Introduction to probability

Nemesh N. T.

1 Probability

1.1 Foundations

1.1.1 Events

Explanation and notation

1. $\varnothing$ — 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;
- $A \triangle B$ — 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 $\Omega = \{\omega_1, \omega_2, \ldots .\omega_6\}$ be 6 possible outcomes and let $A = \{\omega_1, \omega_3\}$ and $B = \{\omega_3, \omega_4, \omega_5\}$ be some events. Then*

$$A \cap B = \{\omega_3\} \tag{1}$$
$$A \cup B = \{\omega_1, \omega_3, \omega_4, \omega_5\} \tag{2}$$
$$A \setminus B = \{\omega_1\} \tag{3}$$
$$B \setminus A = \{\omega_4, \omega_4\} \tag{4}$$
$$\overline{A} = \{\omega_4, \omega_5, \omega_6, \omega_2\} \tag{5}$$
$$\overline{B} = \{\omega_1, \omega_2, \omega_6\} \tag{6}$$

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 $\mathbb{P} : \mathcal{F} \to [0, 1]$ — is a probability function satisfying axioms of pobability.

Axioms of probability:

$$\mathbb{P}(\varnothing) = 0 \tag{12}$$
$$\mathbb{P}(\Omega) = 1 \tag{13}$$
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \text{ whenever } A, B \in \mathcal{F} \quad A \cap B = \varnothing \tag{14}$$

Properties

$$\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A) \tag{15}$$
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \tag{16}$$
$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(B \cap C) - \mathbb{P}(C \cap A) + \mathbb{P}(A \cap B \cap C) \tag{17}$$

Inclusion exclusion principle

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \mathbb{P}(A_1 \cup A_2 \cup \ldots \cup A_n) &= \sum_{i_1} \mathbb{P}(A_{i_1}) && (18)\\
&- \sum_{i_1 < i_2} \mathbb{P}(A_{i_1} \cap A_{i_2}) && (19)\\
&+ \sum_{i_1 < i_2 < i_3} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3}) && (20)\\
&\ldots && (21)\\
&+ (-1)^{k-1} \sum_{i_1 < i_2 < \ldots < i_k} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) && (22)\\
&\ldots && (23)\\
&+ (-1)^{n-1} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_n}) && (24)
\end{aligned}$$

1.1.3 Examples of probability spaces

Example. *A fair dice with six sides.*

We have

$$\Omega = \{\omega_1, \omega_2, \ldots . \omega_6\}$$

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

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

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

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

$$\Omega = \{\omega_1, \omega_2, \ldots . \omega_6\}$$

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

$$\mathbb{P}(\omega_1) = 0.9 \tag{25}$$
$$\mathbb{P}(\omega_2) = \ldots = \mathbb{P}(\omega_6) = 0.02 \tag{26}$$

Example. *Bernoulli trials.*

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

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

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

(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,

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

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

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*

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

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

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A \cap B_1) + \mathbb{P}(A \cap B_2) + \ldots + \mathbb{P}(A \cap B_n) && (28)\\
&= \mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) + \ldots + \mathbb{P}(A|B_n)\mathbb{P}(B_n) && (29)\\
& && (30)
\end{aligned}$$

Corollaries

$$\begin{aligned}
\mathbb{P}(A|C) &= \mathbb{P}(A|B_1, C)\mathbb{P}(B_1) + \mathbb{P}(A|B_2, C)\mathbb{P}(B_2) + \ldots + \mathbb{P}(A|B_n, C)\mathbb{P}(B_n) && (31)\\
&= \mathbb{P}(A \cap B_1|C) + \mathbb{P}(A \cap B_2|C) + \ldots + \mathbb{P}(A \cap B_n|C) && (32)
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|\overline{B})\mathbb{P}(\overline{B}) && (33)\\
&= \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \overline{B}) && (34)
\end{aligned}$$

Bayes' rules

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

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

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*

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

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 $X : \Omega \to \mathbb{R}$.

1.2.1 Distribution

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

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

Properties of CDF

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

It is clear from definition, that

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

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

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

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

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

- if X is a DRV, then

$$\sum_t f_X(t) = 1 \tag{42}$$

Transition from PDF to CDF

- If X is a CRV, then

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

- If X is a DRV, then

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

1.2.2 Independent random variables

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

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

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

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

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

1.2.3 Joint and conditional distribution

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

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

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

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

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

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

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

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

1.2.4 Expected value and around it

Definition. *Expected value of an RV X is defined as*

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

In particular for any DRV we have

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

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

$$\begin{aligned}
\mathbb{E}[X] &= X(\omega_1)\mathbb{P}(\omega_1) + X(\omega_2)\mathbb{P}(\omega_2) + \ldots + X(\omega_6)\mathbb{P}(\omega_6) && (53)\\
&= X(\omega_1) \cdot \frac{1}{6} + X(\omega_2) \cdot \frac{1}{6} + \ldots + X(\omega_6) \cdot \frac{1}{6} && (54)\\
&= \frac{X(\omega_1) + \ldots + X(\omega_6)}{6} && (55)\\
&= \frac{1^2 + 2^2 + \ldots + 6^2}{6} && (56)\\
&= \frac{91}{6} && (57)\\
&\approx 15.16 && (58)
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}[X] &= X(\omega_1)\mathbb{P}(\omega_1) + X(\omega_2)\mathbb{P}(\omega_2) + \ldots + X(\omega_6)\mathbb{P}(\omega_6) && (59)\\
&= X(\omega_1) \cdot 0.9 + X(\omega_2) \cdot 0.02 + \ldots + X(\omega_6) \cdot 0.02 && (60)\\
&= 2.7 && (61)\\
& && (62)
\end{aligned}$$

Properties of expectation

$$\begin{aligned}
\mathbb{E}[X+Y] &= \mathbb{E}[X] + \mathbb{E}[Y] && (63)\\
\mathbb{E}[aX] &= a\mathbb{E}[X] \quad a = \text{const} && (64)\\
\mathbb{E}[a] &= a \quad a = \text{const} && (65)
\end{aligned}$$

Efficient ways to find expectations

- If X is a CRV, then

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

- If X is a DRV, then

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

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

$$\begin{aligned}
\mathbb{E}[g(X)] &= \sum_{i=1}^{\infty} y_i \mathbb{P}(Y = y_i) && (68)\\
&= \sum_{i=1}^{\infty} g(x_i)\mathbb{P}(X = x_i) && (69)
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}[g(X)] &= \int_{-\infty}^{\infty} t f_Y(t) dt && (70)\\
&= \int_{-\infty}^{\infty} g(t) f_X(t) dt && (71)\\
& && (72)
\end{aligned}$$

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

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

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

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

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

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

$$\tag{77}$$

Now we are able to find PDF of Y

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

$$\tag{79}$$

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

$$= \frac{1}{3}\frac{t^{3/2}}{3/2}\Bigg|_0^1 + \frac{1}{6}\frac{t^{3/2}}{3/2}\Bigg|_1^4 = \frac{2}{9}t^{3/2}\Bigg|_0^1 + \frac{2}{18}t^{3/2}\Bigg|_1^4 = 1 \tag{81}$$

1.2.5 Variance, deviation, moments

Definition. *Let X be an RV then*

- *its variance is defined as*

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

- *its standard deviation is defined as*

$$\mathbb{SD}[X] = \sqrt{\mathbb{V}[X]} \tag{83}$$

- *its k-th moment is defined as*

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

Other ways to find variance and moments

- For any RV holds

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

- If X is a CRV, then

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

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

- If X is a DRV, then

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

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

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

- *Their covariation is defined as*

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

- *Their correlation is defined as*

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

Other ways to find correlation and covariation

- If X and Y are RVs, then

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

  In particular,

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

- If X and Y are CRVs, then

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

- If X and Y are DRVs, then

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

Properties of variance and covariation

- For any RVs $X_1, \ldots, X_n$ holds

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

  In particular,

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

- If X and Y are independent RVs, then

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

$$\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] \tag{99}$$

1.2.6 Moment generating function

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

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

Other ways to compute MGF

- If X is a CRV, then

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

- If X is a DRV, then

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

Why the name?

Proposition.

$$\mu_k(X) = \frac{d^k}{dt^k} M_X(t) \bigg|_{t=0} \tag{103}$$

Proof. On the one hand

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

On the other hand, by Taylor series expansion

$$M_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \frac{d^k}{dt^k} M_X(t) \bigg|_{t=0} \tag{105}$$

And we get the desired result. □

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

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

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

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

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

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

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