〈1.1. Probability and its traits〉

Axioms of Probability

(1) (range of probability) : $P(A) \geq 0$

(2) (Probability of whole) : For sample space S, $P(S) = 1$

(3) Countable additivity : For events A_i and A_j s.t. $A_i \cap A_j = \emptyset$ $(i \neq j)$
$\Rightarrow P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$

[Theorem 1.1.1 Basic Properties of Probability]

(a) For each event A, $0 \leq P(A) \leq 1$ and $P(\emptyset) = 0$

(b) $P(A^c) = 1 - P(A)$

(c) If $A \subseteq B$, $P(A) \leq P(B)$

(Proof)
$A \cap A^c = \emptyset$ $\therefore$ (axiom (c)) $\Rightarrow P(A \cup A^c) = P(A) + P(A^c)$
we also know that $A \cup A^c = S$
$\therefore P(S) = P(A) + P(A^c) = 1 \Rightarrow P(A^c) = 1 - P(A)$ ··· (b)

$\because P(A), P(A^c) \geq 0$, it naturally follows from $P(A) + P(A^c) = 1$ that $0 \leq P(A) \leq 1$ and $0 \leq P(A^c) \leq 1$ ··· (a)-1

$P(\emptyset) = P(S^c) = 1 - P(S) = 1 - 1 = 0$ ··· (a)-2

(c) if $A \subseteq B$, $B = A \cup (B \cap A^c)$, $A \cap (B \cap A^c) = \emptyset$

$\therefore P(B) = P(A) + P(B \cap A^c) \geq P(A)$ □

[Textbook proof] use the following

$A \cap A^c = \emptyset$, $A \cup A^c = S$, $\emptyset = S^c$
If $A \subseteq B$, $A \cap (B - A) = \emptyset$, $A \cup (B - A) = B$

Example 1.1.2 Bernoulli Trial

A single random experiment with exactly two outcomes, usually "success" and "failure", where the probability of success (p) stays the same for every trial, and each trial's outcome is independent of the other.

① random experiment
② Two outcomes (success/failure)
③ p stays fixed for every trial.

〈Theorem 1.1.2 Probability of Union of Events〉

(a) $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$

(b) $P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$

(c) $P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - \cdots - P(A_{n-1} \cap A_n) + \cdots + (-1)^{n-1} P(A_1 \cap A_2 \cap \cdots \cap A_n)$

(d) (가산반가법성 ; Countable Subadditivity)
$P(A_1 \cup A_1 \cup \cdots) \leq P(A_1) + P(A_2) + \cdots$

(proof)
(a) $A_1 \cup A_2 = (A_1 \cap A_2^c) \cup (A_1 \cap A_2) \cup (A_1^c \cap A_2)$

we know that $(A_1 \cap A_2^c) \cap (A_1 \cap A_2) = \emptyset$
$(A_1 \cap A_2^c) \cap (A_1^c \cap A_2) = \emptyset$
$(A_1 \cap A_2) \cap (A_1^c \cap A_2) = \emptyset$

$\therefore P(A_1 \cup A_2) = P(A_1 \cap A_2^c) + P(A_1 \cap A_2) + P(A_1^c \cap A_2)$ (Axiom (c))
$= P(A_1 \cap A_2^c) + P(A_1 \cap A_2) + P(A_1 \cap A_2) - P(A_1 \cap A_2) + P(A_1^c \cap A_2)$
$= P(A_1) + P(A_2) - P(A_1 \cap A_2)$

* $P(A_1 \cap A_2^c) + P(A_1 \cap A_2) = P((A_1 \cap A_2^c) \cup (A_1 \cap A_2^c))$ $\because (A_1 \cap A_2^c) \cap (A_1 \cap A_2) = \emptyset$
$= P(A_1 \cap S)$
$= P(A_1)$

(b) and (c) can be proved by mathematical induction

$P(A_1 \cup A_2 \cup A_3) = P(A_1 \cup B)$, $B = (A_2 \cup A_3)$

$P(A_1 \cup B) = P(A_1) + P(B) - P(A_1 \cap B)$ ··· (1)
* $P(B) = P(A_2) + P(A_3) - P(A_2 \cap A_3)$
* $P(A_1 \cap B) = P(A_1 \cap (A_2 \cup A_3))$

$B = (A_2 \cup A_3) = (A_2 \cap A_3^c) \cup (A_2 \cap A_3) \cup (A_2^c \cap A_3)$ (mutually exclusive)
$\therefore P(A_1 \cap B) = P(A_1 \cap A_2 \cap A_3^c) + P(A_1 \cap A_2 \cap A_3) + P(A_1 \cap A_2^c \cap A_3)$
$= P[A_1 \cap A_2] + P[A_1 \cap A_3] - P(A_1 \cap A_2 \cap A_3)$

insert to (1) : $P(A_1) + \underbrace{P(A_2) + P(A_3) - P(A_2 \cap A_3)}_{P(B)} - \underbrace{[P(A_1 \cap A_2) + P(A_1 \cap A_3) - P(A_1 \cap A_2 \cap A_3)]}_{P(A_1 \cap B)}$

$= P(A_1) + P(A_2) + P(A_3) - P(A_2 \cap A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$ □

prove (c) later

〈Theorem 1.1.3 Continuity of Probability measure〉

(a) $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n \subseteq \cdots$ 이면 $P(\bigcup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} P(A_n)$

(b) $B_1 \supseteq B_2 \supseteq \cdots \supseteq B_n \supseteq \cdots$ 이면 $P(\bigcap_{n=1}^{\infty} B_n) = \lim_{n\to\infty} P(B_n)$

(proof)

$C_1 = A_1$, $C_2 = A_2 - A_1$, $\cdots$, $C_n = A_n - A_{n-1}$

$C_i \cap C_j = \emptyset \quad (i \neq j)$

$\sum_{i=1}^{n} C_j = A_n$

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P(C_1 \cup \cdots \cup C_n \cup \cdots) = \sum_{n=1}^{\infty} C_n = \lim_{n\to\infty} \sum_{k=1}^{n} C_k = \lim_{n\to\infty} A_n \quad \square$$

$$\lim_{n\to\infty} B_n = \bigcap_{n=1}^{\infty} B_n \Rightarrow P(\lim_{n\to\infty} B_n) = P(\bigcap_{n=1}^{\infty} B_n)$$

$$\Rightarrow \lim_{n\to\infty} P(B_n) = P(\bigcap_{n=1}^{\infty} B_n)$$

* Q: $P(\lim_{n\to\infty} B_n) = \lim_{n\to\infty} P(B_n)$?
what's the justification?

〈1.2 Conditional Probability and Independence〉

Definition: conditional probability of event B given event A

$$P(B|A) = \frac{P(A \cap B)}{P(B)} \quad (\text{단, } P(A) > 0)$$

〈Theorem 1.2.1 Properties of Conditional Probability〉

(a) [곱셈공식; multiplication formula) if $P(A) > 0$, $P(B) > 0$

$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$

(b) [전확률 공식; Total Probability Rule)

if $A_i \cap A_j = \emptyset \ (i \neq j)$, $A_1 \cup A_2 \cup \cdots = S$, and $P(A_i) > 0$

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots$$

(proof)

(a) (a) can be easily derived from the definition of conditional probability

(b) $B = (B \cap A_1) \cup (B \cap A_2) \cup \cdots$

$$P(B) = P((B \cap A_1) \cup (B \cap A_2) \cup \cdots) = P(B \cap A_1) + P(B \cap A_2) + \cdots = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots$$

$\because (B \cap A_1) \cap (B \cap A_2) = \emptyset$, $(B \cap A_i) \cap (B \cap A_j) = \emptyset \ (i \neq j)$

〈Ex. 1.2.1 Laplace's Probability of the Sun Rising〉

$N+1$ boxes labelled 0 to N.

k^{th} box contains k red balls and $N-k$ white balls.

Randomly pick a box and a ball inside it. Check the color and return it n times. They were all red.

* you're selecting the ball from the same box

What is the probability of choosing a red ball again and the probability of choosing a red ball in terms of N.

A_k: event of choosing the kth box

$P(A_k) = 1/(N+1)$

B: event of choosing a red ball n times from the kth box.

$P(B|A_k) = (k/N)^n$

$$P(B) = \sum_{k=0}^{N} P(B|A_k) \cdot P(A_k) = \sum_{k=0}^{N} \left(\frac{k}{N}\right)^n \frac{1}{N+1}$$

C: event of choosing a red ball for the $(n+1)^{th}$ time

$$P(B \cap C) = \sum_{k=0}^{N} P(B \cap C|A_k) \cdot P(A_k) = \sum_{k=0}^{N} \left(\frac{k}{N}\right)^{n+1} \frac{1}{N+1}$$

$$\therefore P(C|B) = \frac{P(C \cap B)}{P(B)} = \frac{\sum_{k=0}^{N} \left(\frac{k}{N}\right)^{n+1}}{\sum_{k=0}^{N} \left(\frac{k}{N}\right)^{n}}$$

If N is large, $\frac{1}{N} \sum_{k=0}^{N} \left(\frac{k}{N}\right)^m \simeq \int_0^1 x^m dx = \frac{1}{m+1}$

↳ Lower Darboux Sum; $L(f; P_N)$

P_N: $h_n = \frac{(b-a)}{n}$, $b=1$, $a=0$, $f(x) = x^m$

$$\therefore P(C|B) = \frac{\frac{1}{n+2}}{\frac{1}{n+1}} = \frac{n+1}{n+2} \quad \square$$

⟨ Theorem 1.2.2 Bayes Theorem ⟩

Let events $A_1, A_2, \cdots$ be the partition of the sample space S such that $A_i \cap A_j = \emptyset \ (i \neq j)$.

If $P(A_i) > 0$ and $P(B) > 0$,

$$P(A_j|B) \propto P(B|A_j)P(A_j) \quad (j=1,2,\cdots)$$

(proof)

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots}$$

$$C = \frac{1}{P(B)} \quad (??)$$

[Textbook]

If C is a proportionality factor, and add it to the proportional expression,

$$P(A_1|B) + P(A_2|B) + \cdots = C\{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots\}$$
$$P(A_1 \cup A_2 \cup \cdots | B) = C\{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots\}$$
$$1 = P(S|B) = C\{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots\}$$

$$C = 1/\{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots\}$$

$P(A_1), P(A_2), \cdots$: 여러 모형의 가능성
$P(A_1|B), P(A_2|B), \cdots$: 실험결과를 뜻하는 B의 관측 후에 각 모형의 가능성
$P(A_j)$: Prior probability
$P(A_j|B)$: Posterior probability

$$\text{Bayes Theorem}: P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

Ex. 1.2.2.

In a factory, there are 3 machines: M_1, M_2, M_3

$$\text{production} := \begin{cases} 20\% & (M_1) \\ 30\% & (M_2) \\ 50\% & (M_3) \end{cases}$$

$$\text{defect} := \begin{cases} 3\% & (M_1) \\ 2\% & (M_2) \\ 1\% & (M_3) \end{cases}$$

If a defective product is found, what are the respective probabilities that it has been produced from M_1, M_2 and M_3.

E: event that a defective product is made

$$\begin{aligned} P(E) &= P(E \cap M_1) + P(E \cap M_2) + P(E \cap M_3) \\ &= (0.2 \cdot 0.03) + (0.3 \cdot 0.02) + (0.5 \cdot 0.01) \\ &= 6 \cdot 10^{-3} + 6 \cdot 10^{-3} + 5 \cdot 10^{-3} \\ &= 17 \cdot 10^{-3} \\ &= 1.7 \times 10^{-2} \end{aligned}$$

$P(M_1|E) = P(M_1 \cap E)/P(E) = \frac{6}{17}$ ✓✓
$P(M_2|E) = 6/17$ ✓✓
$P(M_3|E) = 5/17$ ✓✓

Ex. 1.2.3

Imagine that you throw two dice

A = { First dice is even }
B = { Second dice is odd }
C = { the "sum" of two dice are odd }

(a) show that events (A,B) events (B,C) and events (C,A) are independent
(b) show that events C and $A \cap B$ are not independent

(Solution)

$A := \{2, 4, 6\}$
$B := \{1, 3, 5\}$
$C := \{\cancel{1}, 3, 5\}$ ↑ ∵ 1st dice + 2nd dice ≥ 2

$(A,B) := \{(1,2),(2,1),(1,4),(4,1),(2,3),(3,2)\}$

To show that two events A and B are independent,

① $P(A|B) = P(A)$
② $P(A \cap B) = P(A)P(B)$

$P(A \cap B) = 1/2 \cdot 1/2 = 1/4$
$P(A) \cdot P(B) = 1/2 \cdot 1/2 = 1/4$
$\therefore P(A \cap B) = P(A)P(B)$

$P(C) = 6/36 = 1/6$
$P(B \cap C) = \|\{(2,1),(4,1),(2,3)\}\|/36 = 1/12$
$P(B)P(C) = 1/2 \cdot 1/6 = 1/12$ $\therefore P(B \cap C) = P(B)P(C)$

$P(C \cap A) = \|\{(2,1),(4,1),(2,3)\}\|/36 = 1/12$
$P(C)P(A) = 1/6 \cdot 1/2 = 1/12$
$\therefore P(C \cap A) = P(C)P(A)$

(b) $P[C \cap A \cap B) = \|\{(2,1),(4,1),(2,3)\}\|/36 = 1/12$
$P(C)P(A \cap B) = 1/6 \cdot 1/2 \cdot 1/2 = 1/24$
$P(C \cap A \cap B) \neq P(C)P(A \cap B)$

[Textbook Solution]
$A \cap B \subset C \Rightarrow A \cap B \cap C = A \cap B$

$\therefore P(C|A \cap B) = P(A \cap B \cap C)/P(A \cap C)$
$= 1$

$\therefore P(C|A \cap B) \neq P(C)$

event $A \cap B$ affects event C hence they are not independent.

<Independence of Events>

(a) (two events are independent):
$P(A \cap B) = P(A)P(B)$

★ (b) (three events are independent):
$P(A \cap B) = P(A)P(B)$, $P(B \cap C) = P(B)P(C)$, $P(A \cap C) = P(A)P(C)$
$P(A \cap B \cap C) = P(A)P(B)P(C)$

★ (c) (n events are independent)
$P(A_i \cap A_j) = P(A_i)P(A_j)$ $[1 \le i \le j \le n)$
$P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k)$ $[1 \le i \le j \le k \le n)$

$(\cdots)$

$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2)\cdots P(A_n)$

Ex. 1.2.4.

You are throwing two dice. Show that the following three events are independent.

A = { 1st dice is even }
B = { 2nd dice is odd }
C = { Sum of 1st and 2nd ≤ 4, difference of 1st, 2nd ≤ 1 }

$A := \{2, 4, 6\}$ $B := \{1, 3, 5\}$

$C := \{(1,1),(1,2),(2,1),(2,2)\}$

$P(A) = 1/2$, $P(B) = 1/2$, $P(C) = 4/36 = 1/9$
$P(A \cap C) = 2/36 = 1/18$, $P(B \cap C) = 1/18$, $P(A \cap B \cap C) = 1/36$
$P(A)P(B)P(C) = 1/36 = P(A \cap B \cap C)$

$\therefore A, B, C$ are independent.

Ex. 1.2.5 Show that if A and B are independent, A and B^c are also independent.

WTS: $P(A \cap B^c) = P(A)P(B^c)$

$* A = (A \cap B) \cup (A \cap B^c)$

$P(A) = P(A \cap B) + P(A \cap B^c)$ $\quad (A \cap B) \cap (A \cap B^c) = \emptyset$

$P(A) = P(A)P(B) + P(A \cap B^c)$ $\quad \therefore P(A) = P(A \cap B) + P(A \cap B^c)$

$P(A) - P(A)P(B) = P(A \cap B^c)$

$P(A)[1 - P(B)] = P(A \cap B^c)$

$P(A)P(B^c) = P(A \cap B^c)$

$\therefore$ A and B^c are independent.

ShinKook's definition of a random variable:

A random variable is a real-valued function whose domain is the set of all possible outcomes in a sample space, and whose range is a subset of the real numbers ($\mathbb{R}$). It is classified as discrete or continuous based on the nature of its range.

Understand the difference between "outcome" and "events".

〈1.3 Random Variables and Probability Distribution〉

〈개념정리〉

Random Experiment:

1. 여러가지 결과 가능
2. 각 결과의 가능성을 확률로 나타낼 수 있음

Random Variable:

표본공간 (Sample Space) 에서 정의된 실수 값 함수

Discrete type random variable:

: 확률변수가 가질 수 있는 값들의 집합을 $\{x_1, x_2, x_3, \cdots\}$ 와 같이 나타낼수 있다 (유한개의 집합)

→ Range of r.v.s are finite

Continuous type random variables:

: 확률변수가 실수 구간의 값들을 가질 수 있고, 그에 관한 확률이 적분으로 주어질 때.

→ Range of r.v.s are infinite

→ probability of r.v.s are given by integral of pdf over $[a,b]$

Ex. 1.3.5

Find the value of c so that f(x) is a pdf, and find $P(1/2 \le X \le 3/4)$ when $X \sim f$

$$f(x) := \begin{cases} cx(1-x) & 0 \le x \le 1 \\ 0 & x < 0,\ x > 1 \end{cases}$$

$$\int_0^1 f(x) = 1 \quad \therefore \int_0^1 cx(1-x)\,dx = \int_0^1 cx - cx^2\,dx = 1$$

$$\left[\frac{1}{2}cx^2 - \frac{1}{3}cx^3\right]_0^1 = \frac{1}{2}c - \frac{1}{3}c = \frac{1}{6}c = 1 \quad \therefore c = 6$$

$$\int_{1/2}^{3/4} f(x)\,dx = \int_{1/2}^{3/4} 6x - 6x^2\,dx = \left[3x^2 - 2x^3\right]_{1/2}^{3/4}$$

$$= \left(3\cdot\frac{9}{16} - 2\cdot\frac{27}{64}\right) - \left(3\cdot\frac{1}{4} - 2\cdot\frac{1}{8}\right)$$

$$= \frac{1}{64}(12 \cdot 9 - 27) - \frac{1}{8}(6 - 2)$$

$$= \frac{1}{64}(108 - 54) - \frac{4}{8}$$

$$= \frac{54}{64} - \frac{1}{2} = \frac{54 - 32}{64} = \frac{22}{64} = \frac{11}{32}$$

* pdf value $f(a) \neq P(X = a)$ and

if $\Delta x \fallingdotseq 0$, $P(a \le X \le a + \Delta x) \fallingdotseq f(a)\Delta x$

* $\fallingdotseq$: approximately equal to.

〈1.4 Properties of Probability Distribution〉

Definition (mean)

$$\mu = \begin{cases} \sum_x x \cdot f(x) & X \text{ is discrete} \\ \int x \cdot f(x) & X \text{ is Continuous} \end{cases}$$

Ex 1.4.1

Find the mean μ of f

$$f(x) := \begin{cases} (-x^2+2x)/2, & 0 \le x \le 1 \\ (-x^2+2x+3)/8, & 1 \le x \le 3 \\ 0 & x<0,\ x>3 \end{cases}$$

$$\mu = \int_0^1 x\cdot(-x^2+2x)/2\,dx + \int_1^3 x\cdot(-x^2+2x+3)/8$$

$$= \frac{1}{2}\int_0^1 -x^3+2x^2\,dx + \frac{1}{8}\int_1^3 -x^3+2x^2+3x\,dx$$

$$= \frac{1}{2}\left[-\frac{1}{4}x^4+\frac{2}{3}x^3\right]_0^1 + \frac{1}{8}\left[-\frac{1}{4}x^4+\frac{2}{3}x^3+\frac{3}{2}x^2\right]_1^3$$

$$= \frac{1}{2}\left[-\frac{1}{4}+\frac{2}{3}\right] + \frac{1}{8}\left[-\frac{81}{4}+\frac{54}{3}+\frac{27}{2}\right] - \frac{1}{8}\left[-\frac{1}{4}+\frac{2}{3}+\frac{3}{2}\right]$$

$$= -\frac{1}{8}+\frac{1}{3}-\frac{81}{32}+\frac{54}{24}+\frac{27}{16}+\frac{1}{32}-\frac{2}{24}-\frac{3}{16} = 1.375 = \frac{11}{8}$$ ✓✓

Definition (Expectation)

If random variable $X \sim f$, for real value function $g(x)$,

$$E[X] := \begin{cases} \sum_x g(x)f(x) & X \text{ is discrete} \\ \int_{-\infty}^{+\infty} g(x)f(x)\,dx & X \text{ is Continuous.} \end{cases}$$

Ex. 1.4.2.

Imagine flipping two separate coins. Let X be the number of heads. If you could get \$$X^2$, what is the expected value?

$P(X=0) = 1/4 \quad P(X=1) = 1/2 \quad P(X=2) = 1/4$

$$\therefore E[X^2] = \sum_{x=0}^{2} x^2 \cdot P(X=x)$$

$$= 0^2\cdot 1/4 + 1^2\cdot 1/2 + 2^2\cdot 1/4 = 3/2$$

Definition (Variance and standard deviation)

If the r.v. $X \sim f$ and $E[X] = \mu$

$$Var(X) = E[(X-\mu)^2] = \begin{cases} \sum_x (x-\mu)^2 f(x) & X \text{ is discrete} \\ \int_{-\infty}^{+\infty} (x-\mu)^2 \cdot f(x) & X \text{ is continuous} \end{cases}$$

$$Sd = \sqrt{Var(X)}$$

* mean and expectation must be absolutely convergent in order for it to be defined.

p. 26 평균: "여기에서 무한 합이나 이상 적분이 절대 수렴하는 경우에만 평균이 실수로 정의되는 것에 유의하여야 한다"

p. 27. 실수로 정의: "여기에서 무한 합이나 이상적분이 실수로 정의되려면 합이나 적분이 절대수렴하여야 한다"

[Absolutely Convergent]

9.1.1 Definition Let $X := (x_n)$ be a sequence in $\mathbb{R}$. We say that the series $\sum x_n$ is **absolutely convergent** if the series $\sum |x_n|$ is convergent in $\mathbb{R}$. A series is said to be **conditionally** (or **nonabsolutely**) **convergent** if it is convergent, but it is not absolutely convergent.

Ex 1.4.3

Find the variance of the following pdfs

(a) $X \sim f_1$ (pdf) (uniform dist.) (b) $X \sim f_2(x)$ (triangle dist.)

$f_1(x) = I_{[0,1]}(x)$ $\qquad f_2(x) = (2-4|x-1/2|)\,I_{[0,1]}(x)$

(a) $\mu = \int_0^1 x\cdot f(x)\,dx = \int_0^1 x\,dx = \left[\frac{1}{2}x^2\right]_0^1 = 1/2$

$$Var(X) = E[(X-\mu)^2] = E[(X-1/2)^2]$$

$$= \int_0^1 (x-1/2)^2\cdot f(x)\,dx$$

$$= \int_0^1 x^2 - x + 1/4\,dx$$

$$= \left[\frac{1}{3}x^3 - \frac{1}{2}x^2 + \frac{1}{4}x\right]_0^1$$

$$= \frac{1}{3}-\frac{1}{2}+\frac{1}{4} = \frac{4-6+3}{12} = \frac{1}{12}$$

(b) $f(x) := \begin{cases} 2-4|x-1/2| & \text{if } 1/2 \le x \le 1 \\ 2-4(1/2-x) & \text{if } 0 \le x \le 1/2 \end{cases} \Rightarrow \begin{matrix} 4-4x & \text{if } 1/2 \le x \le 1 \\ 4x & \text{if } 0 \le x \le 1/2 \end{matrix}$

$$\mu = \int_0^1 x\cdot f(x)\,dx = \int_0^{1/2} x\cdot f(x)\,dx + \int_{1/2}^1 x\cdot f(x)\,dx$$

$$= \int_0^{1/2} 4x^2\,dx + \int_{1/2}^1 4x-4x^2\,dx$$

$$= \left[\frac{4}{3}x^3\right]_0^{1/2} + \left[2x^2-\frac{4}{3}x^3\right]_{1/2}^1$$

$$= \frac{4}{3}\cdot\frac{1}{8} + \left[2-\frac{4}{3}\right) - \left(\frac{1}{2}-\frac{4}{24}\right)$$

$$= 4/24 + 2/3 - 8/24 = 1/6 + 1/3 = 3/6 = 1/2$$

$$\mathrm{Var}(X) = E[(X-\mu)^2] = \int_0^1 (x-\mu)^2 f(x)\,dx$$

$$= \int_0^{1/2} (x-1/2)^2 \cdot 4x\,dx + \int_{1/2}^1 (x-1/2)^2 (4-4x)\,dx$$

$$= \int_0^{1/2} (x^2-x+1/4)\cdot 4x\,dx + \int_{1/2}^1 (x^2-x+1/4)(4-4x)\,dx$$

$$= \int_0^{1/2} 4x^3-4x^2+x\,dx + \int_{1/2}^1 4x^2-4x+1-4x^3+4x^2-x\,dx$$

$$= \left[x^4-\frac{4}{3}x^3+\frac{1}{2}x^2\right]_0^{1/2} + \int_{1/2}^1 -4x^3+8x^2-5x+1\,dx$$

$$= \left[\frac{1}{16}-\frac{4}{3}\cdot\frac{1}{8}+\frac{1}{2}\cdot\frac{1}{4}\right] + \left[-x^4+\frac{8}{3}x^3-\frac{5}{2}x^2+x\right]_{1/2}^1$$

$$= \left[\frac{1}{16}-\frac{1}{6}+\frac{1}{8}\right] + \left[-1+\frac{8}{3}-\frac{5}{2}+1\right] - \left[-\frac{1}{16}+\frac{1}{3}-\frac{5}{8}+\frac{1}{2}\right]$$

$$= \frac{1}{24}$$ ✓✓

[Textbook Solution]

σ_2^2 : Variance of (b)

$$\sigma_2^2 = \int_0^1 (x-1/2)^2 (2-4|x-1/2|)\,dx$$

$$= \int_{-1/2}^{1/2} t^2(2-4|t|)\,dt$$

$t = x-1/2 \quad dx = dt$

$$= 2\int_0^{1/2} t^2(2-4t)\,dt$$

$$= 2\int_0^{1/2} 2t^2-4t^3\,dt = 2\left[\frac{2}{3}t^3-t^4\right]_0^{1/2} = 2\left(\frac{2}{24}-\frac{1}{16}\right)$$

$$= 2\left(\frac{1}{12}-\frac{1}{16}\right)$$

$$= 2\left(\frac{16-12}{192}\right) = \frac{8}{192} = \frac{1}{24}$$ □

〈Theorem 1.4.1〉 Properties of Expectation

(linearity) $E[aX+b] = aE[X]+b$

(linearity) $E[c_1 g_1(X)+c_2 g_2(X)] = c_1 E[g_1(X)] + c_2 E[g_2(X)]$

(uniformity) $g(X_1) \le g(X_2) \Rightarrow E[g(X_1)] \le E[g(X_2)]$

〈Theorem 1.4.2〉 Properties of Variance and ways to calculate it.

(a) $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$ (a, b are constants)

(b) $\mathrm{Var}(X) = E[X^2] - E[X]^2$

(proof)

(a) $\mathrm{Var}(aX+b) = E[(aX+b-\mu^*)^2]$, $\mu^* = E[aX+b] = a\mu+b$

$= E[(aX+b-a\mu-b)^2]$

$= E[a^2(X-\mu)^2]$

$= a^2 E[(X-\mu)^2]$

$= a^2\mathrm{Var}(X)$

(b) $\mathrm{Var}(X) = E[(X-E[X])^2]$

$= E[X^2 - 2XE[X] + E[X]^2]$

$= E[X^2] - 2E[X]E[X] + E[X]^2$

$= E[X^2] - E[X]^2$

Ex 1.4.4 (continuation of Ex. 1.4.3)

(a) For Ex. 1.4.3 (a)

$\mu = E[X] = 1/2$, $E[X^2] = \int_0^1 x^2\,dx = 1/3$

$\therefore \mathrm{Var}(X) = E[X^2] - E[X]^2 = 1/3 - (1/2)^2 = 1/12$

(b) For Ex. 1.4.3 (b)

$\mu = E[X] = 1/2$, $E[X^2] = \int_0^{1/2} x^2(4x)\,dx + \int_{1/2}^1 x^2(4-4x)\,dx = 7/24$

$\therefore \mathrm{Var}(X) = E[X^2] - \mu^2 = 7/24 - (1/2)^2 = 1/24$

Thm 1.4.1, 1.4.2 footnote

: 8) 이러한 선형성과 단조성은 각각의 기대값이 실수로 정의되는 것을 전제로 하는 것이다.

9) 이러한 성질과 공식은 각 변수의 분산과 기대값이 모두 실수라는 전제 하에 성립하는 것이다.

Ex 1.4.5 Standardization of random variable

mean = μ, Standard deviation = σ

$$Z = \frac{X-E[X]}{\sqrt{\mathrm{Var}(X)}} = \frac{X-\mu}{\sigma}$$

$E[Z] = (E[X]-\mu)/\sigma = 0$

$\mathrm{Var}(Z) = \mathrm{Var}(X)/\sigma^2 = 1$

"Z is the standardized random variable of X"

한편, 평균이나 분산이 실수로 정의될 수 없는 확률분포가 있음에 유의하여야 한다. 예를 들어, 다음의 확률밀도함수를 갖는 확률변수에 대하여는 평균이나 분산을 정의할 수 없다.

$$f(k) = P(X=k) = \frac{1}{k(k+1)}, \quad k = 1, 2, \cdots$$

$$E[X] = \sum_{x=1}^{\infty} x \cdot \frac{1}{x(x+1)} = \sum_{x=1}^{\infty} \frac{1}{x+1} \quad : \text{infinitely increasing sequence}$$

Another Example: Cauchy Distribution

$$f(x) := \frac{1}{\pi\gamma\left[1+\left(\frac{x-x_0}{\gamma}\right)^2\right]}, \quad x_0: \text{location parameter}, \quad \gamma: \text{scale parameter}$$

〈1.5 Cumulative Distribution Functions & Generating Functions (생성함수)〉

Definition. Cumulative Distribution Function of discrete random variables

$$f(x_k) = P(X = x_k) \ (k=1,2,3\cdots), \quad f(x) = 0 \quad \forall x : x \neq x_k \ (k=1,2,\cdots)$$

$$F(x) = P(X \le x) = \sum_{t:\, t \le x} f(t)$$

Also, if $x_1 < x_2 < \cdots < x_{n-1} < x_n < \cdots$

$$F(x_n) = \sum_{k=1}^{n} f(x_k), \quad f(x_n) = F(x_n) - F(x_{n-1})$$

Ex 1.5.1 Flipping the coin until you get heads

From 〈Ex. 1.3.3〉 we know that the pdf $f(k)$ is

$$f(k) = P(X=k) = (1/2)^k \quad (k=1,2,\cdots)$$
$$f(x) = 0 \quad \forall x \neq k \quad (k=1,2,\cdots)$$

cdf $F(x)$ is given as

$$F(n) = P(X \le n) = \sum_{k=1}^{n} f(k) = \sum_{k=1}^{n} (1/2)^k = 1 - (1/2)^n \quad (n=1,2,3,\cdots)$$

$$F(x) = \sum_{t:\, t \le x} f(t) = \sum_{k=1 \atop k \le x}^{\infty} f(k) = \begin{cases} 0 & x \le 1 \\ 1-(1/2)^n, & n \le x < n+1 \ (n=1,2,\cdots) \end{cases}$$

Definition. Cumulative Distribution Function of continuous random variable

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\, dt$$

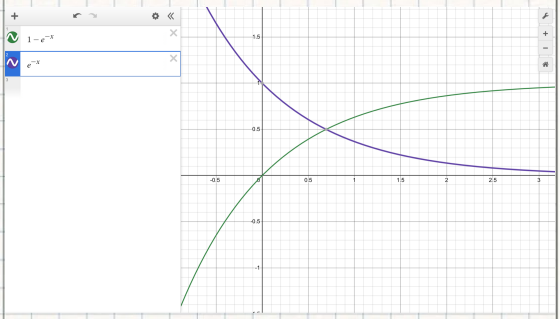
For continuous x, $\frac{d}{dt} F(x) = f(x)$

Ex. 1.5.2 Standard Exponential Distribution (표준지수분포)

$$\text{pdf}: f(x) = e^{-x} I_{[x \ge 0]}(x)$$

$$F(x) = \int_0^x e^{-t}\, dt = \left[-e^{-t}\right]_0^x = 1 - e^{-x}$$

$$F(x) := \begin{cases} 0 & x < 0 \\ \int_0^x e^{-t} dt = 1 - e^{-x} & x \ge 0 \end{cases}$$



〈Theorem 1.5.1 : Properties of Cumulative Distribution Function〉

(a) (increasing) if $x_1 < x_2$, $F(x_1) \le F(x_2)$

(b) (전체 변동) $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to +\infty} F(x) = 1$

☆(c) (오른쪽 연속성) $\lim_{\substack{h \to 0 \\ h > 0}} F(x+h) = F(x)$, $\lim_{x \to c^+} F(x) = F(c)$

(proof)

(a) if $x_1 < x_2$, $(X \le x_1) \subseteq (X \le x_2)$

using Theorem 1.1.1 (c) 확률의 단조성, it is clear that

$P(X \le x_1) \le P(X \le x_2) \Rightarrow F(x_1) \le F(x_2)$

(b)

〈Theorem 1.1.3 Continuity of Probability measure〉

(a) $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n \subseteq \cdots$ 이면 $P(\bigcup_{n=1}^{\infty} A_n) = \lim_{n \to \infty} P(A_n)$

(b) $B_1 \supseteq B_2 \supseteq \cdots \supseteq B_n \supseteq \cdots$ 이면 $P(\bigcap_{n=1}^{\infty} B_n) = \lim_{n \to \infty} P(B_n)$

$$\lim_{x \to -\infty} F(x) = \lim_{x \to -\infty} P(X \le x) = \lim_{x \to \infty} P(X \le -x) = P\left(\bigcap_{x=1}^{\infty} (X \le -x)\right)$$

$$= P(\emptyset) = 0$$

$$\lim_{x \to \infty} F(x) = \lim_{x \to \infty} P(X \le x) = P\left(\bigcup_{x=1}^{\infty} (X \le x)\right) = P(S) = 1$$

(c) 포함관계가 작아지는 사건들 $(X \le x + 1/n)$ $(n = 1, 2, \cdots)$ 에 대하여

$$\lim_{n \to +\infty} (X \le x + 1/n) = \bigcap_{n=1}^{\infty} (X \le x + 1/n) = (X \le x)$$

이므로 확률측도의 연속성과 (a)의 증가성으로부터 오른쪽 연속성이 성립한다.

누적분포함수가 오른쪽 연속임을 나타내기 위하여, 오른쪽 극한을

$$F(a+) = \lim_{\substack{x \to a \\ x > a}} F(x) = \lim_{n \to +\infty} F(a + 1/n)$$

로 나타내어 오른쪽 연속성을 다음과 같이 나타내기도 한다.

$$F(a+) = F(a) = P(X \le a)$$

Q) Why can't cdf be always left continuous?

Suppose that $F(x)$ is left continuous.

$\lim_{n \to \infty} F(a - 1/n) = F(a)$

$$\text{LHS} := \lim_{n \to \infty} F(a - 1/n) = \lim_{n \to \infty} P(X \le a - 1/n) = P\left(\bigcup_{n=1}^{\infty} (X \le a - 1/n)\right)$$

$$= P(X < a) \quad \left[* \text{Note that } \bigcup_{n=1}^{\infty} (X \le a - 1/n) \text{ is exclusive of } a \right]$$

$$\text{RHS} := F(a) = P(X \le a)$$

$\therefore$ LHS $\neq$ RHS

* If we change the definition of $F(x) = P(X < x)$ then $F(x)$ is left continuous and not right continuous.

☆ $* \; F(a) - F(a-) = P(X \le a) - P(X < a) = P(X = a)$

- 누적분포함수의 왼쪽 방향에서의 불연속성의 크기에 대하여 위의 식이 성립한다

* Note:

If $F(x)$ is continuous, $\lim_{x \to c^-} F(x) = \lim_{x \to c^+} F(x) = F(c)$

$\therefore P(X = c) = P(X \le c) - P(X < c) = \lim_{x \to c^+} F(x) - \lim_{x \to c^-} F(x) = 0$

If X is discrete, $P(X \le c) \neq P(X < c)$

$\therefore P(X = c) = P(X \le c) - P(X < c) \neq 0$

A different perspective

Ex. 1.5.3

- Shuttle bus between a hospital and a subway (every 10 minutes)
- hospital ↔ bus : 10 min : bus, 15 min : walk
- A person flips a coin and if head, take bus, if tail, walk.
- Find the cdf of time taken for the person to reach the hospital.

X : minutes taken to arrive

$$P(X \le x \mid \text{Head}) := \begin{cases} 0 & x < 10 \\ (x-10)/10 & 10 \le x \le 20 \\ 1 & x > 20 \end{cases}$$

$$P(X \le x \mid \text{tail}) := \begin{cases} 0 & x < 15 \\ 1 & x \ge 15 \end{cases}$$

$$P(X \le x) = P(X \le x \mid \text{Head})\,P(\text{Head}) + P(X \le x \mid \text{Tail})\,P(\text{Tail})$$

$$:= \begin{cases} 0 & x < 10 \\ (x-10)/20 & 10 \le x < 15 \\ x/20 & 15 \le x \le 20 \\ 1 & x > 20 \end{cases}$$

Definition. Probability Generating Function

음이 아닌 정수의 값을 가질 수 있는 이산형 확률변수 X의 확률생성함수*는

$$G(s) = E[s^X] = \sum_{x=0}^{\infty} s^x P(X = x), \quad -1 < s < 1$$

이고, 그 확률밀도함수는

$$f(k) = P(X = k) = G^{(k)}(0)/k!, \quad k = 0, 1, 2, \cdots$$

Ex. 1.5.4 앞면이 나올 때까지 동전을 던지는 실험

〈Ex 1.3.3〉에서의 앞면이 나올 때까지 동전을 던지게 될 횟수 X의 확률밀도함수는

$$f(k) = P(X = k) = (1/2)^k \quad (k = 1, 2, \cdots), \quad f(x) = 0 \;\; \forall x \neq k \;\; (k = 1, 2, \cdots)$$

$$G(s) = E[s^X] = \sum_{x=1} s^x (1/2)^x = \sum_{x=1}^{\infty} (s/2)^x = \sum_{x=0}^{\infty} (s/2)^x - 1$$

$$\sum_{x=0}^{n} ar^n = \frac{a(r^n - 1)}{r - 1}, \quad = \frac{1 \cdot (0 - 1)}{s/2 - 1} - 1 = \frac{2}{2 - s} - 1 = \frac{s}{2 - s}, \quad -2 < s < 2$$

* pgf footnote: 확률생성함수를 뜻하는 영문은 〈예 1.5.4〉에서와 같이 1보다 큰 s 값에 대해서도 절대 수렴할 수 있다.

Definition: Moment Generating Function

For all values of t in an open interval containing 0, if $E[e^{tX}]$ a real value, the function $M(t) := E[e^{tX}], -h < 0 < h \ (\exists h > 0)$ is called the moment generating function of the random variable X

Ex. 1.5.5 앞면이 나올 때까지 동전을 던지는 실험

Refer to 〈Ex. 1.3.3〉

$$f(k) = P(X=k) = (1/2)^k \quad (k = 1, 2, \cdots)$$

$$E[e^{tX}] = \sum_{k=1}^{\infty} e^{tk} \cdot (\tfrac{1}{2})^k = \sum_{k=1}^{\infty} (\tfrac{e^t}{2})^k = \frac{e^t}{2 - e^t} \quad \text{(Ex. 1.5.4 참고)}$$

range of t: $(e^t/2) < 1$ (for convergence. otherwise $E[e^{tX}] \to +\infty$)

$\log e^t - \log 2 < \log 1$

$t < \log 2$

$\therefore E[e^{tX}] = e^t(2 - e^t), \quad t < \log 2$

* Footnote: $E[e^{tX}] < +\infty \ \forall t: -h < t < h \ (h > 0)$ 가 성립하지 않으면 적률생성함수가 존재하지 않는다고 한다.

Definition: Moment

Given that $E(|X|^k) < +\infty$,

$$m_k = E[X^k] = \begin{cases} \sum_x x^k f(x) & (X \text{ is discrete}) \\ \int_{-\infty}^{+\infty} x^k \cdot f(x)\,dx & (X \text{ is continuous}) \end{cases}$$

m_k is called the kth moment of the random variable X. $(k = 1, 2, \cdots)$

〈Theorem 1.5.2 Properties of Moment Generating Functions〉

(a) (moment generating property)
If the mgf of the random variable X exists, meaning

$$M(t) = E[e^{tX}] < +\infty \quad \forall t: -h < t < h \ (\exists h > 0)$$

all moments of X exists and

$$E[X^k] = M^{(k)}(0), \quad M(t) = \sum_{k=0}^{\infty} \frac{E[X^k]}{k!} t^k, \quad -\epsilon < t < \epsilon \ (\exists \epsilon > 0)$$

(b) (determines the distribution)[14]. If two random variables X and Y's mgfs, $M_X(t)$ and $M_Y(t)$ exists and are equal on an open interval containing 0, meaning

$\therefore M_Y(t) = M_X(t) \quad \forall t: -h < t < h \ (h > 0)$

The distribution of X and Y are equal and so is their pdfs and cdfs

(proof)
The power series expansion [멱급수 전개식] of e^a:

$$e^a := 1 + \frac{a}{1!} + \frac{a^2}{2!} + \cdots + \frac{a^k}{k!} + \cdots$$

From this expansion, we can derive the following inequality

$$|a|^k / k! \le e^{|a|} \le e^a + e^{-a}$$

let $a = tx \Rightarrow |tx|^k / k! \le e^{|tx|} \le e^{tx} + e^{-tx}$

Apply $E[\cdot]$ onto both sides of the inequality.

$$E[|tx|^k / k!] \le E[e^{tx}] + E[e^{-tx}]$$

$$|t|^k E[|X|^k]/k! \le M(t) + M(-t) < +\infty \quad (-h < t < h)$$

$\therefore$ All moments of X exists.

* Footnote:
13) 적률이 실수로 정의되기 위하여는 무한합이나 이상적분이 절대 수렴, 즉 $E(|X|^k) < +\infty$ 이어야 한다

14) 적률생성함수를 일반화한 특성함수(characteristic function)를 이용한 증명은 참고서적 [p. Billingsley] Probability and Measure 3rd ed. p.346 ~ p.347을 참조하시오.

부록I p484 테일러 정리 (Taylor's theorem)

(a) 함수 f가 열린 구간 $(a-h, a+h) \ (h > 0)$에서 n번 미분 가능하고 n가 도함수가 연속함수이면, 그 구간 내의 모든 x에 대하여 다음 등식이 성립한다.

$$f(x) = f(a) + \sum_{r=1}^{n-1} \frac{f^{(r)}(a)}{r!}(x-a)^r + R_n(x,a)$$

$$\star \quad R_n(x,a) = \frac{1}{(n-1)!}\int_0^1 (1-u)^{n-1} f^{(n)}(a + u(x-a))\,du\,(x-a)^n$$

$$= \frac{f^{(n)}(\xi)}{n!}(x-a)^n, \ \exists \xi: |\xi - a| \le |x-a|$$

(b) 함수 f가 열린구간 $(a-h, a+h) \ (h > 0)$에서 n번 미분가능하면, 그 구간 내의 모든 x에 대하여 다음이 성립한다.

$$f(x) = f(a) + \sum_{r=1}^{n} \frac{f^{(r)}(a)}{r!}(x-a)^r + R_{n+1}(x,a), \ \lim_{x \to a} R_{n+1}(x,a)/(x-a)^n = 0$$

(… 이어서) 부록의 테일러 정리를 지수함수에 적용하면

$$e^{tx} - \sum_{k=0}^{n-1} \frac{(tx)^k}{k!} = \frac{1}{(n-1)!}\int_0^1 (1-u)^{n-1} e^{u(tx)}\,du\,(tx)^n = R_n(x, a=0)$$

$$\left| e^{tx} - \sum_{k=0}^{n-1} \frac{(tx)^k}{k!} \right| \le \int_0^1 (1-u)^{n-1} \cdot e^{u|tx|}\,du \cdot \frac{|tx|^{n-1}}{(n-1)!} \cdot \frac{|tx|}{1!}$$

$$\le \int_0^1 (1-u)^{n-1} \cdot e^{|tx|}\,du \cdot e^{|tx|} \cdot e^{|tx|}$$

$$\le \frac{1}{n}\left(e^{3tx} + e^{-3tx}\right)$$

이 부등식의 각 변에 확률밀도함수를 곱하여 합하거나 적분하면

$$\left| M(t) - \sum_{k=0}^{n-1} E[X^k] t^k / k! \right| \le \frac{1}{n} \left\{ M(\beta t) + M(-\beta t) \right\}, \quad -h < \beta t < h$$

as $n \to \infty$

$$M(t) = \sum_{k=0}^{\infty} \frac{E[X^k]}{k!} t^k, \quad -h < \beta t < h \ \left[\Rightarrow -h^* < t < h^*, \ h^* = h/\beta \right]$$

(b) 이 증명은 책의 수준을 넘기므로 생략하기로 한다.

Ex. 1.5.6. Standard exponential distribution's mgf and moments.

$f(x) = e^{-x}, \quad x \in [0, +\infty)$

$$M(t) = E[e^{tX}] = \int_0^{\infty} e^{tx} \cdot e^{-x} dx = \int_0^{\infty} e^{(1-t)\cdot(-x)} dx = \left[\frac{1}{1-t} \cdot (-1) \cdot e^{-x} \right]_0^{\infty}$$

$$= (1-t)^{-1}, \quad t \neq 1 \ \therefore |t| < 1$$

$M'(t) = (-1)\cdot(-1)\cdot(1-t)^{-2} \quad \therefore M'(t=0) = 1$

$M^{(2)}(t) = (-1)\cdot(-1)\cdot(-2)\cdot(-1)\cdot(1-t)^{-3} \quad \therefore M^{(2)}(t=0) = 2$

$\vdots$

$M^{(k)}(t) = k!(1-t)^{-(k+1)} \quad \therefore M^{(k)}(t=0) = k!$

Power series expansion of $M(t)$

$$M(t) = M(t=0) + \frac{M'(t=0)}{1!}(t-0) + \frac{M''(t=0)}{2!}(t-0)^2 + \cdots + \frac{M^{(k)}(t=0)}{k!}(t-0)^k + \cdots$$

$$= \sum_{k=0}^{\infty} t^k, \quad |t| < 1$$

X's k^{th} moment $M^{(k)}(t=0) = E[X^k] = k!$

$$* \int_0^{\infty} x^k e^{-x} dx = \int_0^{\infty} x^k d(-e^{-x}) = \int_0^{\infty} e^{-x} d(x^k) = k \int_0^{\infty} x^{k-1} e^{-x} dx$$

$$= \cdots = k!$$

$$\Gamma(k) = (k-1)! = \int_0^{\infty} x^{k-1} e^{-x} dx$$

Ex. 1.5.7.

If it is known that the mgf of the random variable X exists and its moment $E[X^k] = 2^k k!$ $(k = 0, 1, 2, \cdots)$

by Theorem 1.5.2(a), the mgf of X is:

$$M(t) = \sum_{k=0}^{\infty} \frac{E[X^k]}{k!} t^k = \sum_{k=0}^{\infty} (2t)^k = \frac{1}{1-2t}, \quad |2t| < 1$$

The pdf is: $f(x) := \frac{1}{2} e^{-\frac{1}{2}x} I_{[0,+\infty)}(x)$

$$\because \int_{-\infty}^{+\infty} e^{tx} \cdot \frac{1}{2} e^{-\frac{1}{2}x} I_{[0,+\infty)}(x) dx = \frac{1}{1-2t}, \quad |t| < 1/2$$

"적률생성함수의 분포결정성으로부터 이 함수 $f(x)$가 X의 확률밀도함수이다."[16]

* Footnote: 적률생성함수에 대응하는 확률생성함수를 구하는 일반적인 과정은 라플라스변환의 역변환과정으로서 이 책의 수준에서는 다룰수 없고, 이 책에서는 많은 확률밀도함수에 대응하는 적률생성함수들을 기억하여 역으로 대응관계를 이용하도록 한다.

Definition: Cumulant Generating Function

$$C(t) = \log M(t) = \log E[e^{tX}], \quad -h < t < h \ (h > 0)$$

Power series of $C(t)$: $\sum_{r=0}^{\infty} \frac{C^{(r)}(0)}{r!} t^r, \quad -h < t < h \ (\exists h > 0), \ r = 1, 2, \cdots$

k^{th} cumulant: $C^{(k)}(0) = C_r(X)$ or C_r (symbols)

- Relationship between C_r and m_k

$$\therefore \log\left(\underbrace{1}_{\frac{m_0}{0!}t^0} + \sum_{k=1}^{\infty} \frac{m_k}{k!} t^k \right) = \sum_{r=1}^{\infty} \frac{C_r}{r!} t^r$$

* 로그함수의 멱급수 전개식. logarithmic function's power series expansion

- $\log(1+A) = A - A^2/2 + A^3/3 - A^4/4 + \cdots \quad (-1 < A < 1)$

  $\log(1-A) = A + A^2/2 + A^3/3 + A^4/4 + \cdots \quad (-1 < A < 1)$

  $A = m_1 t + m_2 t^2/2! + m_3 t^3/3! + \cdots$

로그함수의 멱급수 전개를 이용하고 t의 오름차순으로 정리하면

$C_1 = m_1 \quad C_2 = 2!\left\{ m_2/2! - (m_1)^2/2! \right\} = m_2 - (m_1)^2, \cdots$

1st cumulant: mean, 2nd cumulant: variance.

Ex. 1.5.8 Standard exponential distribution's cgf and cumulant

$M(t) = (1-t)^{-1}$, $C(t) = \log M(t) = -\log(1-t) = t + t^2/2 + t^3/3 + \cdots, \ |t| < 1$

$C'(t) = (1-t)^{-1} \quad C'(0) = 1$

$C''(t) = (1-t)^{-2} \quad C''(0) = 1$

$C_r = (r-1)!$

Ex 1.5.9.

Refer to Ex. 1.5.7. $M(t) = (1-2t)^{-1}, \ t < 1/2$

$$C(t) = -\log(1-2t) = \sum_{r=0}^{\infty} \frac{C^{(r)}(t=0)}{r!} t^r = 2t + \frac{2^2}{2} t^2 + \frac{2^3}{3} t^3 + \cdots$$

$$\therefore C_r := (r-1)! \, 2^r$$

〈1.6 Inequalities〉

〈Theorem 1.6.1 : Jensen's Inequality〉

수직선 위의 구간 I 에서의 값을 갖는 확률변수 X의 기대값이 존재하면, 구간 I 에서 볼록한 함수 ϕ 에 대하여 다음 부등식이 성립한다.

(Convex function)

$$E[\phi(X)] \geq \phi(E[X])$$

Convex function: $\phi((1-t)\cdot x_1 + t x_2) \leq (1-t)\phi(x_1) + t\phi(x_2)$ $\quad 0 \leq t \leq 1$, $x_1, x_2 \in I$

(proof) we will prove Theorem 1.6.1 given that $\phi''(x) \geq 0$.

Let $\mu \in E[X]$. $\mu \in I$ and from Taylor's theorem,

$$\phi(x) = \phi(\mu) + \phi'(\mu)(x-\mu) + \phi''(\xi)(x-\mu)^2/2 \quad \exists \xi \in [x, \mu] \text{ or } [\mu, x]$$

Since $\phi''(x) \geq 0 \ \forall x \in I$, $\phi(x) \geq \phi(\mu) + \phi'(\mu)(x-\mu) \ \forall x \in I$

$$\int_{-\infty}^{+\infty} \phi(x)\cdot f(x)\,dx \geq \int_{-\infty}^{+\infty} \phi(\mu)\cdot f(x)\,dx + \int_{-\infty}^{+\infty} \phi'(\mu)(x-\mu)\cdot f(x)\,dx$$

$$\Rightarrow E[\phi(x)] \geq E[\phi(\mu)] + E[\phi'(\mu)(x-\mu)]$$

$\phi'(\mu)[E[x]-\mu] = \phi'(\mu)\cdot 0 = 0$

$$\Rightarrow E[\phi(x)] \geq \phi(\mu) = \phi(E[x]) + 0$$

Ex. 1.6.1 X^2 is a convex function on $(-\infty, +\infty)$

$$\therefore E[\phi(X)] \geq \phi[E[X]] \Rightarrow E[X^2] \geq \mu^2 \quad \forall x \in (-\infty, +\infty)$$

〈Theorem 1.6.2 Lyapunov's Inequality〉 (교과 오타 i ⇒ y)

If $E[|X|^s] < +\infty$ for a random variable X, if $0 < r < s$

$$E[|X|^r]^{1/r} \leq E[|X|^s]^{1/s}$$

(proof) If $p>1$, $\phi(y) = y^p I_{[y \geq 0]}(y)$ is a convex function

1. Handle the trivial case: If $t=0$, $t=1$, $x_1 = x_2$ the equality holds trivially. If $x_1 = 0$ (and $x_2 < 0$, $0 < t < 1$), the inequality becomes $(t x_2)^p \leq t x_2^p$. This simplifies to $t^p \leq t$, which is true because $p>1$ and $0<t<1$.

2. Normalize the inequality: Assume $0 < x_1 < x_2$ and $0 < t < 1$. We can divide the entire inequality by x_1^p to simplify it into a function of a single variable.

$$\frac{((1-t)x_1 + t x_2)^p}{x_1^p} \leq \frac{(1-t)x_1^p + t x_2^p}{x_1^p}$$

$$\left((1-t) + t\frac{x_2}{x_1}\right)^p \leq (1-t) + t\left(\frac{x_1}{x_2}\right)^p$$

Let $u = x_2/x_1$. $\because x_2 > x_1$, $u > 1$

$$\Rightarrow (1-t+tu)^p \leq 1 - t + t u^p$$

3. Define the difference function

$$g(u) := \underbrace{(1-t+tu^p)}_{RHS} - \underbrace{(1-t+tu)^p}_{LHS}$$

4. Analyze the function

$$g(u=1) = (1-t+t) - (1-t+t)^p = 1 - 1^p = 0$$

$$g'(u) = t\cdot p\cdot u^{p-1} - p(1-t+tu)^{p-1}\cdot t$$
$$= pt\left[u^{p-1} - (1-t+tu)^{p-1}\right]$$

5. Determine the Sign of the derivative

Is $u > (1-t+tu)$?

$u - tu > 1 - t$

$u(1-t) > 1 - t$

$u > 1$: True

$$\therefore u > (1-t+tu)$$

$\Rightarrow g'(u) > 0 \quad \forall u > 1 \quad \because p-1 > 0$

6. Conclusion: Since $g(1) = 0$ and $g'(u) > 0 \ \forall u > 1$, the function $g(u)$ is strictly increasing for $u > 1$. This implies $g(u) > 1 \ \forall u > 1$

$$RHS - LHS \geq 0 \Rightarrow LHS \leq RHS$$

Thus the inequality holds and y^p is a convex function

* $f''(x) \geq 0$ guarantees that f is a convex function.

(proof continued)

If $E[Y]$ exists for a random variable Y,

$$E[Y^p] \geq E[Y]^p$$

Let $p = s/r > 1$ and $Y = |X|^r$

$$E[|X|^{r\cdot\frac{s}{r}}] \geq E[|X|^r]^{\frac{s}{r}} \Rightarrow E[|X|^s] \geq E[|X|^r]^{s/r}$$
$$\Rightarrow E[|X|^s]^{1/s} \geq E[|X|^r]^{1/r} \quad \square$$

☆☆ 매우 중요

〈Theorem 1.6.3 : Markov Inequality & Chebyshev Inequality〉

(a) Markov Inequality : If $E[|Z|^r] < +\infty$ for a random variable Z for an arbitrary $K > 0$

$$P(|Z| \geq K) \leq E[|Z|^r] / K^r$$

(proof)

☆ (a) 이 부등식 증명 핵심은 다음과 같이 확률을 지표함수의 기대값으로 나타내는 데에 있다.

$P(|Z| \geq K) = E(I_{(|Z| \geq K)})$

$I_{(|Z| \geq K)} = 1 \cdot I_{(|Z|/K \geq 1)} \leq (|Z|/K)^r \cdot I_{(|Z|/K \geq 1)} \leq |Z|^r / K^r$

$E[I_{(|Z| \geq K)}] = P(|Z| \geq K) \leq E[|Z|^r] / K^r$

(b) 마르코프 부등식에서 다음과 같이 대입하면 체비셰프 부등식을 얻을수 있다.

$Z = X - E[X]$, $r = 2$

$$P(|X - E[X]| \geq K) \leq \frac{E[|X - E[X]|^2]}{K^2} = \frac{Var(X)}{K^2}$$

Ex 1.6.2 probability of X being beyond 6 sigma from mean

$$P(|X - \mu| > 6\sigma) = \frac{Var(X)}{(6\sigma)^2} = \frac{\sigma^2}{36\sigma^2} = \frac{1}{36} = 0.0277\cdots$$

〈Theorem 1.6.4. Meaning of $Var(X) = 0$〉

If the random variable X's mean is $\mu = E[X]$ and $Var(X) = 0$
$P(X = \mu) = 1$

(proof)

확률변수 X가 μ와 다른 값들을 가지는 사건을 다음과 같이 나타낼수 있다

$$(|X - \mu| > 0) = \bigcup_{n=1}^{\infty} \left(|X - \mu| \geq \frac{1}{n}\right)$$

따라서 확률의 가산가법성으로부터

$$P(|X - \mu| > 0) = P\left\{\bigcup_{n=1}^{\infty} \left(|X - \mu| \geq \frac{1}{n}\right)\right\} \leq \sum_{n=1}^{\infty} P\left(|X - \mu| \geq \frac{1}{n}\right)$$

한편, chebyshev's inequality와 $Var(X) = 0$ 으로부터

$0 \leq P(|X - \mu| \geq \frac{1}{n}) \leq Var(X) / (1/n^2) = 0$

$\therefore P(|X - \mu| \geq 1/n) = 0 \quad (n = 1, 2, \cdots)$

$\therefore 0 \leq P(|X - \mu| > 0) \leq \sum_{n=1}^{\infty} P(|X - \mu| \geq \frac{1}{n}) = 0$

$\therefore P(|X - \mu| > 0) = 0$

따라서 $P(X = \mu) = 1 - P(|X - \mu| > 0) = 1$