Mathematical Statistics

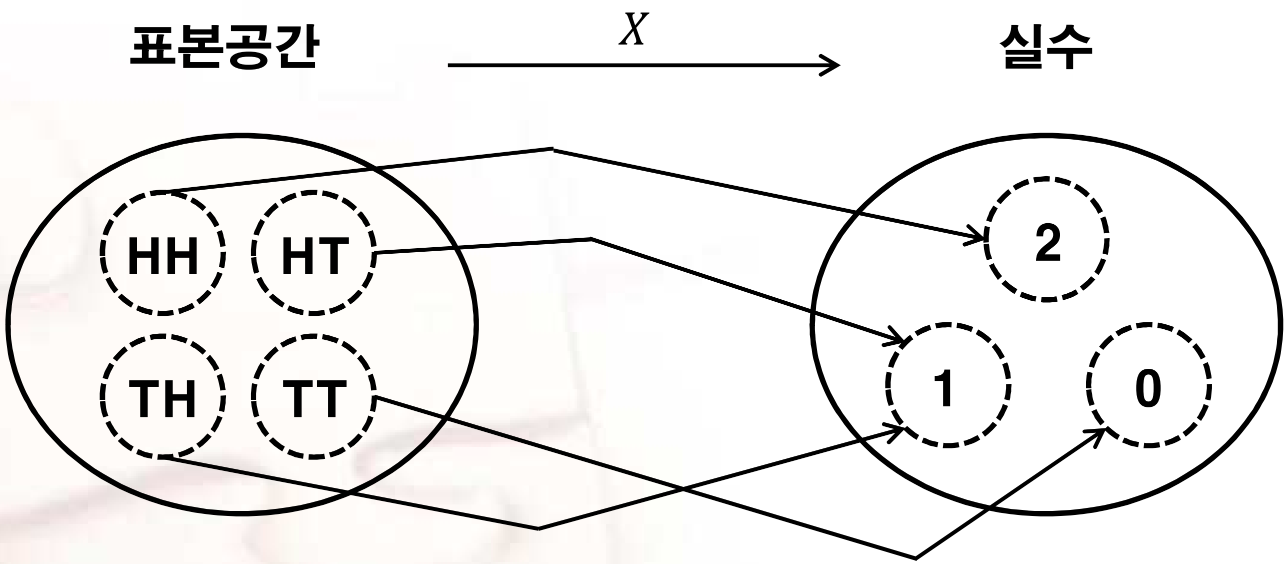
- Implementation with R -

Chap 2. Random Variables and Probability Distributions

2.0 Introduction

Def 2.1 Random Variable(확률변수)

- A function X that assigns to each element s in Ω one and only one real number $X(s) = x$
- 표본공간 내의 각 사건들에 실수 값을 대응시키는 함수



2.0 Introduction

Examples of random variables

(a) 주사위를 던질 때 나오는 눈의 수 V

(b) 어느 공장에서 생산된 1,000개의 제품 중 불량품의 수 W

(c) 대학입시에서 합격여부 X (예 : 합격이면 1, 불합격이면 0)

(d) 동전을 세 번 던질 때 앞면이 나오는 횟수 Y

(e) 어느 공장에서 생산되는 전구의 수명 Z

2.0 Introduction

[예 2.1] X : 앞면이 나올 때까지 동전을 던지는 실험에서
총 동전을 던지는 횟수

- 표본공간의 각 원소가 X에 의하여 어떤 실수 값으로 변환되는지?
- X에 의하여 실수의 부분집합에서 정의된 새로운 확률?

2.0 Introduction

Def. Probability Distribution(확률분포)

- Denotes the possible outcomes of a random variables with its probability
- 확률변수가 취할 수 있는 값과 각 대응하는 확률을 나타낸 것

Def. Cumulative Distribution Function(누적분포함수, CDF)

- Means the probability of random variable X belongs to ($-\infty, x$] that is, $F(x) = P(X \leq x)$

2.0 Introduction

Properties of CDF

(i) $F(x)$ is non-decreasing function (비감소함수),

For $x \le y,\ F(x) \le F(y)$

$(i\ )$ $\lim\limits_{n\to-\infty} F(x) = 0,\ \lim\limits_{n\to\infty} F(x) = 1$

$(i\ \ )$ $F(x)$ is right continuous function (우연속함수),

i.e. $\lim\limits_{n\to\infty} F\left(x + \frac{1}{n}\right) = F(x)$

cf. $\lim\limits_{n\to\infty} F\left(x - \frac{1}{n}\right) = P(X < x) \ne F(x)$

2.0 Introduction

[예 2.2] 확률변수 X의 cdf $F(x)$가 다음과 같을 때, 아래의 확률?

$$F(x) = \begin{cases} 0, & x < 0 \\ x^2, & 0 \le x < \frac{1}{4} \\ \frac{1}{2}, & \frac{1}{4} \le x < \frac{1}{2} \\ x, & \frac{1}{2} \le x < 1 \\ 1, & 1 \le x \end{cases}$$

(a) $P\left(X \le \frac{1}{8}\right)$ (b) $P\left(X < \frac{1}{8}\right)$ (c) $P\left(X = \frac{1}{8}\right)$ (d) $P\left(X \le \frac{1}{4}\right)$

(e) $P\left(X < \frac{1}{4}\right)$ (f) $P\left(X = \frac{1}{4}\right)$ (g) $P\left(\frac{1}{4} < X \le \frac{3}{4}\right)$ (h) $P\left(\frac{1}{4} \le X \le \frac{3}{4}\right)$

2.0 Introduction

Discrete random variable(이산형 확률변수)

- Random variable that takes finite or countable discrete values
- 유한하거나 셀 수 있는 무한개의 이산적인 값을 취하는 확률변수

Probability mass function(p.m.f, 확률질량함수)

X : Discrete r.v, $f(x) \equiv p(x) = P(X = x)$: pmf

$\Leftrightarrow$ (a) For all x, $f(x) \geq 0$

(b) $\sum_x f(x) = 1$

(c) $P(a \leq X \leq b) = \sum_{a \leq x \leq b} f(x)$

2.0 Introduction

[예 2.3] 동전을 2번 던져 나온 앞면의 수를 X라 할 때 X가 가질 수 있는 값과 확률질량함수?

[예 2.4] 앞면이 나올 때까지 동전을 던지는 횟수를 X라 할 때 X가 가질 수 있는 값과 확률질량함수?

2.1.1 Bernoulli distribution (베르누이 분포)

Bernoulli trial :

random experiment with exactly two possible outcomes, "success" and "failure“

Bernoulli random variable :

$$X = \begin{cases} 1, & \text{if 성공} (with\ probability\ p) \\ 0, & \text{if 실패} (with\ probability\ 1-p) \end{cases}$$

Notation & pmf :

$X \sim Bernoulli(p)$ & $p(x) = p^x (1-p)^{1-x}, \quad x = 0$ 또는 1

2.1.2 Binomial distribution (이항분포)

$$\text{When } X_1, X_2, \cdots, X_n \overset{iid}{\sim} Bernoulli(p), \text{ let}$$

$$Y_n = X_1 + X_2 + \cdots + X_n$$

$$\text{then, } Y_n \sim B(n,p) \text{ : Binomial distribution}$$

pmf of Y_n :

$$P(Y_n = y) = \binom{n}{y} p^y q^{n-y}, \quad y = 0, 1, \cdots, n$$

2.1 Discrete Probability Distributions

[예 2.5] 주사위를 5번 던져 3이나 6이 나오는 횟수 : X

[예 2.6]

Note. Discrete Uniform distribution: $U(1, N)$ (이산균일분포)

N : Parameter (positive integer)

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

2.1.3 Geometric distribution (기하분포)

X : the number of trials needed to observe the 1st success
with success probability p

한 번 시행에서 성공률 p인 베르누이 시행을 반복할 때,
첫 번째 성공이 나올 때까지 시행 횟수

$X \sim Geometric(p)$

$$\Leftrightarrow f(x) \equiv P(X=x) = P(x\text{번째 시행에서 처음 성공이 나오는 사건})$$
$$= P(\text{처음 } (x-1)\text{번은 계속 실패, } x\text{번째는 성공})$$
$$= (1-p)^{x-1}p, \quad x = 1, 2, \cdots$$

[예 2.7] 주사위를 던져 3이나 6이 나올 때까지 던지는 시행 횟수 : X,
X의 확률질량함수와 $X = 5$일 확률?

[풀이] 주사위의 눈이 3이나 6이 나올 확률은 $\frac{1}{3}$,

X는 성공(앞면)의 확률 $p = \frac{1}{3}$인 기하분포를 따르는 확률변수.

X의 확률질량함수는

$$p(k) = (1 - \frac{1}{3})^{k-1} \frac{1}{3}, \qquad k = 1, 2, \cdots$$

$$P(X = 5) = p(5) = (1 - \frac{1}{3})^4 \frac{1}{3} = \frac{16}{243}$$

2.1 Discrete Probability Distributions

<Note 1> $X \sim Geometric(p)$ 일 때,

$$P(X > x) = (1-p)^x, \quad x = 1, 2, \cdots$$

<Note 2> Memoryless property (기하분포의 비기억성)

When $X \sim Geometric(p)$, for $x > y$, $P(X > x \mid X > y) = P(X > x - y)$

2.1.4 Negative Binomial (음이항분포)

Bernoulli trial with success probability p : $\begin{cases} S, & w.p \quad p \\ F, & w.p \ 1-p \end{cases}$

X : the number of trials needed to observe the rth success

$$\Rightarrow X \sim NB(r,p), \quad 0 \le p \le 1, \quad r \ge 1$$

p.m.f : $f(x) = P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \ \ x = r, r+1, r+2, \cdots$

e.g. $r = 3$

SSS $\quad X = 3$

FSSS
SFSS $\quad X = 4$
SSFS

$\Rightarrow \binom{x-1}{r-1}$: # of cases putting $(r-1)$ S into $(x-1)$ different boxes

FFSSS
FSFSS
FSSFS $\quad X = 5$
SFFSS
SFSFS
SSFFS

Derivation of pmf of NB(p, r)

$$P(X=k) = P(k\text{번째 시행에서 } r\text{번째 성공이 나오는 사건})$$

$$= P(\text{처음 } (k-1)\text{번 시행에서 성공이 } (r-1)\text{번 나오고}$$

$$\text{마지막 } k\text{번째 시행은 성공인 사건})$$

$$= P(\text{처음 } (k-1)\text{번 시행에서 성공이 } (r-1)\text{번인 사건})$$

$$\times P(\text{마지막 } k\text{번째 시행은 성공인 사건})$$

$$= P(Y=r-1) \times p, \quad Y \sim B(k-1, p)$$

$$= \binom{k-1}{r-1} p^{r-1}(1-p)^{k-r} \times p$$

2.1.5 Hypergeometric distribution (초기하분포)

n : Size of the population items

r : # of items having specific characteristics $(0 \leq \mathrm{M} \leq N)$

m : Sample size $(1 \leq m \leq N)$

X : # of items having specific characteristics among selected sample

Example

n : A 공장에서 오늘 생산된 제품의 수

r : n개의 제품 중 불량 품의 수

$n - r$: n개의 제품 중 양호품의 수

m : n개의 제품 중 비복원추출로 뽑은 표본의 수

X ≡ 뽑힌 표본 가운데 불량품의 수,

$$\max(0, m - n + r) \leq X \leq \min(r, m)$$

$P(X = k)$??

$\binom{n}{m}$: 서로 다른 표본이 뽑히는 경우의 수

$\Longrightarrow$ 각 표본이 뽑힐 확률 : $\frac{1}{\binom{n}{m}}$

2.1 Discrete Probability Distributions

$\binom{r}{k}$: 불량품에서 k 개를 뽑는 경우의 수

$\binom{n-r}{m-k}$: 양호품을 뽑는 경우의 수

==⇒ 초기하분포의 확률질량함수 :

$X \sim HyperGeometric(n, r, m)$일 때,

$$p(k) = P(X = k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}},$$

$$\max(0, m - n + r) \leq k \leq \min(r, m)$$

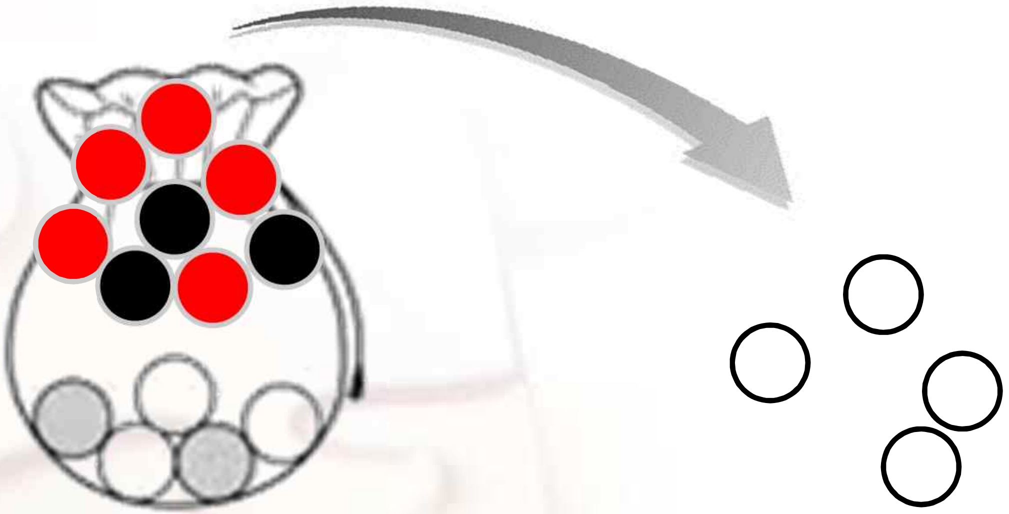
2.1 Discrete Probability Distributions

[예 2.8]

주머니 : 빨간색 구슬 5개, 검정색 구슬 3개

임의로 4개의 구슬을 꺼낼 때 빨간색 구슬의 수 : X

X의 확률질량함수?



2.1.6 Poisson distribution(포아송분포)

- Some experiments result in counting the number of particular events occur in the given times or on given physical objects
- 특정한 사건이 어떤 일정한 시간대에 또는 특정 영역에서 발생한 횟수가 따르는 분포

- **Examples** :
  - the number of phone calls arriving at a switchboard for a given time period
  - the number of flaws in 100 feet of wire
  - the number of customers that arrive at a shop between 10 and 12 AM
  - the number of typos in the given page of a textbook

Requirement for the poisson process(포아송의 우선조건)

1. The probability of occurring two or more events in a sufficiently short interval is essentially zero.
2. The probability of occurring an event in a given interval is proportional to the length of the interval.
3. The number of occurring events in non-overlapping intervals are independent.

Probability mass function (pmf)

When $X \sim Poisson(\lambda)$,

$$f(x) = P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0,\ 1,\ 2,\ \cdots$$

참고

단위(시간, 공간)당 어떤 사건이 평균적으로 λ회 발생된다고 할 때,

X : t단위(시간, 공간)동안 발생하는 사건의 수

$\Rightarrow X \sim Poisson(\lambda t)$이고, X의 확률질량함수는

$$P(X=x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, \quad x = 0,\ 1,\ 2,\ \cdots$$

2.1 Discrete Probability Distributions

포아송분포의 확률질량함수 모양 : [그림 2.4] 참고

[예 2.9] 전화통화 수가 시간당 0.5통인 포아송분포를 따름

(a) 5시간 동안 한 통화도 걸려오지 않을 확률

: t시간 동안 걸려오는 전화통화 수는 Poisson$(0.5t)$ 분포를 따름을 이용

(b) 5시간 동안 걸려오는 통화 수가 1일 확률

2.2 Continuous Random Variable

<u>연속형 확률변수</u> (continuous random variable)

: Random variable with possible outcomes in an interval.

일정한 구간에 속하는 모든 값을 취할 수 있는 확률변수

Probability Density Function (p.d.f, 확률밀도함수)

$f(x)$: 확률밀도함수

$\Leftrightarrow$ $i)\ f(x) \geq 0$

$\boldsymbol{i}\ )\ \int_{-\infty}^{\infty} f(x)dx = 1$

$\boldsymbol{i}\ )\ P(a \leq X \leq b) = P(a < X \leq b) = p(a < X < b) = F(b) - F(a)$

$= \int_{-\infty}^{\infty} f(x)dx$

Here, $F(x)$ is the cdf of X.

2.2 Continuous Random Variable

Relation between pdf and cdf:

$$F(x)=\int_{-\infty}^{x} f(t)dt \qquad \& \qquad f(x)=\frac{d}{dx}F(x)$$

Graphical meaning of the pdf

$$P(x \leq X \leq x+\Delta x)=\int_{x}^{x+\Delta x} f(t)dt \approx f(x)\Delta x$$

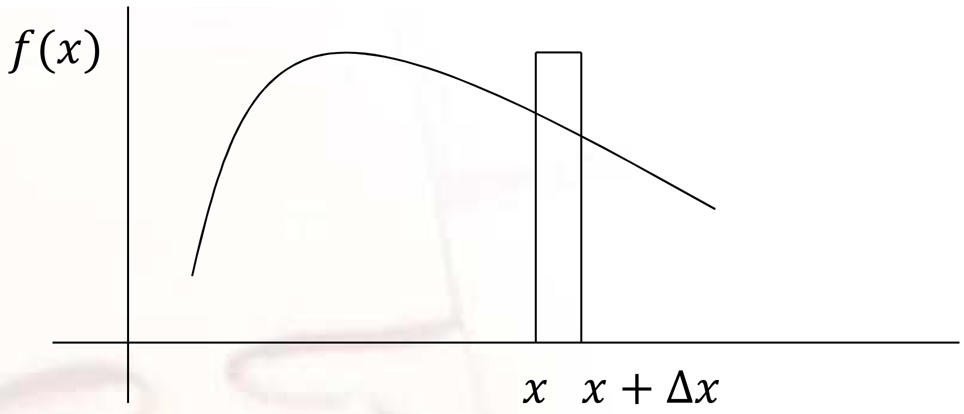


그림 2.5: 구간 $(x, x+\Delta x]$에 속할 확률

2.2 Continuous Random Variable

[예 2.10] X를 [0,1]에서 임의로 선택된 수라 할 때 X의 확률밀도함수

[예 2.11] 연속형 확률변수 X의 확률밀도함수가 $f(x) = 2x, 0 \le x \le 1$일 때,

(a) X의 누적분포함수

(b) $X \in [0,0.5]$일 확률

2.2.1 Uniform distribution (균일분포)

- 연속형 분포 중 가장 간단한 형태
- 균등분포라 불리기도 함
- 유한인 특정한 구간 안에서 동일한 확률 구조를 갖고, 그 구간 밖에서는 확률 0

pdf

확률변수 X가 구간 (a, b)에서 균일분포를 따를 때, X의 확률밀도함수 :

$$f(x) = \frac{1}{b-a}, a < x < b$$

2.2 Continuous Random Variable

Notation

$$X \sim U(a, b)$$

Cdf

$$F(x) = P(a \le X \le x) = \int_a^x \frac{1}{b-a} dx = \frac{x-a}{b-a}, \ a \le x \le b$$

[예 2.12] X가 구간 [2,5]에서 임의로 선택된 수일 때,

(a) X가 4이하일 확률

(b) X의 확률밀도 함수

2.2.2 Exponential distribution (지수분포)

X : Waiting time until the first occurrence of an event for the poisson process with mean λ

평균이 λ인 포아송 분포에서 첫 번째 사건이 일어날 때까지의 대기시간.

$$
\begin{aligned}
P(X \le t) &= 1 - P(X > t) = 1 - P(\text{첫 번째 사건이 발생하는 시간} > t) \\
&= 1 - P([0,t] \text{ 사이에서 발생하는 사건의 수} = 0) \\
&= 1 - P(Poisson(\lambda t) = 0) \\
&= 1 - e^{-\lambda t}
\end{aligned}
$$

2.2 Continuous Random Variable

pdf of X :

$$f(x) = \frac{d}{dx}F(x) = \frac{d}{dx}P(X \le x) = \lambda e^{-\lambda x},\ x \ge 0,\ \lambda > 0$$

Distribution of X : Exponential distribution & $X \sim Exp\,(1/\lambda)$

(H.W.) : Use R

지수분포의 확률밀도함수 형태를 몇 가지 λ값에 대해 그려볼 것

Note 1. Other representation of exponential distribution

지수분포의 확률밀도함수는 $\theta = \frac{1}{\lambda}$로 치환해서

$$\frac{1}{\theta}e^{-\frac{x}{\theta}},\ x \geq 0,\ \theta > 0$$

로 나타내기도 함. 이 경우 표기는 $X \sim Exp(\theta)$이 됨.

Note 2. Memoryless property of exponential distribution (지수분포의 비기억성)

For $0 < t < s,\ P(X > s \mid X > t) = P(X > s - t)$

(*Proof*)

$$P(X > s \mid X > t) = \frac{P(X > s, X > t)}{P(X > t)} = \frac{P(X > s)}{P(X > t)}$$

$$= \frac{e^{-\lambda s}}{e^{-\lambda t}} = e^{-\lambda(s-t)} = P(X > s - t)$$

2.2.3 Gamma Distribution (감마분포)

<u>Gamma function</u> (감마함수)

For , $a > 0, \Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$

Therefore, $f(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x},\ x > 0$ can be a p.d.f.

2.2 Continuous Random Variable

Facts about gamma function :

- $\Gamma(1) = \int_0^\infty e^{-x} dx = 1$
- $\alpha > 1$에 대해, $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$
- 자연수 n에 대해 $\Gamma(n) = (n-1)!$
- $0 < \alpha < 1$에서 $\Gamma(\alpha)$의 값만 있으면 모든 α에 대한 값을 구할 수 있음

예 :

$$1 < \alpha \leqq 2 \Rightarrow \Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$$

$$2 < \alpha \leqq 3 \Rightarrow \Gamma(\alpha) = (\alpha-1)(\alpha-2)\Gamma(\alpha-2)$$

$$\vdots \qquad \vdots$$

(*Proof of the second fact*)

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

$$= x^{\alpha-1}(-e^{-x})|_0^\infty - \int_0^\infty -e^{-x}(\alpha-1)x^{\alpha-2} dx$$

$$= \int_0^\infty (\alpha-1)x^{\alpha-2} e^{-x} dx$$

$$= (\alpha-1)\Gamma(\alpha-1)$$

2.2 Continuous Random Variable

pdf of gamma distribution :

$$f(x|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}}, \quad x > 0,\ \alpha > 0,\ \beta > 0$$

α : shape parameter, β : scale parameter

Notation :

$$X \sim Gamma(\alpha,\beta)$$

Expected value and variance :

When $X \sim Gamma(\alpha, \beta)$,

$$E(X) = \int_0^\infty x \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} dx$$

$$= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{(\alpha+1)-1} e^{-\frac{x}{\beta}} dx$$

$$= \frac{\Gamma(\alpha+1)\beta^{\alpha+1}}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty \frac{1}{\Gamma(\alpha+1)\beta^{\alpha+1}} x^{(\alpha+1)-1} e^{-\frac{x}{\beta}} dx$$

$$= \frac{\Gamma(\alpha+1)\beta^{\alpha+1}}{\Gamma(\alpha)\beta^\alpha}$$

$$= \alpha\beta$$

$Var(X) = \alpha\beta^2$ *(H.W.)*

Subset of gamma distribution :

① Exponential distn (지수분포)

$X \sim Gamma(\alpha, \beta)$ 일 때, $\alpha = 1$이면 $X \sim Exp(\beta)$

$$f(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}, x > 0$$

② Chi-square distribution (카이제곱분포)

$X \sim Gamma(\alpha, \beta)$일 때, $\alpha = \frac{r}{2}, \beta = 2$ (r은 양의 정수)이면

$X \sim \chi^2(r)$ (즉, 자유도가 r인 카이제곱분포)

$$f(x) = \frac{1}{\Gamma(\frac{r}{2})2^{r/2}} x^{r/2-1} e^{-x/2}, \; x > 0$$

$$E(X) = r/2 \times 2 = r$$

$$Var(X) = r/2 \times 2^2 = 2r$$

2.2.4 Normal distribution (정규분포)

pdf of normal distribution (정규분포의 확률밀도함수) :

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty, \ \sigma^2 > 0$$

Notation (정규분포의 표기) :

$$X \sim N(\mu, \sigma^2)$$

Expected value and variance (기댓값과 분산) :

$$E(X) = \mu, \quad Var(X) = \sigma^2$$

<u>Properties of normal distribution (정규분포의 특징)</u> :

- 연속형 분포 가운데 가장 중요
- 많은 분포형태가 정규분포에 가까움
- 모집단에 관계없이 표본평균의 분포는 정규분포로 근사됨 (CLT)
- 다른 확률분포들을 정규분포로 근사시킬 수 있음

<u>Facts related to normal distribution (정규분포와 관련된 사실)</u> :

(1) $X \sim N(\mu, \sigma^2)$일 때,

$$P(a < X < b) = \int_a^b \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx:$$ 직접계산 불가

$\Rightarrow$ 표를 이용

(2) $X \sim N(\mu, \sigma^2)$일 때, $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$

$Z = \dfrac{X-\mu}{\sigma}$: X를 표준화

(3) $X \sim N(\mu, \sigma^2)$일 때,

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right),$$

여기서, $\Phi(\cdot)$는 $N(0,1)$의 cdf. 즉, $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$

2.2 Continuous Random Variable

(4) 이항분포의 정규근사

$X \sim B(n,p)$이고, n이 충분히 클 때 $(\min(np, n(1-p)) \geq 5)$

$X \overset{\cdot}{\sim} N(\mu, \sigma^2)$, 여기서, $\mu = np, \sigma^2 = np(1-p)$.

예제 : $P(X=2) \approx P(1.5 \leq X \leq 2.5)$, 0.5 : 수정항

$X \sim B(25, 0.6) \Longrightarrow X \overset{\cdot}{\sim} N(15, 6)$

$$P(X \leq 13) \approx P(X \leq 13.5) = P\left(\frac{X-15}{\sqrt{6}} \leq \frac{13.5-15}{\sqrt{6}}\right)$$

$$= P(Z \leq -0.61) = 0.271$$

실제로는 $P(X \leq 13) = 0.267$

2.2.5 베타분포 (beta distribution)

<u>베타분포의 확률밀도함수</u> :

$$f(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1},\ 0 < x < 1,\ \alpha,\beta > 0$$

<u>베타분포의 표기</u> :

$$X \sim Beta\ (\alpha,\beta)$$

<u>베타분포의 기댓값과 분산</u> :

$$E(X) = \frac{\alpha}{(\alpha+\beta)},\ Var(X) = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$$

(증명)

$$EX^n = \int_0^1 x^n \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{n+\alpha-1}(1-x)^{\beta-1}dx$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(n+\alpha)\Gamma(\beta)}{\Gamma(n+\alpha+\beta)} \int_0^1 \frac{\Gamma(n+\alpha+\beta)}{\Gamma(n+\alpha)\Gamma(\beta)} x^{n+\alpha-1}(1-x)^{\beta-1}dx$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{\Gamma(n+\alpha)}{\Gamma(n+\alpha+\beta)}$$

2.2 Continuous Random Variable

n이 양의 정수 일 때,

$$EX^n = \frac{(n+\alpha-1)(n+\alpha-2)\cdots\alpha\Gamma(\alpha)\Gamma(\alpha+\beta)}{\Gamma(\alpha)(n+\alpha+\beta-1)(n+\alpha+\beta-2)\cdots(\alpha+\beta)\Gamma(\alpha+\beta)}$$

$$= \Pi_{i=0}^{n-1}\frac{(\alpha+i)}{(\alpha+\beta+i)}$$

$$\Longrightarrow \quad E(X) = \frac{\alpha}{\alpha+\beta}, \qquad E(X^2) = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$$

<u>베타분포의 확률밀도함수 형태</u> :

몇 개의 α, β 조합에 대해 그려볼 것 ***(H.W.)***

2.2.6 코쉬분포(Cauchy distribution)

<u>코쉬분포의 확률밀도함수</u> :

$$f(x|\theta) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2},\ -\infty < x < \infty,\ -\infty < \theta < \infty$$

<u>코쉬분포의 표기</u> :

$$X \sim C(\theta, 1)$$

<u>코쉬분포의 기댓값과 분산</u> : 존재하지 않음

2.2.7 이중지수분포(double exponential distribution)

이중지수분포의 확률밀도함수 :

$$f(x|\mu,\sigma) = \frac{1}{2\sigma} e^{-\frac{|x-\mu|}{\sigma}}, \quad -\infty < x < \infty, \ -\infty < \mu < \infty, 0 < \sigma$$

이중지수분포의 표기 :

$$X \sim DE(\mu, \sigma)$$

이중지수분포의 기댓값과 분산 :

$$E(X) = \mu, \ Var(X) = 2\sigma^2$$ ***(H.W.)***

H.W. : $N(0,1)$, $C(0,1)$과 $DE(0,1)$의 확률밀도함수를 그리고 그 특성을 비교 서술하라.

2.3 Variable transformation : Change-of-Variable

When the distribution of X is given, what is the prob. distn. of $Y = g(X)$?

Ex : Given the p.d.f of X, what is the p.d.f of $Y = X^2$ or $\mathrm{Y} = e^X$?

2.3 Variable transformation

2.3.1 Discrete case

X : Discrete r.v. which can take $x_1, x_2, \cdots$

Then, $Y = g(X)$ is the r.v. which can take $g(x_1)$, $g(x_2)$, $\cdots$

$p_X(x)$: p.m.f of X

$p_Y(y)$: p.m.f of $Y = g(X)$

$$
\begin{aligned}
p_y(y) = P(Y = y) &= P(g(X) = y) \\
&= P\left(X \in g^{-1}(y)\right) \\
&= \sum_{x \in g^{-1}(y)} p_X(x)
\end{aligned}
$$

2.3 Variable transformation

Example : p.m.f of X

x	-1	0	1	2
$p_X(x)$	0.1	0.2	0.3	0.4

p.m.f of $Y = X^2$

y	0	1	4
$p_Y(y)$	0.2	0.4	0.4

2.3.2 Continuous case

$f_X(x), F_X(x)$: p.d.f and c.d.f of X

Method for finding the p.d.f of conti. r.v. $Y = g(X)$:

- Find the c.d.f of Y, $F_Y(y)$, and take the derivative
- Example : Simple linear transformation case

2.3 Variable transformation

(i) When $Y = 2X$,

$$F_Y(y) = P(Y \le y) = P(2X \le y) = P\left(X \le \frac{y}{2}\right) = F_X\left(\frac{y}{2}\right)$$

$$\therefore f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{1}{2}f_X\left(\frac{y}{2}\right)$$

(ii) When $Y = aX + b$, $(a > 0)$

$$F_Y(y) = P(Y \le y) = P(aX + b \le y) = P\left(X \le \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right)$$

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{1}{a}f_X\left(\frac{y-b}{a}\right)$$

2.3 Variable transformation

(iii) When $Y = aX + b$, $(a < 0)$

$$F_Y(y) = P(Y \le y) = P(aX + b \le y) = P\left(X \ge \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right)$$

$$f_Y(y) = \frac{d}{dy}F_Y(y) = -\frac{1}{a}f_X\left(\frac{y-b}{a}\right)$$

From (ii) and (iii), the p.d.f of Y when $Y = aX + b$:

$$f_Y(y) = \left|\frac{1}{a}\right| f_X\left(\frac{y-b}{a}\right)$$

2.3 Variable transformation

Ex 2.1 When $X \sim N(\mu, \sigma^2)$, what is the p.d.f of $Y = aX + b$? What about $Y = \frac{X-\mu}{\sigma}$?

[Sol.]

p.d.f of Y :

$$f_Y(y) = \left|\frac{1}{a}\right| f_X\left(\frac{y-b}{a}\right)$$

$$= \frac{1}{|a|}\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left(\frac{y-b}{a}-\mu\right)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2 a^2}} e^{-\frac{(y-a\mu-b)^2}{2\sigma^2 a^2}}$$

Therefore, $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$. When $a = \frac{1}{\sigma}$ and $b = -\frac{\mu}{\sigma}$, $a\mu + b = 0$, $a^2\sigma^2 = 1$. Hence, $Y = \frac{X-\mu}{\sigma}$ follows $N(0, 1)$.

2.3 Variable transformation

Note 1 Direct derivation of the above results using c.d.f.

pf)

$$
\begin{aligned}
P(Z \le z) &= P(\frac{X-\mu}{\sigma} \le z) \\
&= P(X \le z\sigma + \mu) \\
&= \int_{-\infty}^{z\sigma+\mu} \frac{1}{\sqrt{2\pi}\sigma}\, exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx
\end{aligned}
$$

$$
\begin{aligned}
\therefore\ f(z) = \frac{dP(Z \le z)}{dz} &= \frac{1}{\sqrt{2\pi}\sigma}\, exp\left\{-\frac{z^2\sigma^2}{2\sigma^2}\right\} \cdot \sigma \\
&= \frac{1}{\sqrt{2\pi}}\, exp\left(-\frac{z^2}{2}\right)
\end{aligned}
$$

i.e., $Z \sim N(0,1)$.

2.3 Variable transformation

General method for finding the p.d.f. of $Y = g(X)$

When the function $g(\cdot)$ is invertible and differentiable

(i) When $g(\cdot)$ is monotone increasing :

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P\left(X < g^{-1}(y)\right) = F_x(g^{-1}(y))$$

$$\therefore\ f_y(y) = \frac{d}{dy}F_Y(y) = f_x(g^{-1}(y)) \cdot \frac{d}{dy}g^{-1}(y)$$

2.3 Variable transformation

(i) When $g(\cdot)$ is monotone decreasing : Note that $\{g(X) \leq y\} \equiv \{X \geq g^{-1}(y)\}$

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P\big(X \geq g^{-1}(y)\big) = 1 - F_x(g^{-1}(y))$$

$$\therefore\ f_y(y) = \frac{d}{dy} F_Y(y) = -f_x(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y)$$

From (i) and (ii)

$$f_y(y) = f_x(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right|$$

2.3 Variable transformation

Thm 2.1 $f_X(x)$: pdf of X, let $Y = g(X)$ with $g(\cdot)$: invertible and differentiable.

Then, the pdf of Y :

$$f_y(y) = f_x(g^{-1}(y)) \cdot \left|\frac{d}{dy} g^{-1}(y)\right|$$

Ex 2.2 When X~Uniform[0,1], what is the pdf of $Y = \frac{1}{X}$?

[Sol.]

$f_X(x) = 1,\ 0 \le x \le 1.\ g^{-1}(y) = \frac{1}{y}$: diff'ble. By Thm 2.1,

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left|\frac{d}{dy} g^{-1}(y)\right| = f_X\left(\frac{1}{y}\right) \cdot \left|\frac{d}{dy}\frac{1}{y}\right| = \frac{1}{y^2}, \quad 1 \le y < \infty$$

2.3 Variable transformation

Method based on using CDF

Ex 2.3 $X \sim N(0, 1)$일 때 $Y = X^2$ 의 확률분포?

[Sol.]

$$F_Y(y) = P(Y \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

pdf of Y :

$$\begin{aligned} f_y(y) &= \frac{d}{dy} F_Y(y) = f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} + f_X(-\sqrt{y}) \frac{1}{2\sqrt{y}} \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} \frac{1}{\sqrt{y}} = \frac{1}{\sqrt{\pi}\ 2^{1/2}} y^{\frac{1}{2}-1} e^{-\frac{y}{2}}, \quad y > 0 \end{aligned}$$

: This corresponds to the Gamma distribution with $\alpha = \frac{1}{2}$, $\lambda = \frac{1}{2}$, i.e. pdf of $\chi^2(1)$

2.3 Variable transformation

Note 2 CDF 계산을 통한 직접 유도

$$X \sim N(\mu, \sigma^2),\ \sigma^2 > 0 \quad \Rightarrow \quad V = \left(\frac{X-\mu}{\sigma}\right)^2 \equiv Z^2 \sim \chi^2(1)$$

pf) $G(v)$: 확률변수 V의 분포함수, $v \geq 0$.

그러면, $G(v) = P(V \leq v) = P(Z^2 \leq v) = P(-\sqrt{v} \leq Z \leq \sqrt{v})$.

$$G(v) = \int_{\sqrt{v}}^{\sqrt{v}} \frac{1}{\sqrt{2\pi}}\, exp\left(-\frac{z^2}{2}\right) dz = 2\int_{0}^{\sqrt{v}} \frac{1}{\sqrt{2\pi}}\, exp\left(-\frac{z^2}{2}\right) dz$$

$z = \sqrt{y}$라 하자. 그러면, $dz = \dfrac{1}{2\sqrt{y}}dy$.

2.3 Variable transformation

$$\therefore \quad G(v) = \int_0^v \frac{1}{\sqrt{2\pi y}}\, exp\left(-\frac{y}{2}\right) dy, \; v \ge 0$$

$$\therefore \quad g(v) = G'(v) = \frac{1}{\sqrt{\pi}\sqrt{2}}\, v^{1/2-1} e^{-v/2}, \; 0 < v < \infty$$

$g(v)$: V의 pdf 이므로 $\int_0^\infty g(v)dv = 1$을 만족해야 함. $x = v/2$라 하면,

$$\frac{1}{\sqrt{\pi}} \int_0^\infty x^{1/2-1} e^{-x} dx = \frac{1}{\sqrt{\pi}}\, \Gamma\left(\frac{1}{2}\right) = 1 \;\Rightarrow\; \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

$$\therefore \quad g(v) = G'(v) = \frac{1}{\Gamma(1/2) 2^{1/2}}\, v^{1/2-1} e^{-v/2}, \; 0 < v < \infty$$

: $\chi^2(1)$의 확률밀도함수 $\equiv$ Gamma$(\frac{1}{2}, 2)$의 확률밀도함수

2.3 Variable transformation

Thm 2.2 F(·) : *X* 의 cdf일 때, *Y=F(X)* ~ Uniform(0,1)

[증명]

$$F_Y(y) = P(Y \le y) = P(F(X) \le y),\ 0 < y < 1.\ \ (F(x) : \text{단조증가함수})$$
$$= P(X \le F^{-1}(y)) = F(F^{-1}(y)) = y$$

$$\therefore Y \sim \text{Uniform}(0,1)$$

Note 3 X 가 평균이 $\frac{1}{\lambda}$인 지수분포를 따를 때, $F(x) = 1 - e^{-\lambda x}$이므로
$Y = 1 - e^{-\lambda X}$ ~ Uniform(0,1)

2.3 Variable transformation

Thm 2.3 $U \sim$ Uniform(0,1), $\mathrm{F}(\cdot)$: 연속형 확률분포의 cdf ⇒ $Y = F^{-1}(U) \sim \mathrm{F}(\cdot)$

[증명] Y의 cdf는

$$F_Y(y) = P(Y \leq y) = P(F^{-1}(U) \leq y) = P\big(U \leq F(y)\big) = F(y)$$

Note 4 $X \sim \mathrm{Exp}(\lambda)$일 때, $F(x) = 1 - e^{-\lambda x}$이고 $F^{-1}(x) = -\frac{1}{\lambda}\log(1-x)$이므로
$U \sim$ Uniform(0,1)일 때,

$$Y = -\frac{1}{\lambda}\log(1-U) \sim \mathrm{Exp}(\lambda)$$

난수발생 방법 설명