Tutorial 2 – the outline

- **Example-1 from linear algebra**
- **Conditional probability**
- **Example 2: Bernoulli Distribution**
- **Bayes' Rule**
- **Example 3**
- **Example 4 The game of three doors**

Example – 1: Linear Algebra

- A line can be written as $ax+by=1$. You are given a number of example points:

$$P = \begin{bmatrix} x_1 & y_1 \\ \vdots & \vdots \\ x_n & y_n \end{bmatrix}$$

Let $M = \begin{bmatrix} a \\ b \end{bmatrix}$

- **(A)** Write a single matrix equation that expresses the constraint that each of these points lies on a single line
- **(B)** Is it always the case that some M exists?
- **(C)** Write an expression for M assuming it does exist.

Example – 1: Linear Algebra

- **(A)** For all the points to lie on the line, *a* and *b* must satisfy the following set of simultaneous equations:

$$a x_1 + b y_1 = 1$$
$$\vdots$$
$$a x_N + b y_N = 1$$

This can be written much more compactly in the matrix form (linear regression equation):

$\boldsymbol{PM}=\mathbf{1}$ where **1** is an Nx1 column vector of 1's.

- **(B)** An $\boldsymbol{M}$ satisfying the matrix equation in part **(A)** will not exist unless al the points are collinear (i.e. fall on the same line). In general, three or more points may not be collinear.
- **(C)** If $\boldsymbol{M}$ exists, then we can find it by finding the left inverse of $\boldsymbol{P}$, but since $\boldsymbol{P}$ is in general not a square matrix $\boldsymbol{P}^{-1}$ may not exist, so we need the pseudo-inverse $(\boldsymbol{P}^T\boldsymbol{P})^{-1}\boldsymbol{P}^T$. Thus $\boldsymbol{M}=(\boldsymbol{P}^T\boldsymbol{P})^{-1}\boldsymbol{P}^T\mathbf{1}$.

Conditional probability

- When 2 variables are statistically dependent, knowing the value of one of them lets us get a better estimate of the value of the other one. This is expressed by the conditional probability of *x* given *y*:

$$\Pr\{x = v_i \mid y = w_j\} = \frac{\Pr\{x = v_i, y = w_j\}}{\Pr\{y = w_j\}}, \text{ or } P(x \mid y) = \frac{P(x, y)}{P_y(y)}$$

- If x and y are statistically independent, then $P(x \mid y) = P_x(x).$

Bayes' Rule

- The law of total probability: If event A can occur in m different ways $A_1, A_2, \ldots, A_m$ and if they are mutually exclusive,then the probability of A occurring is the sum of the probabilities $A_1, A_2, \ldots, A_m$.

$$P(y) = \sum_{x \in \chi} P(x, y).$$

From definition of condition probability

$$P(x,y) = P(x|y)P(y) = P(y|x)P(x)$$

Bayes' Rule

$$P(x \mid y) = \frac{P(y \mid x) P_x(x)}{P(y)}$$

or

$$P(x \mid y) = \frac{P(y \mid x) P(x)}{\sum_{x \in \chi} P(x, y)} = \frac{P(y \mid x) P(x)}{\sum_{x \in \chi} P(y \mid x) P(x)}$$

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

Bayes' rule – continuous case

- For continuous random variable we refer to densities rather than probabilities; in particular,

$$p(x \mid y) = \frac{p(x, y)}{p(y)}$$

- The Bayes' rule for densities becomes:

$$p(x \mid y) = \frac{p(y \mid x)\, p(x)}{\int_{-\infty}^{\infty} p(y \mid x)\, p(x)\, dx}$$

Bayes' formula - importance

- Call *x* a ‘cause’, y an effect. Assuming *x* is present, we know the likelihood of *y* to be observed
- The Bayes’ rule allows to determine the likelihood of a cause *x* given an observation *y*.

  (Note that there may be many causes producing *y*).
- The Bayes’ rule shows how probability for *x* changes from **prior** p(*x*) before we observe anything, to **posterior** p(x| y) once we have observed *y*.

Example – 2: Bernoulli Distribution

- A random variable X has a Bernoulli distribution with parameter θ if it can assume a value of 1 with a probability of θ and the value of 0 with a probability of $(1-\theta)$. The random variable X is also known as a Bernoulli variable with parameter θ and has the following probability mass function:

$$\mathbf{p}(x,\theta) = \begin{cases} \theta & x = 1 \\ 1-\theta & x = 0 \end{cases}$$

- The mean of a random variable X that has a Bernoulli distribution with parameter p is

$$E(X) = 1(\theta) + 0(1-\theta) = \theta$$

The variance of X is

Example – 2: Bernoulli Distribution

$$Var(X) = E(X^2) - [E(X)]^2 = 1^2(\theta) + 0^2(1-\theta) - \theta^2 = \theta - \theta^2 = \theta(1-\theta)$$

A random variable whose value represents the outcome of a coin toss (1 for heads, 0 for tails, or vice-versa) is a Bernoulli variable with parameter θ, where θ is the probability that the outcome corresponding to the value 1 occurs. For an unbiased coin, where heads or tails are equally likely to occur, $\theta = 0.5$.

For Bernoulli rand. variable x_n the probability mass function is:

$$P(x_n \mid \theta) = P_\theta(x_n) = \theta^{x_n}(1-\theta)^{1-x_n}, \qquad x_n = 0,1$$

For N independent Bernoulli trials we have random sample

$$\mathbf{x} = (x_0, x_1, \ldots, x_{N-1})$$

Example – 2: Bernoulli Distribution

The distribution of the random sample is:

$$P_\theta(\mathbf{x}) = \prod_{n=0}^{N-1} \theta^{x_n}(1-\theta)^{1-x_n} = \theta^k(1-\theta)^{N-k}$$

$$k = \sum_{n=0}^{N-1} x_n \qquad \text{number of ones} \in x = (x_0, x_1, ..., x_{N-1}).$$

The distribution of the number of ones in N independent Bernoulli trials is:

$$P_\theta(k) = \binom{N}{k}\theta^k(1-\theta)^{N-k}$$

The joint probability to observe the sample x and the number k

$$P_\theta(\mathbf{x}, k) = \begin{cases} P_\theta(\mathbf{x}), & k = \text{number of ones} \in \mathbf{x} \\ 0, & \text{otherwise} \end{cases}$$

Example – 2: Bernoulli Distribution

The conditional probability of $\boldsymbol{x}$ given the number k of ones:

$$P_\theta(\mathbf{x} \mid k) = \frac{P_\theta(\mathbf{x}, k)}{P_\theta(k)} = \frac{\theta^k (1-\theta)^{N-k}}{\binom{N}{k} \theta^k (1-\theta)^{N-k}} = \frac{1}{\binom{N}{k}}$$

Thus

$$P_\theta(\mathbf{x}) = P_\theta(\mathbf{x} \mid k) P_\theta(k) = \frac{1}{\binom{N}{k}} P_\theta(k)$$

Example - 3

- Assume that *X* is distributed according to the Gaussian density with mean $\mu=0$ and variance $\sigma^2=1$.
  - **(A)** What is the probability that $x=0$?
- Assume that *Y* is distributed according to the Gaussian density with mean $\mu=1$ and variance $\sigma^2=1$.
  - **(B)** What is the probability that $y=0$?
- Given a distribution: $Pr(Z=z)=1/2Pr(X=z)+1/2Pr(Y=z)$

  known as a mixture (i.e. ½ of the time points are generated by the *X* process and ½ of the time points by the *Y* process).
  - **(C)** If $Z=0$, what is the probability that the *X* process generated this data point ?

Example – 3 solutions

- **(A)** Since *p(x)* is a continuous density, the probability that $x=0$ is

$$\Pr(0 < x \le 0) = \int_0^0 p(x)dx = 0.$$

- **(B)** As in part **(A)**, the probability that $y=0$ is

$$\Pr(0 < y \le 0) = \int_0^0 p(y)dy = 0.$$

- **(C)** Let ω_0 (ω_1) be the state where the *X* (*Y*) process generates a data point. We want Pr(ω_0 |$Z=0$). Using Bayes' rule and working with the probability densities to get the total probability:

Example - 3

$$\Pr(\omega_0 \mid Z=0) = \frac{p(Z=0\mid\omega_0)\Pr(\omega_0)}{p(Z=0)} = \frac{p(Z=0\mid\omega_0)\Pr(\omega_0)}{p(Z=0\mid\omega_0)\Pr(\omega_0) + p(Z=0\mid\omega_1)\Pr(\omega_1)}$$

$$= \frac{0.5\,p_X(X=0)}{0.5\,p_X(X=0) + 0.5\,p_Y(Y=0)} = \frac{p_X(X=0)}{p_X(X=0) + p_Y(Y=0)}$$

$$= \frac{0.3989}{0.3989 + 0.2420} = 0.6224$$

- where

$$p_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad p_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-1)^2}{2}}$$

Example 4 The game of three doors

- A game: 3 doors, there is a prize behind one of them. You have to select one door.
- Then one of the other two doors is opened (not revealing the prize).
- At this point you may either stick with your door, or switch to the other (still closed) door.
- *Should one stick with his initial choice, or switch, and does your choice make any difference at all?*

The game of three doors

- Let H_i denote the hypothesis "the prize is behind the door i ".
- Assumption: $\Pr(H_1) = \Pr(H_2) = \Pr(H_3) = \frac{1}{3}$
- Suppose w.l.o.g.: initial choice of door 1,then door 3 is opened. We can stick with 1 or switch to 2.
- Let D denote the door which is opened by the host. We assume:

$$\Pr(D=2 \mid H_1) = \frac{1}{2}, \Pr(D=2 \mid H_2) = 0, \Pr(D=2 \mid H_3) = 1,$$

$$\Pr(D=3 \mid H_1) = \frac{1}{2}, \Pr(D=3 \mid H_2) = 1, \Pr(D=3 \mid H_3) = 0$$

- By Bayes' formula:

$$\Pr(H_i \mid D=3) = \frac{\Pr(D=3 \mid H_i)\Pr(H_i)}{\Pr(D=3)},$$

$$\Pr(H_1 \mid D=3) = \frac{\frac{1}{2}\frac{1}{3}}{\Pr(D=3)}, \Pr(H_2 \mid D=3) = \frac{1 \cdot \frac{1}{3}}{\Pr(D=3)}, \Pr(H_3 \mid D=3) = 0$$

The game of three doors-the solutiion

- The denominator $\Pr(D=3)=\frac{1}{2}$ is a normalizing factor.
- So we get

$$\Pr(H_1 \mid D=3)=\frac{1}{2},$$

$$\Pr(H_2 \mid D=3)=\frac{2}{3},$$

$$\Pr(H_3 \mid D=3)=0,$$

which means that we are more likely to win the prize if we switch to the door 2.

Further complication

- A violent earthquake occurs just before the host has opened one of the doors; door 3 is opened accidentally, and there is no prize behind it. The host says “it is valid door, let’s let it stay open and go on with the game”.
- What should we do now?
- First, any number of doors might have been opened by the earthquake. There are 8 possible outcomes, which we assume to be equiprobable: d=(0,0,0),…,d=(1,1,1).
- A value of *D* now consists of the outcome of the quake, and the visibility of the prize; e.g., <(0,0,1),NO>
- We have to compare $\Pr(H_1|D)$ vs. $\Pr(H_2|D)$.

The earthquake-continued

- $\Pr(D|H_i)$ hard to estimate, but we know that $\Pr(D|H_3)=0$.
- Also from $\Pr(H_3, D) = \Pr(H_3 \mid D)\Pr(D) = \Pr(D|H_3)\Pr(H_3)$ and from $\Pr(D) \neq 0$ we have $\Pr(H_3 \mid D)=0$.

- Further, we have to assume that $\Pr(D|H_1)= \Pr(D|H_2)$ (we don't know the values, but we assume they are equal).
- Now we have

The earthquake-continued

$$\Pr(H_1 \mid D) = \frac{\Pr(D \mid H_1)\Pr(H_1)}{\Pr(D)} = \frac{1}{2},$$

$$\Pr(H_2 \mid D) = \frac{\Pr(D \mid H_2)\Pr(H_2)}{\Pr(D)} = \frac{1}{2}$$

(why they are ½?)

(Because $\Pr(H_1) = \Pr(H_2)$ and $\Pr(D \mid H_1) = \Pr(D \mid H_2)$ we get $\Pr(H_1 \mid D) = \Pr(H_2 \mid D)$ Also $\Pr(H_1 \mid D) + \Pr(H_2 \mid D) + \Pr(H_3 \mid D) = 1$)

- So, we might just as well stick with our choice.
- We have different outcome because the data here is of different nature (although looks the same).