× Lecture 5, oct. 6, 04

Mohsen had covered in lecture 4, Sections 4.1 – 4.5. I do a brief review and also talk briefly about entropy.

Information we get about an event is inversely related to its probability. The more improbable an event is the more uncertain we are about its occurence and the more information is that we obtain by knowing the outcome.

Need for logarithmic function: Additivity of information.

$$I(X=k) = \log \frac{1}{P[X=k]} = -\log P[X=k]$$

example: Information we get about the outcome of a coin toss:

$$I(X=H) = -\log_2 1/2 = 1$$ (1 bits of informat

example

$$I(X_1=H, X_2=T) = -\log_2 1/4 = 2 \text{ bits}$$

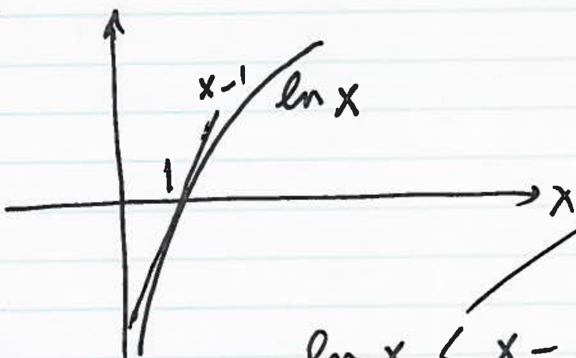
When the base of log is 2, we call the unit bit if it is e, i.e., $-\ln P(X=k)$ we call it <u>nat</u> :

<u>Entropy</u> : is the expected value of the information, i.e.,

$$H(X) = E\left[\log \frac{1}{P(X=k)}\right] = -\sum_k P(X=k) \log P(X=k)$$

or $$H(X) = -\sum_k p_k \ln p_k \quad \text{nats}$$

or $$H(X) = -\sum_k p_k \log_2 p_k \quad \text{bits}$$



$$\ln \frac{1}{x} \le \frac{1}{x} - 1$$

$$\ln x \ge 1 - \frac{1}{x}$$

$$\ln x \le x - 1 \qquad \text{with equality at } x=$$

$$\cancel{\ln K \;\; H(X) = \ln K \sum_k p_k \ln p_k = \sum_k p_k \ln \frac{Kp_k}{} \le \sum_k p_k \left(\frac{1}{p_k} - 1\right)}$$

$$\ln K - H(X) = \sum_k p_k \ln \frac{p_k}{1/K} \ge \sum_k p_k \left(1 - \frac{1/K}{p_k}\right)$$

$$= \sum_k p_k - \sum_K 1/K = 0$$

$$\ln K - H(X) \ge 0 \implies H(X) \le \ln K$$

in general $H(X) \le \log K$

Example:

Assume that an experiment has four outcomes $\{A, B, C, D\}$ or $\{1, 2, 3, 4\}$ with probabilities $\{1/2, 1/4, \frac{1}{8}, \frac{1}{8}\}$

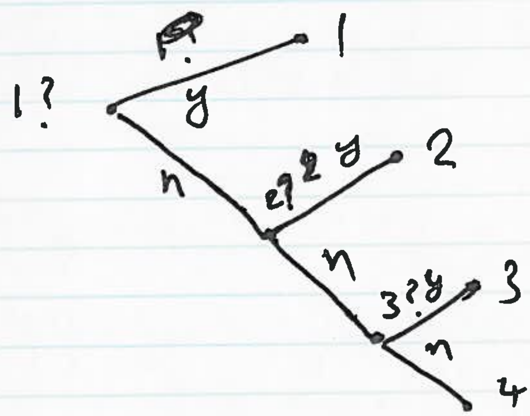
The entropy is

$$H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log 1/8 - \frac{1}{8}\log 1/8$$

or

$$H(X) = \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{3}{8} = 1\,3/4 = \frac{7}{4}$$

we can know the outcome using the following questioning strategy:



$$\text{average \# of questions} = \frac{1}{2} + \frac{1}{4}\times 2 + \frac{1}{8}\times 3 + \frac{1}{8}\times 3$$
$$= 1\,3/4$$

<u>Source Coding Theorem</u>: The average number of bits required to encode a source X is $H(X)$.

Differential Entropy

For a continuous random variable, we have

$$H(X) = -\int_{-\infty}^{\infty} f_X(x) \log f_X(x)\, dx$$

example: For the Gaussian random ~~var~~ variable

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

$$H(X) = -\int_{-\infty}^{\infty} f_X(x) \log[f_X(x)]\, dx$$

$$= -\int_{-\infty}^{\infty} \left(\log \frac{1}{\sqrt{2\pi}\,\sigma}\right) f_X(x)\, dx + \int_{-\infty}^{\infty} \frac{(x-m)^2}{2\sigma^2} f_X(x)\, dx$$

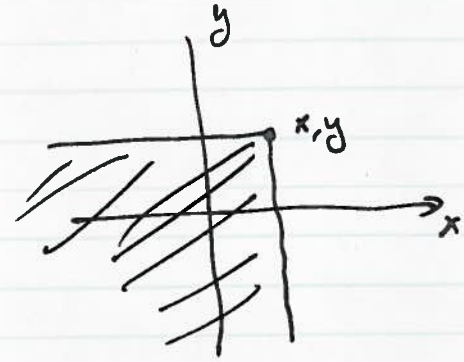
$$= \frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2} = \frac{1}{2}\log(2\pi e \sigma^2)$$

Multiple random variables:

A) Two random variables:

Joint CDF

$$F_{X,Y}(x,y) = P[X \le x, Y \le y]$$



Properties of Joint CDF:

1) $F_{X,Y}(x_1, y_1) \le F_{X,Y}(x_2, y_2)$ if $x_1 \le x_2$ & $y_1 \le y_2$

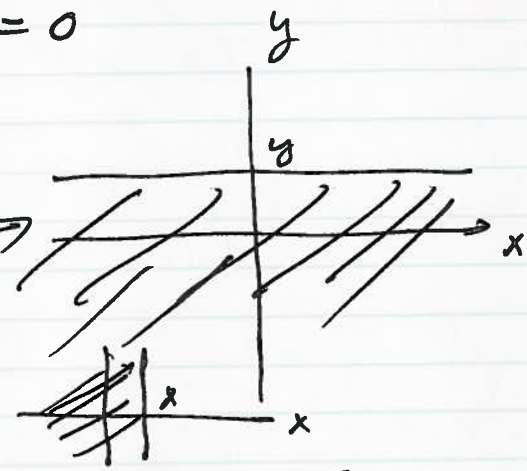
2) $F_{X,Y}(-\infty, y) = F_{X,Y}(x, -\infty) = 0$

3) $F_{X,Y}(\infty, \infty) = 1$

4) $F_{X,Y}(x, \infty) = F_X(x)$

and

$F_Y(y) = F_{X,Y}(\infty, y)$



Proof:

$$F_{X,Y}(x, \infty) = P[X \le x, Y < \infty] = P[X \le x]$$

$F_X(x)$ and $F_Y(y)$ are called the marginal CDF's.

Joint pdf

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

$$P[a \le X \le b,\ c \le Y < d] = \int_a^b \int_c^d f_{X,Y}(x,y)\,dx\,dy$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$$

in general:

$$P[(X,Y) \in A] = \iint_A f_{X,Y}(x,y)\,dx\,dy$$

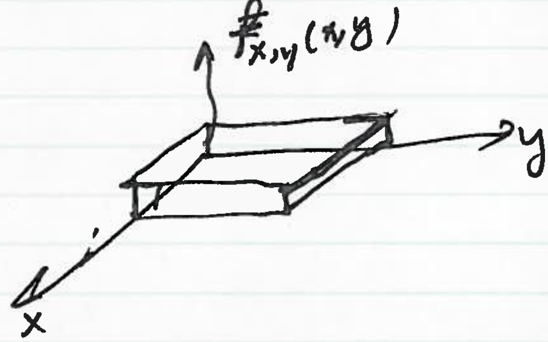
marginal pdf

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$$

Example:

$$f_{X,Y}(x,y) = \begin{cases} 1 & 0 \le x \le 1 \ , \ 0 \le y \le 1 \\ 0 & \text{elsewhere} \end{cases}$$



find $F_{X,Y}(x,y)$

1) $x \le 0 \ , \ y \le 0$

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} 0 \, dx \, dy = 0$$

2) $0 \le x \le 1$ and $0 \le y \le 1$

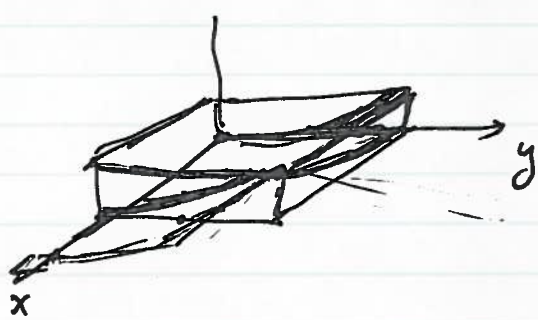
then

$$F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} 1 \, dx \, dy = xy$$

3) $0 \le x \le 1$ and $y > 1$

then

$$F_{X,Y}(x,y) = \int_{0}^{x} \int_{0}^{1} dx \, dy = x$$



4) $x > 1$ and $0 \le y \le 1$

$$F_{X,Y}(x,y) = y$$

5) $x > 1 \ , \ y > 1$

$$F_{X,Y}(x,y) = 1$$

Two random variables are independent

iff:

$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$

or

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

Conditional CDF:

$$F_Y(y|x) = \frac{P[Y \leq y,\ X = x]}{P[X = x]} \quad (\text{if } x \text{ is discrete})$$

$$f_Y(y|x) = \frac{d}{dy} F_Y(y|x)$$

if x is continuous

$$F_Y(y|x) = \frac{P[\ Y \leq y,\ \ x < X \leq x+h]}{P[\cancel{x}\ x \cancel{\leq} X \leq x+h]}$$

$$= \frac{\int_{-\infty}^{y} \int_{x}^{x+h} f_{X,Y}(x',y')\,dx'\,dy'}{\int_{x}^{x+h} f_X(x')\,dx'}$$

$$\simeq \frac{\int_{-\infty}^{y} f_{X,Y}(x,y')\,dy'\,h}{f_X(x)\,h}$$

or

$$F_y(y|x) = \frac{\int_{-\infty}^{y} f_{x,y}(x,y')\,dy'}{f_x(x)}$$

$$f_y(y|x) = \frac{d}{dy} F_y(y|x) = \frac{f_{x,y}(x,y)}{f_x(x)}$$

or

$$\boxed{f_{x,y}(x,y) = f_x(x)\, f_y(y|x)}$$

But, we could similarly have shown that

$$f_{x,y}(x,y) = f_y(y)\, f_x(x|y)$$

So:

$$f_y(y) f_x(x|y) = f_x(x)\, f_y(y|x)$$

i.e.,

$$\boxed{f_x(x|y) = \frac{f_x(x)\, f_y(y|x)}{f_y(y)}}$$

This is called the Bayes Rule.

XX Next time: Talk about Detection (MAP & ML)

Example:

A random variable X is selected uniformly in [0,1] and the Y is selected uniformly in [0,x]. Find $f_Y(y)$.

$$F_Y(y) = P[Y \le y] = \int_0^{\infty} P[Y \le y \mid X = x]\, f_X(x)\, dx$$

when $X = x$ then

$$P[Y \le y \mid X = x] = \begin{cases} \frac{y}{x} & 0 \le y \le x \\ 1 & x < y \end{cases}$$

$$F_Y(y) = \int_0^{y} P[Y \le y \mid X = x]\, f_X(x)\, dx$$

$$+ \int_y^{1} P[Y \le y \mid X = x]\, f_X(x)\, dx$$

$$= \int_0^{y} 1\, dx + \int_y^{1} \frac{y}{x}\, dx = y - y \ln y$$

$$\boxed{f_Y(y) = \frac{d}{dy} F_Y(y) = -\ln y \qquad 0 \le y \le 1}$$