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, union, intersection, complement, mutually exclusive events
- Conditional probability, independent events, law of total probability
- Expected value of a random variable, probability mass function (PMF), probability density function (PDF), joint probability density of two random variables
- Conditional probability density, independent random variables, conditional expectation, 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 more puzzling when we think about it carefully.

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), the 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 ($18.5 \le \text{BMI} \le 24.9$) }

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 a different subpopulation leads to a different value 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, the concept of 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 n flips as ONE experiment.

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

x_i = 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. For that purpose, we have to include the heights of n tosses in the outcome.

Outcome: $\omega = v_1 v_2 \cdots v_n$

$$v_i = \begin{pmatrix} x_i \\ h_i \end{pmatrix}, \qquad x_i = \text{side facing up}, \quad h_i = \text{height}$$

Sample space of an experiment

Ω = { all possible outcomes }

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

The sample space is

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 n flips as ONE experiment.

Let 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 3 flips as ONE experiment

A = "exactly 2 heads in 3 flips"

= {THH, HTH, HHT}

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

In this example, event A is conveniently described by random variable $N(\omega)$.

Not all events are conveniently described by a random variable.

Example:

$B = \{ \omega \mid$ no consecutive heads in $\omega \}$

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

Event B is not conveniently described by a random variable.

Probability of an event:

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

Remark: Mathematically, Pr(A) is the measure of set A in space Ω.

Example: Flip a fair coin 3 times and view the sequence of 3 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}

Special cases:

$\Pr(\Omega) = 1$ ($\omega \in \Omega$ is always true)

$\Pr(\emptyset) = 0$ ($\omega \in \emptyset$ is never true where $\emptyset \equiv$ empty set)

Intersection of two events A and B: both A and B are true

$AB = \{ \omega \mid \omega \in A \text{ and } \omega \in B \}$ **Draw a Venn diagram to show it.**

Alternative notation for intersection: $A \cap B$

Union of two events A and B: at least one of A and B is true

$A+B = \{ \omega \mid \omega \in A \text{ or } \omega \in B \text{ or both} \}$ **Draw a Venn diagram to show it.**

Alternative notation for union: $A \cup B$

Complement of event A: A is false

$A^C = \{ \omega \mid \omega \notin A \}$ **Draw a Venn diagram to show it.**

For complement, we always have

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

Conditional probability: Pr(A | B)

Repeat the experiment M times.

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

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

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

Thus, we obtain

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

Example:

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

= 1/8 { HTH } out of all 8 outcomes

This is the probability of an intersection.

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

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

This is a conditional probability.

Independent events:

Intuition:

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

<==> $\dfrac{\Pr(AB)}{\Pr(B)} = \Pr(A)$

<==> Pr(AB) = Pr(A) Pr(B)

Definition (independent events):

Events A and B are called independent if

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

Mutually exclusive events:

Definition (mutually exclusive events):

Events A and B are called mutually exclusive if $AB = \emptyset$.

Draw a Venn diagram to show it.

Note: Two mutually exclusive events are definitely not independent.

Law of total probability

Definition: (Partition of sample space Ω)

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

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

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

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

Draw a Venn diagram to show it.

Example: A simple partition of Ω.

$\Omega = A + A^C$

Theorem (the law of total probability)

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

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

Proof: $\Pr(A) = \sum_n \Pr(AB_n) = \sum_n \Pr(A \mid B_n)\Pr(B_n)$.

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

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, $\{\omega_j, j = 1, 2, \ldots, M\}$

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

Properties of expected value:

- $E(\alpha X) = \alpha E(X)$
- $E(X+Y) = E(X) + E(Y)$ for all X and Y

  Proof: $\sum_{j=1}^{M}\left(X(\omega_j)+Y(\omega_j)\right) = \sum_{j=1}^{M} X(\omega_j) + \sum_{j=1}^{M} Y(\omega_j)$

Note: Operationally, this definition is not practical for calculating $E(X)$. To establish a convenient mathematical formulation we introduce probability mass function.

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

Random variable: $N(\omega)$

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

$p_N(k) \equiv \Pr(N(\omega) = k)$

Note: The statistical behavior of a discrete random variable is completely described by its probability mass function (PMF).

Example:

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

$$\Pr(N=0) = (1/2)^n$$

$$\Pr(N=1) = n(1/2)^n$$

$$\Pr(N=k) = C(n,k)(1/2)^n, \qquad k = 0,1,2,\ldots,n$$

where $C(n, k)$ = # of ways of choosing an unordered subset of k elements from n elements. $C(n, k)$ has the expression

$$C(n,k) = \frac{k!(n-k)!}{n!}$$

$C(n, k)$ is called the binomial coefficient because it appears in …

$$(a+b)^n = a^n + C(n,1)a^{n-1}b^1 + C(n,2)a^{n-2}b^2 + \cdots$$

Expected value (of a discrete random variable) in terms of PMF

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

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

We obtain:

$$\boxed{E(N) = \sum_k k\, p_N(k)}$$

Expected value of $f(N)$

$$E(f(N)) = \sum_k f(k) p_N(k)$$

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

Random variable: $X(\omega)$

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

Note: Here we use $\Pr\big(x < X(\omega) \le x+\Delta x\big)$ for its simplicity. To properly accommodate both continuous and discrete random variables, we should use

$\Pr\big(x-\Delta x/2 < X(\omega) \le x+\Delta x/2\big)$.

Cumulative distribution function (CDF)

$$F_X(x) = \Pr\big(X(\omega) \le x\big)$$

Connection between CDF and PDF:

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

We obtain:

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

Expected value (of a continuous random variable) in terms of PDF

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

We divide the real axis into bins of size Δx. Let $x_i = x_0 + i\,\Delta x$.

$$E(X) = \lim_{\Delta x \to 0} \lim_{M \to \infty} \frac{\sum_i x_i \left(\#\text{ of } \omega_j \text{ satisfying } x_i < X(\omega_j) \le x_{i+1} \right)}{M}$$

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

We obtain:

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

Expected value of $f(X)$

$$E(f(X)) = \int f(x) \rho_X(x) dx$$

Remark: With the notation of δ function, we can treat a discrete random variable as continuous where $\rho_X(x)$ is a sum of δ functions.

Example:

$$X = \begin{cases} 1, & \text{prob} = 0.6 \\ 0, & \text{prob} = 0.4 \end{cases}, \qquad \rho_X(x) = 0.4\delta(x) + 0.6\delta(x-1)$$

Joint density of two random variables (X, Y)

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

Conditional probability density: $\rho_X(x|Y=y)$

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

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

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

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

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

(again, we should use $\rho_X\left(x \middle| y - \Delta y/2 < Y \le y + \Delta y/2 \right)$).

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

We obtain

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

<u>Independent random variables</u>

<u>Intuition:</u>

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

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

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

<u>Definition</u> (independent 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),   E(X|Y=y)

We first study E(X|B).

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

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

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

(We divide the real axis into bins of size Δx. Let $x_i = x_0 + \Delta x$).

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

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

We obtain

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

The above works when Pr(B) > 0.

Next we study E(X|Y=y). When Pr(B) = 0, we find a way to work around it.

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

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

We obtain

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

Remarks:

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

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

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

Law of total expectation

Theorem (the law of total expectation):

$$\boxed{E(X) = E\left(E(X|Y)\right)}$$

This is called the law of total expectation.

Proof:

$$E\left(E(X|Y)\right) = \int E(X|Y=y)\rho_Y(y)dy = \int\left(\int x\rho_X(x|Y=y)dx\right)\rho_Y(y)dy$$

$$= \int x(\int \underbrace{\rho_X(x|Y=y)\rho_Y(y)}_{\rho_{(X,Y)}(x,y)}dy)dx = \int x(\int \rho_{(X,Y)}(x,y)dy)dx = E(X)$$

<u>A special case</u> of the law of total expectation:

Suppose $\{B_n, n = 1, 2, \ldots\}$ is a partition of Ω. We define random variable Y as

$$Y(\omega) = n \text{ where } \omega \in B_n$$

Recall that $E(X|Y)$ is a function of discrete random variable Y. We calculate its expected value directly using $E\big(f(Y)\big) = \sum_n f(n) p_Y(n)$.

$$E\big(E(X|Y)\big) = \sum_n E(X|Y = n)\Pr(|Y = n) = \sum_n E(X|B_n)\Pr(B_n)$$

From the law of total expectation, we obtain

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