

Introduction to probability

Nemesh N. T.

1 Probability

1.1 Foundations

1.1.1 Events

Explanation and notation

1. \emptyset — is an event that never happens
2. Ω — all elementary events
3. Some σ -algebra on Ω . Its elements are called events.
4. Let A, B be some events (may be composite events). Then
 - $A \cap B$ — both A and B happened;
 - $A \cup B$ — at least A or B or both happened;
 - $A \setminus B$ — A happened, but B didn't happen;
 - \overline{A} — A didn't happen;
 - $A \Delta B$ — either A or B happened but not both.
4. Note 2^Ω denotes all events, or in other words all subsets in Ω .

Example. Suppose we are given a fair dice with six sides. Let $\Omega = \{\omega_1, \omega_2, \dots, \omega_6\}$ be 6 possible outcomes and let $A = \{\omega_1, \omega_3\}$ and $B = \{\omega_3, \omega_4, \omega_5\}$ be some events. Then

$$A \cap B = \{\omega_3\} \tag{1}$$

$$A \cup B = \{\omega_1, \omega_3, \omega_4, \omega_5\} \tag{2}$$

$$A \setminus B = \{\omega_1\} \tag{3}$$

$$B \setminus A = \{\omega_4, \omega_5\} \tag{4}$$

$$\overline{A} = \{\omega_2, \omega_4, \omega_5, \omega_6\} \tag{5}$$

$$\overline{B} = \{\omega_1, \omega_2, \omega_6\} \tag{6}$$

Proposition. De Morgan's laws

$$\overline{\bigcup_{i=1}^n A_i} = \bigcap_{i=1}^n \overline{A_i} \tag{7}$$

$$\overline{\bigcap_{i=1}^n A_i} = \bigcup_{i=1}^n \overline{A_i} \tag{8}$$

$$\tag{9}$$

In particular,

$$\overline{A \cup B} = \overline{A} \cap \overline{B} \tag{10}$$

$$\overline{A \cap B} = \overline{A} \cup \overline{B} \tag{11}$$

1.1.2 Probability space

Probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω — is a set of elementary events and $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ — is a probability function satisfying axioms of probability.

Axioms of probability:

$$\mathbb{P}(\emptyset) = 0 \quad (12)$$

$$\mathbb{P}(\Omega) = 1 \quad (13)$$

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \text{ whenever } A, B \in \mathcal{F} \quad A \cap B = \emptyset \quad (14)$$

Properties

$$\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A) \quad (15)$$

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \quad (16)$$

$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(B \cap C) - \mathbb{P}(C \cap A) + \mathbb{P}(A \cap B \cap C) \quad (17)$$

Inclusion exclusion principle

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i_1} \mathbb{P}(A_{i_1}) \quad (18)$$

$$- \sum_{i_1 < i_2} \mathbb{P}(A_{i_1} \cap A_{i_2}) \quad (19)$$

$$+ \sum_{i_1 < i_2 < i_3} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3}) \quad (20)$$

$$\dots \quad (21)$$

$$+ (-1)^{k-1} \sum_{i_1 < i_2 < \dots < i_k} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) \quad (22)$$

$$\dots \quad (23)$$

$$+ (-1)^{n-1} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_n}) \quad (24)$$

1.1.3 Examples of probability spaces

Example. *A fair dice with six sides.*

We have

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_6\}$$

where ω_i — is an elementary event where we scored i points.

$$\mathcal{F} = 2^\Omega = \{\emptyset, \{\omega_1\}, \dots, \{\omega_6\}, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \dots, \{\omega_5, \omega_6\}, \{\omega_1, \omega_2, \omega_3\}, \dots, \{\omega_1, \omega_2, \dots, \omega_6\}\}$$

$$\mathbb{P}(\omega_1) = \mathbb{P}(\omega_2) = \dots = \mathbb{P}(\omega_6) = \frac{1}{6}$$

As the consequence

$$\mathbb{P}(\{\omega_1, \omega_4, \omega_2\}) = \mathbb{P}(\omega_1) + \mathbb{P}(\omega_4) + \mathbb{P}(\omega_2) = \frac{3}{6} = \frac{1}{2}$$

Example. *Assume we are given a **nonfair** dice with six sides.*

We have

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_6\}$$

where ω_i — is an elementary event where we scored i points.

$$\mathbb{P}(\omega_1) = 0.9 \quad (25)$$

$$\mathbb{P}(\omega_2) = \dots = \mathbb{P}(\omega_6) = 0.02 \quad (26)$$

Example. *Bernoulli trials.*

(a) *Tossing a coin (1 trial). Probability space*

$$\Omega = \{0, 1\}, \quad \mathcal{F} = 2^\Omega$$

$$\mathbb{P}(\{0\}) = p, \quad \mathbb{P}(\{1\}) = 1 - p$$

(b) *Tossing a coin (n trials). Probability space*

Ω — all n -tuples of zeros and ones, that is

$$\Omega = \{(0, 0, \dots, 0, 0), (0, 0, \dots, 0, 1), (0, 0, \dots, 1, 0), (0, 0, \dots, 1, 1), \dots, (1, 1, \dots, 1)\}$$

$$\mathcal{F} = 2^\Omega$$

There are 2^n elementary events here.

Let A be an elementary event, say $A = (0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, \dots, 0)$ with k heads and $n - k$ — tails.

$$\mathbb{P}(A) = p^k (1 - p)^{n-k}$$

Consider event B_k : there were k heads after n tosses. What's probability of B ?

$$\mathbb{P}(B_k) = C_n^k p^k (1 - p)^{n-k}$$

Because B_k consist of C_n^k elementary events each with k heads and $n - k$ tails.

Consider event C_k : there were at least k heads with n tosses. Clearly,

$$C_k = B_k \cup B_{k+1} \cup \dots \cup B_n$$

$$\mathbb{P}(C_k) = \mathbb{P}(B_k \cup B_{k+1} \cup \dots \cup B_n) = \sum_{i=k}^n \mathbb{P}(B_i) = \sum_{i=k}^n C_n^i p^i (1 - p)^{n-i}$$

Consider event D_k : there were less than k heads with n tosses.

$$\mathbb{P}(D_k) = \sum_{i=0}^{k-1} C_n^i p^i (1 - p)^{n-i}$$

1.1.4 Conditional probability

Definition. Conditional probability $\mathbb{P}(A|B)$ of event A given B has happened defined as

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad (27)$$

Law of total probability. Assume $B_1 \cup B_2 \cup \dots \cup B_n = \Omega$ and all B_i 's are disjoint

$$\mathbb{P}(A) = \mathbb{P}(A \cap B_1) + \mathbb{P}(A \cap B_2) + \dots + \mathbb{P}(A \cap B_n) \quad (28)$$

$$= \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) + \dots + \mathbb{P}(A|B_n)\mathbb{P}(B_n) \quad (29)$$

$$(30)$$

Corollaries

$$\mathbb{P}(A|C) = \mathbb{P}(A|B_1, C)\mathbb{P}(B_1) + \mathbb{P}(A|B_2, C)\mathbb{P}(B_2) + \dots + \mathbb{P}(A|B_n, C)\mathbb{P}(B_n) \quad (31)$$

$$= \mathbb{P}(A \cap B_1|C) + \mathbb{P}(A \cap B_2|C) + \dots + \mathbb{P}(A \cap B_n|C) \quad (32)$$

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|\bar{B})\mathbb{P}(\bar{B}) \quad (33)$$

$$= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \bar{B}) \quad (34)$$

Bayes' rules

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} \quad (35)$$

$$\mathbb{P}(A|B, C) = \frac{\mathbb{P}(B|A, C)\mathbb{P}(A|C)}{\mathbb{P}(B|C)} \quad (36)$$

1.1.5 Independent events

Definition. We say that events A_1, \dots, A_n are independent if for any distinct indices $1 < i_1 < \dots < i_k < n$ holds

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_k}) \quad (37)$$

In particular, events A and B are independent if and only if one of the following holds

- $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
- $\mathbb{P}(A|B) = \mathbb{P}(A)$
- $\mathbb{P}(B|A) = \mathbb{P}(B)$

1.2 Random variables

Random variable (RV for short) is an \mathcal{F} -measurable function $X : \Omega \rightarrow \mathbb{R}$.

1.2.1 Distribution

Definition. Cumulative distribution function (CDF) of an RV X is defined as

$$F_X(t) = \mathbb{P}(\{X \leq t\}) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \leq t\}) \quad (38)$$

Properties of CDF

- $F_X(t)$ is a non-decreasing function
- $F_X(t) \rightarrow 0$ as $t \rightarrow -\infty$
- $F_X(t) \rightarrow 1$ as $t \rightarrow 1$
- for any c we have $F(t) \rightarrow F(c)$ as $t \rightarrow c + 0$

It is clear from definition, that

$$\mathbb{P}(\{a < X \leq b\}) = F_X(b) - F_X(a) \quad (39)$$

An RV is called continuous random variable (CRV) if its CDF is absolutely continuous. If an RV is not continuous it is called discrete random variable (DRV).

For any CRV X we always have $\mathbb{P}(X = t) = 0$, so

$$\mathbb{P}(X \leq t) = \mathbb{P}(X < t) = F_X(t) \quad (40)$$

Definition. Let X be a CRV, then its probability density function (PDF) is defined as

$$f_X(t) = \frac{d}{dt}F_X(t)$$

Let X be a DRV, then its probability density function (PDF) is defined as

$$f_X(t) = \mathbb{P}(X = t)$$

Properties of PDF

- $f_X(t)$ is non-negative
- If X is a CRV, then

$$\int_{-\infty}^{\infty} f_X(t) dt = 1 \quad (41)$$

- if X is a DRV, then

$$\sum_t f_X(t) = 1 \quad (42)$$

Transition from PDF to CDF

- If X is a CRV, then

$$F_X(t) = \int_{-\infty}^t f_X(s) ds \quad (43)$$

- If X is a DRV, then

$$F_X(t) = \sum_{s < t} f_X(s) \quad (44)$$

1.2.2 Independent random variables

Definition. We say that two RV's X and Y are independent if

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \quad (45)$$

for all $A \subset \mathbb{R}, B \subset \mathbb{R}$.

In particular two DRV X and Y are independent if and only if

$$\mathbb{P}(X = t, Y = s) = \mathbb{P}(X = t)\mathbb{P}(Y = s) \quad (46)$$

1.2.3 Joint and conditional distribution

Definition. Joint CDF of two RV's X, Y is defined as

$$F_{X,Y}(t, s) = \mathbb{P}(X < t, Y < s) \quad (47)$$

It is immediate from definition, that for independent RV's we have

$$F_{X,Y}(t, s) = F_X(t)F_Y(s) \quad (48)$$

Definition. A joint PDF of two CRV's X, Y is defined as

$$f_{X,Y}(t, s) = \frac{d}{dt} \frac{d}{ds} F_{X,Y}(s, t) = \frac{d}{ds} \frac{d}{dt} F_{X,Y}(s, t) \quad (49)$$

A joint PDF of two DRV's X, Y is defined as

$$f_{X,Y}(t, s) = \mathbb{P}(X = t, Y = s) \quad (50)$$

1.2.4 Expected value and around it

Definition. Expected value of an RV X is defined as

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) \mathbb{P}(\omega) \quad (51)$$

In particular for any DRV we have

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\omega) \quad (52)$$

Example. A fair dice with six sides. Let X be a random variable such that $X(\omega_i) = i^2$, then

$$\mathbb{E}[X] = X(\omega_1)\mathbb{P}(\omega_1) + X(\omega_2)\mathbb{P}(\omega_2) + \dots + X(\omega_6)\mathbb{P}(\omega_6) \quad (53)$$

$$= X(\omega_1) \cdot \frac{1}{6} + X(\omega_2) \cdot \frac{1}{6} + \dots + X(\omega_6) \cdot \frac{1}{6} \quad (54)$$

$$= \frac{X(\omega_1) + \dots + X(\omega_6)}{6} \quad (55)$$

$$= \frac{1^2 + 2^2 + \dots + 6^2}{6} \quad (56)$$

$$= \frac{91}{6} \quad (57)$$

$$\approx 15.16 \quad (58)$$

Example. A *nonfair* dice with six sides. Let X be a random variable such that $X(\omega_i) = i^2$, then

$$\mathbb{E}[X] = X(\omega_1)\mathbb{P}(\omega_1) + X(\omega_2)\mathbb{P}(\omega_2) + \dots + X(\omega_6)\mathbb{P}(\omega_6) \quad (59)$$

$$= X(\omega_1) \cdot 0.9 + X(\omega_2) \cdot 0.02 + \dots + X(\omega_6) \cdot 0.02 \quad (60)$$

$$= 2.7 \quad (61)$$

$$(62)$$

Properties of expectation

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] \quad (63)$$

$$\mathbb{E}[aX] = a\mathbb{E}[X] \quad a = \text{const} \quad (64)$$

$$\mathbb{E}[a] = a \quad a = \text{const} \quad (65)$$

Efficient ways to find expectations

- If X is a CRV, then

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} t f_X(t) dt \quad (66)$$

- If X is a DRV, then

$$\mathbb{E}[X] = \sum_t t f_X(t) = \sum_t t \mathbb{P}(X = t) \quad (67)$$

Proposition. Law of unconscious statistician (LOTUS for short) For X — DRV with values $\{x_1, x_2, \dots\}$. Consider DRV $Y = g(X)$ with values $\{y_1, y_2, \dots\}$. Then

$$\mathbb{E}[g(X)] = \sum_{i=1}^{\infty} y_i \mathbb{P}(Y = y_i) \quad (68)$$

$$= \sum_{i=1}^{\infty} g(x_i) \mathbb{P}(X = x_i) \quad (69)$$

For X — CRV with PDF $f_X(t)$ consider CRV $Y = g(X)$ with PDF $f_Y(t)$. Then

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} t f_Y(t) dt \quad (70)$$

$$= \int_{-\infty}^{\infty} g(t) f_X(t) dt \quad (71)$$

$$(72)$$

Example. Let $X \sim U[-1, 2]$. Find $\mathbb{E}[X^2]$ using LOTUS and without it. Using LOTUS

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} t^2 f_X(t) dt = \int_{-1}^2 t^2 \frac{1}{2 - (-1)} dt = \frac{1}{3} \frac{t^3}{3} \Big|_{-1}^2 = 1 \quad (73)$$

In the second case we had to derive PDF for $Y = X^2$

$$F_Y(t) = \mathbb{P}(Y < t) = \mathbb{P}(X^2 < t) = \begin{cases} \mathbb{P}(-\sqrt{t} < X < \sqrt{t}) & \text{if } t > 0 \\ \mathbb{P}(\emptyset) & \text{if } t \leq 0 \end{cases} = \begin{cases} \mathbb{P}(-\sqrt{t} < X < \sqrt{t}) & \text{if } t > 0 \\ \mathbb{P}(\emptyset) & \text{if } t \leq 0 \end{cases} \quad (74)$$

$$= \begin{cases} \mathbb{P}(\Omega) & \text{if } t \geq 4 \\ \mathbb{P}(-1 \leq X < \sqrt{t}) & \text{if } 1 \leq t \leq 4 \\ \mathbb{P}(-\sqrt{t} < X < \sqrt{t}) & \text{if } 0 < t < 1 \\ \mathbb{P}(\emptyset) & \text{if } t \leq 0 \end{cases} = \begin{cases} 1 & \text{if } t \geq 4 \\ \int_{-1}^{\sqrt{t}} f_X(s) ds & \text{if } 1 \leq t \leq 4 \\ \int_{-\sqrt{t}}^{\sqrt{t}} f_X(s) ds & \text{if } 0 < t < 1 \\ 0 & \text{if } t \leq 0 \end{cases} \quad (75)$$

$$= \begin{cases} 1 & \text{if } t \geq 4 \\ \int_{-1}^{\sqrt{t}} \frac{1}{3} ds & \text{if } 1 \leq t \leq 4 \\ \int_{-\sqrt{t}}^{\sqrt{t}} \frac{1}{3} ds & \text{if } 0 < t < 1 \\ 0 & \text{if } t \leq 0 \end{cases} = \begin{cases} 1 & \text{if } t \geq 4 \\ \frac{\sqrt{t}+1}{3} & \text{if } 1 \leq t \leq 4 \\ \frac{2\sqrt{t}}{3} & \text{if } 0 < t < 1 \\ 0 & \text{if } t \leq 0 \end{cases} \quad (76)$$

$$(77)$$

Now we are able to find PDF of Y

$$f_Y(t) = \frac{d}{dt} F_Y(t) = \begin{cases} 0 & \text{if } t \geq 4 \\ \frac{1}{6\sqrt{t}} & \text{if } 1 \leq t \leq 4 \\ \frac{1}{3\sqrt{t}} & \text{if } 0 < t < 1 \\ 0 & \text{if } t \leq 0 \end{cases} \quad (78)$$

$$(79)$$

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} t f_Y(t) dt = \int_0^1 t \frac{1}{3\sqrt{t}} dt + \int_1^4 t \frac{1}{6\sqrt{t}} dt = \frac{1}{3} \int_0^1 \sqrt{t} dt + \frac{1}{6} \int_1^4 \sqrt{t} dt \quad (80)$$

$$= \frac{1}{3} \left. \frac{t^{3/2}}{3/2} \right|_0^1 + \frac{1}{6} \left. \frac{t^{3/2}}{3/2} \right|_1^4 = \frac{2}{9} \left. t^{3/2} \right|_0^1 + \frac{2}{18} \left. t^{3/2} \right|_1^4 = 1 \quad (81)$$

1.2.5 Variance, deviation, moments

Definition. Let X be an RV then

- its variance is defined as

$$\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] \quad (82)$$

- its standard deviation is defined as

$$\mathbb{SD}[X] = \sqrt{\mathbb{V}[X]} \quad (83)$$

- its k -th moment is defined as

$$\mu_k(X) = \mathbb{E}[X^k] \quad (84)$$

Other ways to find variance and moments

- For any RV holds

$$\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mu_2(X) - \mu_1(X)^2 \quad (85)$$

- If X is a CRV, then

$$\mathbb{V}[X] = \int_{-\infty}^{\infty} (t - \mathbb{E}[X])^2 f_X(t) dt = \int_{-\infty}^{\infty} t^2 f_X(t) dt - \left(\int_{-\infty}^{\infty} t f_X(t) dt \right)^2 \quad (86)$$

$$\mu_k(X) = \int_{-\infty}^{\infty} t^k f_X(t) dt \quad (87)$$

- If X is a DRV, then

$$\mathbb{V}[X] = \sum_t (t - \mathbb{E}[X])^2 \mathbb{P}(X = t) = \sum_t t^2 \mathbb{P}(X = t) - \left(\sum_t t \mathbb{P}(X = t) \right)^2 \quad (88)$$

$$\mu_k(X) = \sum_t t^k \mathbb{P}(X = t) \quad (89)$$

Definition. Let X and Y be two RV's, then

- Their covariation is defined as

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \quad (90)$$

- Their correlation is defined as

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\mathbb{V}[X]\mathbb{V}[Y]}} \quad (91)$$

Other ways to find correlation and covariation

- If X and Y are RVs, then

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \quad (92)$$

In particular,

$$\text{Cov}(X, X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{V}[X] \quad (93)$$

- If X and Y are CRVs, then

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ts f_{X,Y}(ts) dt ds - \left(\int_{-\infty}^{\infty} t f_X(t) dt \right) \left(\int_{-\infty}^{\infty} s f_Y(s) ds \right) \quad (94)$$

- If X and Y are DRVs, then

$$\text{Cov}(X, Y) = \sum_t \sum_s ts \mathbb{P}(X = t, Y = s) - \left(\sum_t t \mathbb{P}(X = t) \right) \left(\sum_s s \mathbb{P}(Y = s) \right) \quad (95)$$

Properties of variance and covariation

- For any RVs X_1, \dots, X_n holds

$$\mathbb{V}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{V}[X_i] + 2 \sum_{i < j} \text{Cov}(X_i, X_j) \quad (96)$$

In particular,

$$\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\text{Cov}(X, Y) \quad (97)$$

- If X and Y are independent RVs, then

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0 \quad (98)$$

$$\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] \quad (99)$$

1.2.6 Moment generating function

Definition. Let X be a RV, then its moment generating function (MGF for short) is defined as

$$M_X(t) = \mathbb{E}[e^{tX}] \quad (100)$$

Other ways to compute MGF

- If X is a CRV, then

$$M_X(t) = \int_{-\infty}^{\infty} e^{ts} f_X(s) ds \quad (101)$$

- If X is a DRV, then

$$M_X(t) = \sum_s e^{ts} \mathbb{P}(X = s) \quad (102)$$

Why the name?

Proposition.

$$\mu_k(X) = \left. \frac{d^k}{dt^k} M_X(t) \right|_{t=0} \quad (103)$$

Proof. On the one hand

$$M_X(t) = \mathbb{E}[e^{tX}] = \mathbb{E}\left[\sum_{k=0}^{\infty} \frac{t^k X^k}{k!}\right] = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathbb{E}[X^k] = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mu_k(X) \quad (104)$$

On the other hand, by Taylor series expansion

$$M_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \left. \frac{d^k}{dt^k} M_X(t) \right|_{t=0} \quad (105)$$

And we get the desired result. \square

Example. Let $X \sim \text{Expo}(\lambda)$, that is

$$f_X(t) = \begin{cases} \lambda e^{-\lambda t} & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases} \quad (106)$$

Find moment generating function of X . Find first and second moments of X , find variance of X .

$$M_X(t) = \int_{-\infty}^{\infty} e^{ts} f_X(s) ds = \int_0^{\infty} e^{ts} \lambda e^{-\lambda s} ds = \lambda \int_0^{\infty} e^{(t-\lambda)s} ds = \lambda \frac{1}{t-\lambda} e^{(t-\lambda)s} \Big|_0^{\infty} = \begin{cases} \frac{\lambda}{\lambda-t} & \text{if } t < \lambda \\ +\infty & \text{if } t > \lambda \end{cases} \quad (107)$$

$$\mu_1(X) = \mathbb{E}[X] = \left. \frac{d}{dt} M_X(t) \right|_{t=0} = \left. \frac{\lambda}{(\lambda-t)^2} \right|_{t=0} = \frac{1}{\lambda} \quad (108)$$

$$\mu_2(X) = \mathbb{E}[X^2] = \left. \frac{d^2}{dt^2} M_X(t) \right|_{t=0} = \left. \frac{2\lambda}{(\lambda-t)^3} \right|_{t=0} = \frac{2}{\lambda^2} \quad (109)$$

$$\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mu_2(X) - \mu_1(X)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2} \quad (110)$$