

# Probability


## Review

definition  $\left( \begin{array}{l} \Omega : \text{sample space} \\ E : \text{event} \\ \mathcal{F} : \text{set of events} \\ P : \mathcal{F} \rightarrow [0, 1] = \{x \in \mathbb{R} : 0 \leq x \leq 1\} \end{array} \right.$

Ex. Fair coin flip  $\rightarrow$  probability of head, tail

$\Omega : \{H, T\}$   $\left| \begin{array}{l} \mathcal{F} : \{\emptyset, \{H\}, \{T\}, \{H, T\}\} \\ \downarrow \\ Pr(\emptyset) = 0 \quad Pr(\{H\}) = \frac{1}{2} \quad Pr(\{T\}) = \frac{1}{2} \quad Pr(\{H, T\}) = 1 \end{array} \right.$

Ex 5x5 binary

  $= \begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}$   $\Omega = \{0, 1\}^{5 \times 5}$   
 $|\Omega| = 2^{25}$   
 $|\mathcal{F}| = 2^{25}$

## Random variable

$X : \Omega \rightarrow \mathbb{R}$

example: toss coin 5 times

$\Omega : \{H, T\}^5 = \{HHHHH, HHHHT, \dots\}$

$X = \# \text{ of heads}$

$X(HHTHT) = 3$

## [2] Information Theory ( $\approx$ Information Measures)

### • Entropy

$X$ : r.v

$\mathcal{X} = \{0, 1, \dots, M-1\}$

$P_X(x)$ : pmf

### • def Surprisal of $x$

$$S(x) = \log_2 \frac{1}{P_X(x)} = -\log_2 P_X(x)$$

ex:  $x=1$  if earthquake  
0 else

# • def Entropy of random variable $X$

$$H(X) = \mathbb{E}[S(X)]$$

$$= \mathbb{E}\left[\log_2 \frac{1}{P_X(X)}\right] = \sum_{x \in \mathcal{X}} \left(\log_2 \frac{1}{P_X(x)}\right) \cdot P_X(x)$$

ex )  $X$ : binary random variable

Bern( $p$ )

$$\Pr(X=1) = p$$

$$P_X(1) = p, P_X(0) = 1-p$$

$$H(X) = \mathbb{E}\left[\log_2 \frac{1}{P_X(X)}\right] = \sum_{x \in \mathcal{X}} P_X(x) \log_2 \frac{1}{P_X(x)}$$

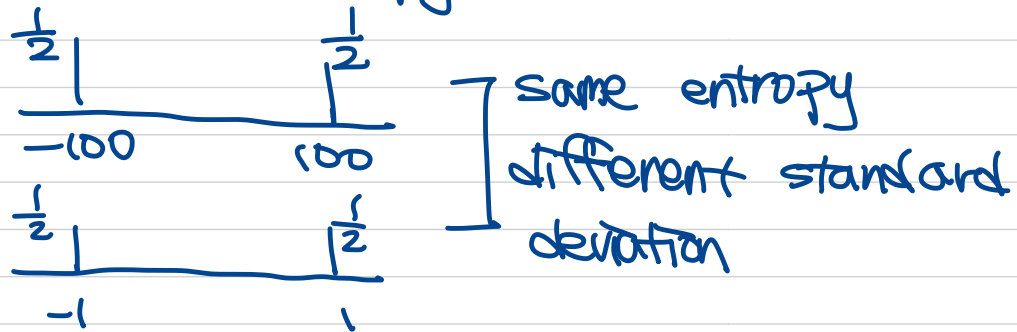
$$= p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p}$$

Entropy = measure of uncertainty  
(randomness)

= amount of information

⇒ Units of Entropy: bits

std vs entropy



\* Entropy can also be used as image hardness

ex>  $\begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix}$  A: p of dog  
B: p of cat  
C: p of fish  
D: p of table

image 1:  $\begin{bmatrix} 0.8 \\ 0.05 \\ 0.15 \\ 0.05 \end{bmatrix}$

image 2:  $\begin{bmatrix} 0.6 \\ 0.2 \\ 0.1 \\ 0.1 \end{bmatrix}$

①  $H(x)$  property: non-negativity

$$H(x) \geq 0$$

&  $H(x) = 0 \Leftrightarrow x$ : deterministic  
( $\Pr(x=x) = 1$ )

Ex>  $p_X(x) = \left(\frac{1}{2}\right)^x$  ( $x \geq 1$ , integer)

Q.  $H(X) = ?$

A.  $H(X) = \mathbb{E} \left[ \log_2 \frac{1}{p_X(x)} \right] = \mathbb{E} \left[ \log_2 2^x \right] = \mathbb{E} [x] = \sum_{x=1}^{\infty} x \cdot \left(\frac{1}{2}\right)^x$

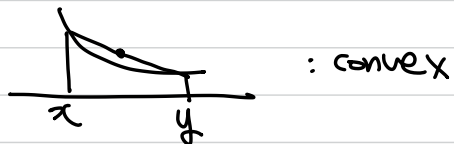
$\sum_{x=1}^{\infty} p^x = \frac{p}{1-p}$  when  $|p| < 1$

$\frac{ds}{dp} = \sum_{x=1}^{\infty} x \cdot p^{x-1} = \frac{d}{dp} \left( -1 + \frac{1}{1-p} \right) = \frac{1}{(1-p)^2} = 4$

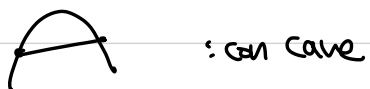
$\therefore \sum_{x=1}^{\infty} x \cdot p^x = 4 \cdot p = 2$

## • Convexity (Concavity)

$f: \mathcal{X} \rightarrow \mathbb{R}$  is convex if  $\eta f(x) + (1-\eta)f(y) \geq f(\eta x + (1-\eta)y)$   $\forall 0 \leq \eta \leq 1, \forall x, y \in \mathcal{X}$



$f: \mathcal{X} \rightarrow \mathbb{R}$  is concave if  $\eta f(x) + (1-\eta)f(y) \leq f(\eta x + (1-\eta)y)$



ex)  $X \sim \text{Bern}(p)$

$$H(X) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} = H_2(p) \quad : \text{concave.}$$

## • Jensen's Inequality

$f: \mathcal{X} \rightarrow \mathbb{R}$  concave

$$\mathbb{E}[f(X)] \leq f(\mathbb{E}[X])$$

$\Rightarrow$  Max Entropy when uniform

## • Mismatch

$$H(X) = \mathbb{E} \left[ \log \frac{1}{P_X(x)} \right] = \sum_x P_X(x) \frac{1}{\log \frac{1}{P_X(x)}}$$

Suppose!  $X \sim Q$

$$H(X) \leq \mathbb{E}_{P_X} \left[ \log \frac{1}{P_X(x)} \right]$$

$\Rightarrow$  KL-Divergence

- **KL-Divergence** (= relative entropy)  $\rightarrow$  non negative by mismatch

def

$$D(P||Q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}$$

\* Not distance or metric but divergence

i) unstable: certain event  $\Rightarrow$  divergence  $\rightarrow \infty$

ii) symmetry  $\times$

\* Earth Mover dist (= Wasserstein dist): symmetric ~~etc~~

— generate realistic image

GAN: true distribution  $\text{img} = P_X(\text{r.v.})$

$\rightarrow$  image  $Q$

minimize dist between  $P_X$  and  $Q$

## • **Cross Entropy**

def  $\sum_{x \in X} \underbrace{P_X(x)}_{\text{true}} \log \underbrace{\frac{1}{Q(x)}}_{\text{ass'n}}$

indep def)  $X$  &  $Y$  are indep

$$\text{iff } P_{X,Y}(x,y) = P_X(x) P_Y(y) \quad \forall x,y$$

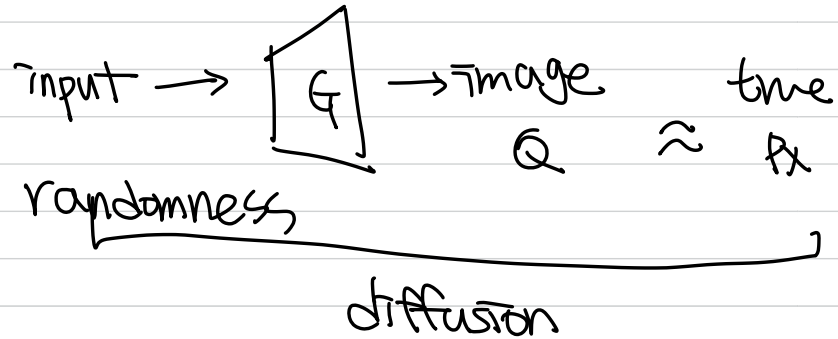
def **Random Vector**

$$X^n = (X_1, X_2, \dots, X_n) \quad x^n = (x_1, \dots, x_n)$$

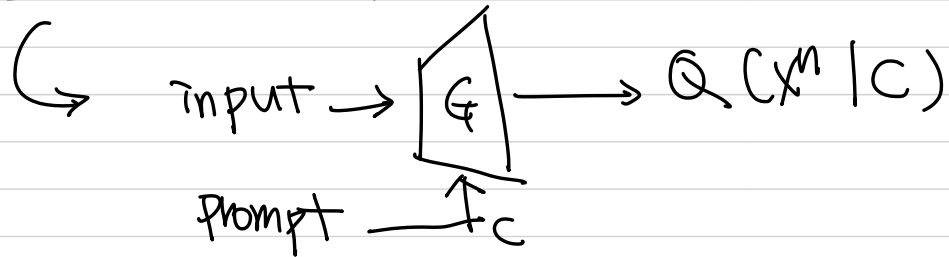
joint pmf

$$P_{x^n}(x^n) = \Pr(X^n=x^n) = \Pr(x_1=x_1, \dots, x_n=x_n)$$

## def Diffusion Overview



train process



MNIST  
 $Q(x^n)$

MNIST  $\eta$   
 $Q(x^n | \eta)$



## • Joint Entropy

$$H(X_1, X_2) = \sum_{(x_1, x_2)} P_{X_1, X_2}(x_1, x_2) \log \frac{1}{P_{X_1, X_2}(x_1, x_2)}$$

$$X_1 \in \mathcal{X}_1 = \{1, 2, \dots, M_1\}$$

$$X_2 \in \mathcal{X}_2 = \{1, 2, \dots, M_2\}$$

$\Rightarrow$  total uncertainty of pair  $(X_1, X_2)$

: total amount of information

Special case:  $X_1 \perp\!\!\!\perp X_2 \Leftrightarrow P_{X_1, X_2}(x_1, x_2) = P_{X_1}(x_1) P_{X_2}(x_2)$

①  $H(X_1, X_2) = H(X_1) + H(X_2)$  (when indep)

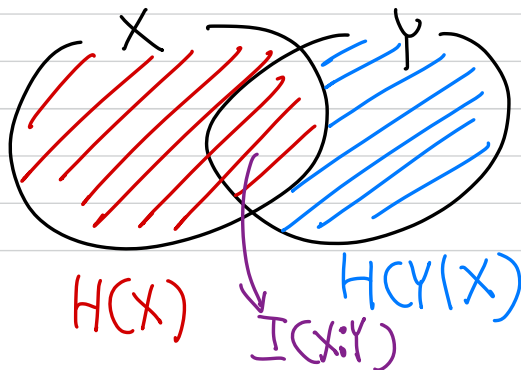
②  $H(X_1, X_2) = H(X_1) + H(X_2|X_1)$

**= Conditional Entropy**

## • Conditional Entropy

def  $H(X|Y)$  = amount of info in  $X$  when  $Y$  is given?

$$= \mathbb{E} \left[ \log \frac{1}{P_{X|Y}(X|Y)} \right]$$



$I(X:Y)$  : Mutual Information

$\hookrightarrow$  nonnegative

$\hookrightarrow$  symmetric

## • Data processing inequality

①  $f: X \rightarrow \mathbb{R}$   
 $H(X) \geq H(f(X))$

②  $f: X \rightarrow \mathbb{R}$   
 $I(X; Y) \geq I(f(X); Y)$

③ Markovity: seq of r.v

Markov triplet  $X-Y-Z$  : iff  $X$  &  $Z$  are indep given  $Y$

$$I(Z; X) \geq I(Z; Y)$$

## • Random Process

① iid

② Markov Process (1st order)

def 
$$P_{X_i | X_{i-1}, \dots, X_1} = P_{X_i | X_{i-1}}$$

(namely, Probability in  $i$ th stage only depend on previous stage ( $i-1$ )th process)

③ kth order Markov Process

def  $P_{X_i | X_{i-1} \dots X_i} = P_{X_i | X_{i-1} \dots X_{i-k}}$

④ **Stationary**

$$X_i^n = X_{i+1}^{i+n} \quad \forall i, n$$

## • Continuous random variable

$F_X(x) = \Pr(X \leq x)$  : cumulative distribution function

$f_X(x) = \frac{dF_X(x)}{dx}$  : probability density function

Jacobian?

If  $Y = h(X) \Rightarrow |dy f_Y(y)| = |dx f_X(x)| = \text{probability}$

$$\rightarrow f_Y(y) = \left| \frac{dx}{dy} \right| f_X(x) = \left| \frac{\partial h^{-1}(y)}{\partial y} \right| f_X(h^{-1}(y))$$

$$(Y_1, Y_2) = H(X_1, X_2)$$

$$f_{Y_1, Y_2}(y_1, y_2) = \underbrace{\left| \frac{dx_1 dx_2}{dy_1 dy_2} \right|}_{\text{Jacobian}} f_{X_1, X_2}(x_1, x_2)$$

$$= \det \left( \frac{\partial H^{-1}}{\partial y} \right) : \text{Jacobian}$$

- Gaussian Distribution

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Jointly Gaussian  $\sim N(\mu, \Sigma)$

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

$$f_{X_1, X_2} = \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} \exp\left(-\frac{1}{2}(X-\mu)^T \Sigma^{-1} (X-\mu)\right)$$

$$f_{X^n}(x^n) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(x^n - \mu)^T \Sigma^{-1} (x^n - \mu)\right)$$

# • Differential Entropy

$$\begin{aligned} \left( \begin{array}{l} \text{when discrete: } D(P \parallel Q) = \mathbb{E}_P \left[ \log \frac{P(X)}{Q(X)} \right] = \int f(x) \log \frac{f(x)}{g(x)} dx \quad : \text{ratio} \\ \text{when continuous: } D(P \parallel Q) = \mathbb{E}_P \left[ \log \frac{P(X)}{Q(X)} \right] = \sum_x p(x) \log \frac{p(x)}{q(x)} \quad : \text{quantity} \end{array} \right. \end{aligned}$$

→ In continuous case

$$\text{differential entropy } h(X) := \mathbb{E}_{f_X} \left[ \log \frac{1}{f_X(X)} \right]$$

Properties

$$\textcircled{1} h(X) \neq h(aX) = h(X) + \log |a|$$

$$\textcircled{2} h(X_1, X_2) = h(X_1) + h(X_2 | X_1)$$

$$\textcircled{3} I(X; Y) = h(X) + h(Y) - h(X, Y)$$

$$h(X, Y) = \mathbb{E}_{f_{X,Y}} \left[ \log \frac{1}{f_{X,Y}(X, Y)} \right]$$