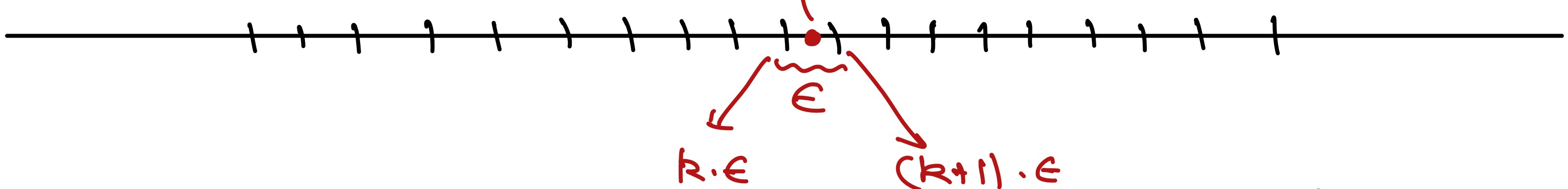
$p(y) = 2$

$Z = X^2$: $h(Z) =$

The weirdness of differential entropy ($\mathcal{X} = \mathbb{R}$)

density $p(x) : \mathbb{R} \longrightarrow \mathbb{R}_{\geq 0}$

$\exists\, x_k$ s.t. $p(x_k) = \frac{1}{\epsilon} \int_{k\cdot\epsilon}^{(k+1)\cdot\epsilon} p(x)\, dx$



Consider discrete Y : $\mathbb{P}[Y = x_k] = \epsilon \cdot p(x_k) = \int_{k\cdot\epsilon}^{(k+1)\cdot\epsilon} p(x)\, dx$

$$\sum_k \mathbb{P}[Y = x_k] = 1$$

$$H(Y) = \sum_k \epsilon \cdot p(x_k) \cdot \log \frac{1}{\epsilon \cdot p(x_k)}$$

$$= \underbrace{\sum_k \epsilon \cdot p(x_k) \cdot \log \frac{1}{p(x_k)}}_{\to\; h(x)} + \underbrace{\log \frac{1}{\epsilon} \sum_k \epsilon \cdot p(x_k)}_{\to\; \infty \text{ as } \epsilon \to 0}$$