AM216 Stochastic Differential Equations

Lecture 01

List of topics in this lecture

- The framework of repeated experiments for probability
- Outcome, sample space, random variable, event, probability of an event
- Conditional probability, independence of two events, two random variables
- Union, intersection, complement, law of total probability
- Expected value, law of total expectation

Review of probability theory

Question: What is probability?

Example 1: Flip a fair coin

Prob(head) = 50%

What is the exact meaning of this statement?

While most people won't be struggling with this statement, the next example is somewhat different.

Example 2:

I go to see my doctor. Before doing any test she tells me

Prob(I have cancer) = 15%

What is the exact meaning of this statement?

I either have or not have cancer. The answer is deterministic, which is unknown at the moment and which theoretically will be known after a sequence of comprehensive tests. When we restrict the scope of consideration to one person (me), my cancer status is not uncertain. It can be fully determined. It is just unknown for the time being.

So the probability "Prob(I have cancer)" has to be interpreted in a proper framework …

The framework of repeated experiments (with uncertain outcomes):

Probability of an event

= fraction of repeated experiments with the event occurring

$$= \frac{\text{\# of repeats with the event occurring}}{\text{\# of repeats}}$$

Example 1: Flip a fair coin. Repeat this M times.

$$\text{Prob(head)} = \lim_{M\to\infty} \frac{\text{\# of heads}}{M \text{ repeats}}$$

Example 2: Select a subject randomly from a sub-population S. Repeat this M times.

$$\text{Prob(cancer)} = \lim_{M\to\infty} \frac{\text{\# of subjects having cancer}}{M \text{ repeats}}$$

Question: What is the sub-population S?

There are many possibilities.

S = { all men }

S = { all persons over 50 years old }

S = { all Asian Americans }

S = { all foreign-born }

S = { all university professors }

S = { all persons with BMI in the normal range (18.5 ≤ BMI ≤ 24.9) }

Different doctors may view a given patient as a member of different sub-populations. As a result, different doctors may have different interpretation of Prob(cancer) for the same patient. This is why probability can be subjective. The subjectiveness is in specifying how to repeat the experiment.

So the probability "Prob(I have cancer)" makes sense only when the subject (me) is viewed as a member of a subpopulation. The selection of different subpopulations leads to different values of the probability, which is mathematically correct. The full statement should be something like the one below:

"When viewed as a random member of the subpopulation of all Asian Americans, the probability that I have cancer is … "

Observation:

Without specifying how the experiment is repeated, probability does not make sense.

With the concept of probability established in the framework of repeated experiments, we introduce terminology associated with probability.

Outcome of an experiment:

= Full description of the relevant result

Example: Flip a coin n times and view the sequence of flips as ONE experiment.

Outcome: $\omega = x_1\, x_2 \cdots x_n$

$x_j = 1$ (H, head) or 0 (T, tail)

This form of description is adequate for answering most questions.

However, if we want to study the possible connection between the height of toss and the landing result, this form of outcome is inadequate; we have to include in the outcome a sequence of n heights for the n tosses.

Sample space of an experiment

Ω = { all possible outcomes }

Example: Flip a coin 3 times and view the sequence of flips as ONE experiment.

Ω = { TTT, TTH, THT, THH, HTT, HTH, HHT, HHH }

Random variable:

= A function of outcpme.

Full notation: $X(\omega)$

Short notation: X

Example: Flip a coin n times and view the sequence of flips as ONE experiment.

N = # of heads in n flips

N is a random variable.

Full notation: $N(\omega) = N(x_1\, x_2 \cdots x_n) = \Sigma_j\, x_j$

Event:

= A subset of sample space

Example: Flip a coin 3 times and view the sequence of flips as ONE experiment

A = "exactly 2 heads in 3 flips"

= {THH, HTH, HHT}

Full notation: $A = \{\, \omega \mid N(\omega) = 2 \,\} = \{\, \omega \mid \omega \text{ contains exactly 2 heads} \,\}$

The description A = { ω | N(ω) = 2 } connects event A to random variable N(ω).

Not all events are associated with a random variable.

Example:

B = { ω | no consecutive heads in ω }

= {HTT, THT, TTH, HTH, TTT}

Probability of an event:

Pr(A) ≡ Pr(outcome ω ∈ A)

$$= \lim_{M\to\infty} \frac{\text{\# of } \omega \in A}{M \text{ repeats}}$$

Example: Flip a coin 3 times and view the sequence of flips as ONE experiment.

Pr(exactly 2 heads in 3 flips)

= 3/8 {THH, HTH, HHT}

Pr(no consecutive heads in 3 flips)

= 5/8 {HTT, THT, TTH, HTH, TTT}

Conditional probability:

Pr(A | B)

Repeat the experiment M times.

Consider only those repeats with ω ∈ B.

$$\Pr(A|B) = \lim_{M\to\infty} \frac{\text{\# of } \omega \in A \text{ (out of those with } \omega \in B)}{\text{\# of } \omega \in B}$$

$$= \lim_{M\to\infty} \frac{\dfrac{\text{\# of } \left(\omega \in A \text{ and } \omega \in B\right)}{M}}{\dfrac{\text{\# of } \omega \in B}{M}} = \frac{\Pr(AB)}{\Pr(B)}$$

Thus, we obtain

$$\boxed{\Pr(A|B) = \frac{\Pr(AB)}{\Pr(B)}}$$

Intersection: both A and B are true

AB = { ω | ω ∈ A and ω ∈ B }

Alternative notation for intersection: A ∩ B

Union: at least one of A and B is true

$A+B = \{ \omega \mid \omega \in A \text{ or } \omega \in B \text{ or both} \}$

Alternative notation for union: $A \cup B$

Complement: A is false

$A^C = \{ \omega \mid \omega \notin A \}$

For complement, we always have

$\Pr(A^C) = 1 - \Pr(A)$

Example:

Pr(exactly 2 heads in 3 flips AND no consecutive heads)

= 1/8 { HTH }

Pr(exactly 2 heads in 3 flips | no consecutive heads)

= 1/5 { HTH } out of {HTT, THT, TTH, HTH, TTT}

Independence of two events:

Intuition:

$\Pr(A|B) = \Pr(A)$, probability of A is not affected by occurrence of B

$$\Longleftrightarrow \quad \frac{\Pr(AB)}{\Pr(B)} = \Pr(A)$$

$\Longleftrightarrow \quad \Pr(AB) = \Pr(A)\Pr(B)$

Definition (independence of two events):

Events A and B are called independent if

$\Pr(AB) = \Pr(A)\Pr(B)$.

Expected value of a random variable

Random variable: $X(\omega)$

Notation for expected value:

$E(X)$, $E[X]$, $\langle X \rangle$

Repeat the experiment M times. Collect M outcomes, $\{\omega_j, j = 1, 2, \ldots, M\}$

$$E(X) = \lim_{M\to\infty} \frac{\sum_{j=1}^{M} X(\omega_j)}{M}$$

Probability mass function (PMF) (of a **discrete** random variable)

Random variable: $N(\omega)$

PMF of random variable $N(\omega)$:

$f_N(k) = \Pr(N(\omega) = k)$

Expected value in terms of PMF

$$E(N) = \lim_{M\to\infty} \frac{\sum_{j=1}^{M} N(\omega_j)}{M} = \lim_{M\to\infty} \frac{\sum_k k \times \left(\#\text{ of } N(\omega_j) = k\right)}{M}$$

$$= \sum_k k \times \left(\lim_{M\to\infty} \frac{\#\text{ of } N(\omega_j) = k}{M} \right)$$

$$= \sum_k k \Pr\left(N(\omega) = k\right) = \sum_k k\, f_N(k)$$

Example:

N = # of heads in n flips of a fair coin

$$\Pr(N = 0) = \left(\frac{1}{2}\right)^n$$

$$\Pr(N = 1) = n\left(\frac{1}{2}\right)^n$$

$$\Pr(N = k) = C(n,k)\left(\frac{1}{2}\right)^n$$

Probability density function (PDF) (of a **continuous** random variable)

Random variable: $X(\omega)$

$$\rho_X(x) = \lim_{\Delta x\to 0} \frac{\Pr\left(x \le X(\omega) < x + \Delta x\right)}{\Delta x}$$

Cumulative distribution function (CDF)

$$F_X(x) = \Pr\left(X(\omega) \le x\right)$$

Connection between CDF and PDF:

$$\rho_X(x) = \lim_{\Delta x \to 0} \frac{\Pr\left(x \le X(\omega) < x + \Delta x\right)}{\Delta x} = \lim_{\Delta x \to 0} \frac{F_X(x+\Delta x) - F_X(x)}{\Delta x} = \frac{d}{dx} F_X(x)$$

We obtain:

$$\boxed{\rho_X(x) = \frac{d}{dx} F_X(x)}$$

Expected value in terms of PDF

$$E(X) = \lim_{M \to \infty} \frac{\sum_{j=1}^{M} X(\omega_j)}{M} = \lim_{\substack{M \to \infty \\ \Delta x \to 0}} \frac{\sum_i x_i \left(\#\text{ of } x_i \le X(\omega_j) < x_{i+1}\right)}{M}, \quad x_i = x_0 + i\Delta x$$

$$= \lim_{\Delta x \to 0} \sum_i x_i \Pr\left(x_i \le X(\omega) < x_{i+1}\right)$$

$$= \lim_{\Delta x \to 0} \sum_i x_i \rho_X(x_i) \Delta x = \int x \rho_X(x) dx$$

We obtain:

$$\boxed{E(X) = \int x \rho_X(x) dx}$$

Law of total probability

Definition: (Partition of sample space Ω)

If $\{B_n, n = 1, 2, \dots\}$ satisfies

i) $B_1 + B_2 + \cdots = \Omega$

ii) $B_i B_j = \emptyset$ for all $i \neq j$

then $\{B_n, n = 1, 2, \dots\}$ is called a partition of Ω.

Theorem (the law of total probability)

Suppose $\{B_n, n = 1, 2, \dots\}$ is a partition of Ω. Then we have

$$\Pr(A) = \sum_n \Pr(AB_n)$$

$$\boxed{\Pr(A) = \sum_n \Pr(A | B_n) \Pr(B_n)}$$

This is called the law of total probability. This law is useful for calculating probability when conditional probabilities are easy to find.

Joint density:

$$\rho_{(X,Y)}(x,y)=\lim_{\substack{\Delta x\to 0\\ \Delta y\to 0}}\frac{\Pr\left(x\le X(\omega)<x+\Delta x\ \text{AND}\ y\le Y(\omega)<y+\Delta y\right)}{(\Delta x)(\Delta y)}$$

Conditional probability density:

$$\rho_X(x|B)=\lim_{\Delta x\to 0}\frac{\Pr\left(x\le X(\omega)<x+\Delta x\middle|B\right)}{\Delta x}$$

$$=\lim_{\Delta x\to 0}\frac{\Pr\left(x\le X(\omega)<x+\Delta x\ \text{AND}\ \ \omega\in B\right)}{(\Delta x)}\cdot\frac{1}{\Pr(B)}=\rho\left(X=x\ \&\ B\right)\cdot\frac{1}{\Pr(B)}$$

The above works when $\Pr(B) > 0$.

When $\Pr(B) = 0$, we can find a way to work around it.

$$\rho_X\left(x\middle|Y=y\right)=\lim_{\Delta y\to 0}\rho_X\left(x\middle|y\le Y<y+\Delta y\right)$$

$$=\lim_{\Delta x\to 0}\frac{\dfrac{\Pr\left(x\le X(\omega)<x+\Delta x\ \text{AND}\ y\le Y<y+\Delta y\right)}{(\Delta x)(\Delta y)}}{\dfrac{\Pr\left(y\le Y<y+\Delta y\right)}{(\Delta y)}}=\frac{\rho_{(X,Y)}(x,y)}{\rho_Y(y)}$$

We obtain

$$\boxed{\rho_X\left(x\middle|Y=y\right)=\frac{\rho_{(X,Y)}(x,y)}{\rho_Y(y)}}$$

Independence of two random variables

Intuition:

$\rho_X\left(x\middle|Y=y\right)=\rho_X(x)$, density of X is not affected by the value of Y.

<==> $\dfrac{\rho_{(X,Y)}(x,y)}{\rho_Y(y)}=\rho_X(x)$

<==> $\rho_{(X,Y)}(x,y)=\rho_X(x)\rho_Y(y)$

Definition (independence of two random variables):

Random variables X and Y are called independent if

$$\rho_{(X,Y)}(x,y)=\rho_X(x)\rho_Y(y).$$

<u>Conditional expectation:</u>

E(X|B)

Repeat the experiment M times. Collect M outcomes, $\{\omega_j, j = 1, 2, \ldots, M\}$

Consider only those repeats with $\omega_j \in B$.

$$E(X|B)=\lim_{M\to\infty}\frac{\sum_{\omega_j\in B}X(\omega_j)}{\#\text{ of }\omega_j\in B}$$

$$=\lim_{\substack{M\to\infty\\ \Delta x\to 0}}\frac{\sum_i x_i\dfrac{\#\text{ of }\left(x_i\le X(\omega_j)<x_{i+1}\text{ AND }\omega_j\in B\right)}{M}}{\dfrac{\#\text{ of }\omega_j\in B}{M}},\quad x_i=x_0+i\Delta x$$

$$=\lim_{\Delta x\to 0}\frac{\sum_i x_i\Pr\left(x_i\le X(\omega)<x_{i+1}\text{ AND }\omega\in B\right)}{\Pr(B)}=\int x\rho_X(x|B)dx$$

We obtain

$$\boxed{E(X|B)=\int x\rho_X(x|B)dx}$$

The above works when $\Pr(B) > 0$.

When $\Pr(B) = 0$, we can find a way to work around it.

Next we study E(X|Y=y)

$$E\left(X\middle|Y=y\right)=\lim_{\Delta y\to 0}E\left(X\middle|y\le Y<y+\Delta y\right)$$

$$=\lim_{\Delta y\to 0}\int x\rho_X\left(x\middle|y\le Y<y+\Delta y\right)dx=\int x\rho_X\left(x\middle|Y=y\right)dx$$

We obtain

$$\boxed{E(X|Y=y)=\int x\rho_X(x|Y=y)dx}$$

<u>Remarks:</u>

i) Notice that

$$\rho_X(x|Y=y)=\lim_{\Delta y\to 0}\rho_X\left(x\middle|y\le Y<y+\Delta y\right)$$

$$E(X|Y=y)=\lim_{\Delta y\to 0}E\left(X\middle|y\le Y<y+\Delta y\right)$$

ii) E(X|Y=y) is a function of y. When we apply this function to random variable Y, we get s derived random variable

$$E(X|Y) \equiv E(X|Y=y)\Big|_{y=Y}$$

is a function of Y, a derived random variable.

We can consider the expected value of random variable E(X|Y).

<u>Law of total expectation</u>

<u>Theorem</u> (the law of total expectation):

$$E(X) = E\Big(E(X|Y)\Big)$$

This is called the law of total expectation.

<u>A special case:</u>

Suppose { B_n, n = 1, 2, … } is a partition of Ω.

We define random variable *Y* as

$$Y(\omega) = j \text{ if } \omega \in B_j$$

Recall that *E*(*X*|*Y*) is a function of *Y*. We calculate *E*(*E*(*X*|*Y*))

$$E\Big(E(X|Y)\Big) = \sum_{j=1}^{n} E(X|Y=j)\Pr(|Y=j) = \sum_{n} E(X|B_n)\Pr(B_n)$$

Thus, we obtain

$$E(X) = \sum_{n} E(X|B_n)\Pr(B_n)$$