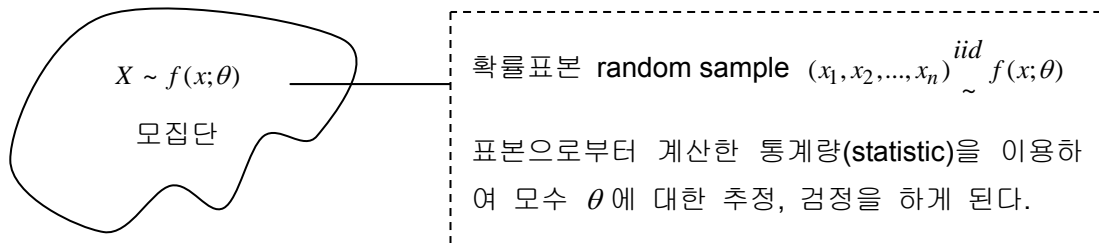


Chapter 6 확률변수 함수

확률변수(X , random variable)를 정의하고, 확률변수가 가지는 값에 대응하는 확률을 정의한 표나 식인 확률밀도함수($f(x)$, probability density function)를 구했다. 일정 구간의 확률을 적분한 (확률)분포함수($F(x)$, distribution function)도 정의하였다. 연속형인 경우 분포함수와 확률밀도함수 간에는 $F'(x) = f(x)$ 성립한다. 확률변수가 두 개 이상인 경우 결합 확률밀도함수(joint density function, $f(x_1, x_2, \dots, x_n)$), 주변확률밀도함수($f_k(x_k)$, marginal density function), 조건부 확률밀도함수($f(x_k | x_m)$) 등을 정의하였다. 또한 확률변수의 기대값을 정의하여 평균과 분산을 구하였다.

통계학(Statistics)은 모집단의 확률밀도함수 $f(x; \theta)$ 에 대한 추론(inferences)을 하거나 변수들 간의 관계 $Y = f(X_1, X_2, \dots, X_p)$ 을 알아내는 도구이다.



통계량을 이용하여 모수에 대한 검정을 위해서는 통계량의 확률밀도함수를 알아야 한다. 통계량은 확률표본의 함수로 이루어져 있으므로 확률변수의 함수에 대한 확률밀도함수를 구하는 방법에 대해 여기서 다루기로 한다. 통계량의 분포를 표본 분포(sampling distribution)이라 한다. 사실 우리는 이미 MGF의 유일성(uniqueness)에 의해 \bar{X} , $\sum X_i$ 의 확률밀도함수를 구하는 방법을 배웠다. 여기서는 자세히 다룰 것이다.

모집단의 평균 $\theta = \mu$ 에 대한 가설 검정을 위하여 사용되는 (검정)통계량은 $T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$

이다. 이를 위하여 우리는 \bar{X} 의 분포를 알아야 한다.

확률변수 X_1, X_2, \dots, X_n 의 함수 $Y = g(X_1, X_2, \dots, X_n)$ 의 확률밀도함수를 구하는 방법은 다음 3가지 방법으로 요약된다.

○분포함수 $F(x)$ 이용하기: 우선 $F(y) = P(Y \leq y) = P(g(X_1, X_2, \dots, X_n) \leq y)$ 을 구하고 이를 y 에 대해 미분하여 $f(y)$ 얻는다.

○변환(Transformation): $f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|$ where $Y = g(X)$, $X = g^{-1}(Y)$

○MGF 이용하기: MGF의 유일성(동일 MGF이면 동일한 pdf 갖는다)을 이용한다.

6.1 분포함수 이용하기 (The method of Distribution Functions)

X_1, X_2, \dots, X_n 의 함수 $Y = h(X_1, X_2, \dots, X_n)$ 라 하자.

- ① $Y = h(X_1, X_2, \dots, X_n)$ 의 영역(확률변수 y 가 가질 수 있는 영역)을 구한다.
- ② 적분에 의해 $F(y) = P(Y \leq y) = P(h(X_1, X_2, \dots, X_n) \leq y)$ 을 구한다.
- ③ $F(y)$ 을 y 에 대해 미분하여 y 의 확률밀도함수 $f(y)$ 구한다.



EXAMPLE 6.1

$f(x) = 2x, 0 \leq x \leq 1$ 일 경우 $Y = 3X - 1$ 의 확률밀도함수를 구하고 $E(Y)$ 구하시오.

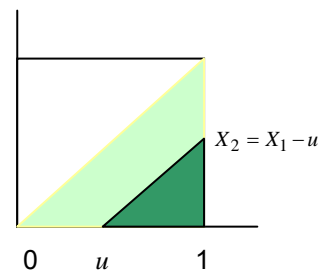
- ① Y 의 영역(range)은 $[-1, 2]$.
- ② $F(y) = \left(\frac{y+1}{3}\right)^2, -1 \leq y \leq 2$
- ③ Therefore, $f(y) = F'(y) = \frac{2}{9}(y+1), -1 \leq y \leq 2$



EXAMPLE 6.2

$f(x_1, x_2) = 3x_1, 0 \leq x_2 \leq x_1 \leq 1$ 일 경우 $U = (X_1 - X_2)$ 의 확률밀도함수를 구하시오.

- ① U 의 범위는 $0 \leq u \leq 1$.
- ② $F(u) = \frac{1}{2}(3u - u^3), 0 \leq u \leq 1$
- ③ Therefore $f(u) = F'(u) = \frac{3}{2}(1 - u^2), 0 \leq u \leq 1$



**EXAMPLE 6.3**

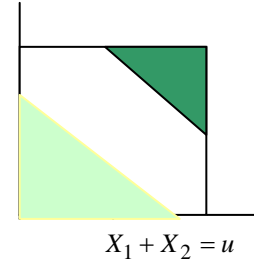
확률변수 X_1 과 X_2 은 $U(0,1)$ 으로부터 확률표본이다. $X_1 + X_2$ 의 확률밀도함수를 구하시오.

X_1, X_2 의 결합확률밀도함수는 $f(x_1, x_2) = 1, 0 \leq x_1, x_2 \leq 1$ 이다.

① U 의 범위는 $0 \leq u \leq 2$.

$$\textcircled{2} F(u) = \begin{cases} u^2/2, & 0 \leq u \leq 1 \\ -u^2/2 + 2u - 1, & 1 < u \leq 2 \end{cases}$$

$$\textcircled{3} f(u) = \begin{cases} u, & 0 \leq u \leq 1 \\ 2-u, & 1 < u \leq 2 \end{cases}$$

**EXAMPLE 6.4**

$f(x) = \frac{x+1}{2}, -1 \leq x \leq 1$ 일 경우 $U = X^2$ 의 확률밀도함수를 구하시오.

① U 의 범위는 $0 \leq u \leq 1$ 이다.

$$\textcircled{2} F(u) = P(U \leq u) = P(X^2 \leq u) = P(-\sqrt{u} \leq X \leq \sqrt{u}) = F_X(\sqrt{u}) - F_X(-\sqrt{u}).$$

$$\textcircled{3} f(u) = \frac{1}{2\sqrt{u}}, 0 < u \leq 1.$$

[TIP] 식 ②을 u 에 대해 미분하면 다음과 같다. $f(u) = \frac{1}{2\sqrt{u}} [f_Y(\sqrt{u}) + f_Y(-\sqrt{u})]$

**EXAMPLE 6.5**

확률변수 X 는 $Uniform(0,1)$ 을 갖는다. X 의 함수 $g(x)$ 가 평균 β 인 지수분포를 따른다고 할 때 $g(x)$ 을 구하시오.

확률변수 X 의 분포함수는 $F_X(x) = x, 0 \leq x \leq 1$ 이다.

$$F(u) = P(U \leq u) = P(g(x) \leq u) = P(X \leq g^{-1}(u)) = g^{-1}(u)$$

$U \sim \exp(\theta)$ 가 지수분포일 경우 $F_U(u) = \int_0^u e^{-t/\beta} dt = 1 - e^{-u/\beta}, 0 \leq u$ 이다.

그러므로 $1 - e^{-u/\beta} = x$ 이다. 즉 $U = g(x) = -\beta \ln(1-x)$ 이다. 이것을 이용하여 지수분포를 생성한다.

**HOMEWORK #16-1**

Let $f(x) = 2(1-x), 0 \leq x \leq 1$. Find the pdf of $U = X^2$ by the method of distribution function.

**HOMEWORK #16-2**

Let $X \sim Normal(0,1)$. Show that the pdf of X^2 is $\chi^2(1)$ by the method of dist. function.

▣ 표준 정규분포의 제곱은 자유도 1인 Chi-square 분포를 따른다.

▣ 표준 정규분포의 제곱의 합은 자유도 n 인 카이제곱 분포를 따른다.

**HOMEWORK #16-3**

DUE 5월 26일

Let $f(x_1, x_2) = e^{-x_1}, 0 \leq x_2 \leq x_1 < \infty$. Find the pdf of $U = X_1 - X_2$ and the expected value of U .

6.2 변수 변환 방법 (The method of Transformation)

확률변수 X 의 함수 $U = h(x)$ 가 비감소(non-decreasing) 혹은 비증가(non-increasing) 함수인 경우 변수 변환 방법에 의해 확률변수 함수의 분포를 구할 수 있다.

① $U = h(x)$ 가 비감소(non-decreasing) 함수인지 비증가(non-increasing) 함수인지 확인한다. 그렇다면 $U = h(x)$ 의 범위를 구한다.

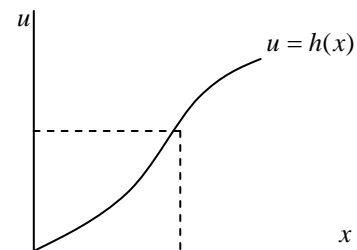
② $x = h^{-1}(u)$ 을 구하고 dx/du (u 에 대해 미분)을 얻는다.

③ U 의 확률밀도함수는 $f_U(u) = f_X(x) \left| \frac{dx}{du} \right|$ 이다.

비감소(Non-decreasing) 함수

$$F_U(u) = P(U \leq u) = P(h(x) \leq u) = P(x \leq h^{-1}(u)) = F_X(h^{-1}(u))$$

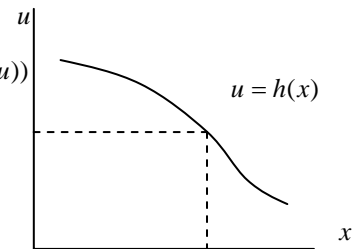
$$\text{Let } x = h^{-1}(u). \text{ Then, } f_U(u) = F'_X(x) = f_X(x) \frac{dx}{du}$$



비증가(Non-increasing) 함수

$$F_U(u) = P(U \leq u) = P(h(x) \geq u) = 1 - P(x \leq h^{-1}(u)) = 1 - F_X(h^{-1}(u))$$

$$\text{Let } x = h^{-1}(u). \text{ Then, } f_U(u) = -F'_X(x) = -f_X(x) \frac{dx}{du}$$



EXAMPLE 6.6

$f(x) = 2x, 0 \leq x \leq 1$ 일 경우 $U = 3X - 1$ 의 확률밀도함수를 구하시오.

①

②

③ $f_U(u) = \frac{2}{9}(u+1), -1 \leq u \leq 2$ 이다.

**EXAMPLE 6.7**

Let $f(x) = 2x, 0 \leq x \leq 1$. Find the pdf of $U = -4X + 3$.

①

②

③ U 의 확률밀도함수는 $f_U(u) = \frac{3-u}{8}, -1 \leq u \leq 3$ 이다.

이변량 결합확률밀도함수를 구할 때도 변수변환 방법을 이용할 수 있다. J 을 JACOBIAN(행렬식)이라 하자.

$$U = h_1(X_1, X_2), V = h_2(X_1, X_2) \rightarrow X_1 = h_1^{-1}(U, V), X_2 = h_2^{-1}(U, V)$$

$$f_{U,V}(u, v) = f_{X_1, X_2}(h_1^{-1}, h_2^{-1}) |J|, \text{ where } J = \begin{vmatrix} \frac{dx_1}{du} & \frac{dx_2}{du} \\ \frac{dx_1}{dv} & \frac{dx_2}{dv} \end{vmatrix}$$

다변량인 경우에도 사용할 수 있으나 Jacobian(행렬식)을 구하는 것이 쉽지 않아 자주 사용하지 않는다.

**EXAMPLE 6.8**

$f(x_1, x_2) = e^{-(x_1+x_2)}, 0 < x_1, x_2$ 일 경우 $U = X_1 + X_2$ 의 확률밀도함수를 구하시오.

$U = X_1 + X_2, V = X_1$ 라 하자. $0 \leq v \leq u$ 이고 $X_1 = V, X_2 = U - V$ 이다.

그러므로 $f_{U,V}(u, v) = e^{-u}, 0 \leq v \leq u$ 이고 U 의 주변밀도함수는 $f_U(u) = ue^{-u}, 0 \leq u$ 이다.

**EXAMPLE 6.9**

$f(x_1, x_2) = 2(1-x_1), 0 \leq x_1, x_2 \leq 1$ 일 경우 $U = X_1X_2$ 의 확률밀도함수를 구하시오.

$U = X_1X_2, V = X_1$ 라 놓자. $0 \leq u \leq v \leq 1$ 이고 $X_1 = V, X_2 = U/V$.

$$f_{U,V}(u, v) = 2(1-v) \frac{1}{v}, 0 \leq u \leq v$$

그러므로 $f_U(u) = 2(u - \ln u - 1), 0 \leq u \leq 1$ 이다.

**EXAMPLE 6.10**

$f(x_1, x_2) = 1, 0 < x_1, x_2 < 1$ 일 경우 $U = X_1 + X_2$ 와 $V = X_1 - X_2$ 의 결합밀도함수를 구하시오.

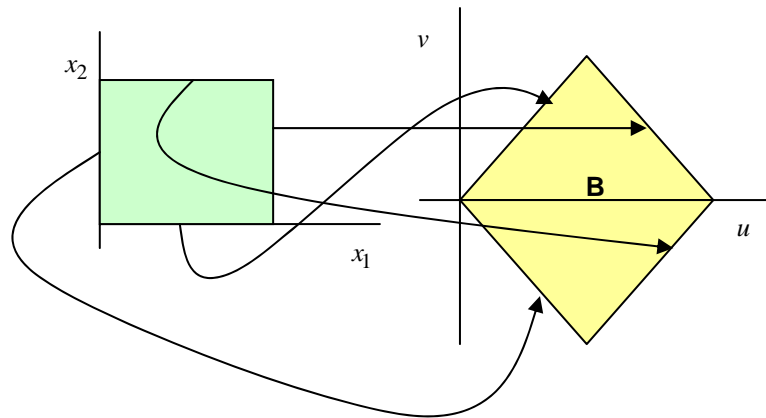
$U = X_1 + X_2, V = X_1 - X_2$.

$$X_1 = \frac{(U+V)}{2}, X_2 = \frac{(U-V)}{2}$$

$$f_{U,V}(u, v) = 1/2, (u, v) \in B$$

$$f_U(u) = \begin{cases} u, & 0 < u \leq 1 \\ 2-u, & 1 < u < 2 \end{cases}$$

$$f_V(v) = \begin{cases} v+1, & -1 < v \leq 0 \\ 2-v, & 0 < v < 1 \end{cases}$$

**HOMEWORK #17-1**

X_1, X_2 are a random sample of size 2 from $f(x) = e^{-x}, 0 < x$.

(1) Find the joint pdf of X_1, X_2 .

(2) Suppose $U = X_1 + X_2, V = X_1 - X_2$. Then find the joint pdf of U, V .

$$f_{U,V}(u, v) = \frac{1}{2} \exp^{-u}, (u, v) \in B$$

(3) Find the marginal pdf of U, V . The pdf of U is a Gamma distribution. The pdf of V is a double exponential distribution.

**HOMEWORK #17-2**

Let $f(x) = 2(1-x), 0 \leq x \leq 1$.

- (1) Find the pdf of $U = 1 - 2X$ by the method of transformation.
 (2) Find the pdf of $U = X^2$ by the method of transformation.

**HOMEWORK #7-6**

Let $f(x_1, x_2) = (1/8)x_1 e^{-(x_1+x_2)/2}, 0 < x_1, x_2$. Find the marginal pdf of $U = X_2 / X_1$.

6.3 MGF 방법

MGF의 유일성을 이용한다.

- ① $U = h(x)$ 의 범위를 구한다.
- ② X 의 MGF를 이용하여 U 의 MGF를 구한다.
- ③ 유명한 확률분포함수 MGF와 비교하여 확률밀도함수를 구한다.

**EXAMPLE 6.11**

$X \sim Normal(\mu, \sigma^2)$. Find the pdf of $Z = \frac{X - \mu}{\sigma}$ (standardization).

① $-\infty < Z = \frac{X - \mu}{\sigma} < \infty$.

② 정규분포 X 의 MGF는 $M_X(t) = e^{\mu t + \sigma^2 t^2 / 2}$ 이다.

그러므로 Z 의 확률밀도함수는 $M_Z(t) = E(e^{tz}) = e^{\sigma^2 t^2 / 2}$ 이다.

③ 그러므로 $Z \sim Normal(0,1)$

**EXAMPLE 6.11**

$Z \sim Normal(0,1)$ 일 경우 $U = Z^2$ 의 확률밀도함수를 구하시오.

① $0 < U < \infty$

② $M_U(t) = E(e^{tu}) = E(e^{tZ^2})$, $M_U(t) = \left(\frac{1}{1-2t}\right)^{1/2}$

③ $U \sim Gamma(\alpha = 1/2, \beta = 2)$, $U = \chi^2(\lambda = 1)$.

THEOREM (정리)

(X_1, X_2, \dots, X_n) 을 표본 크기 n 인 확률표본(random sample)이라 하자. $\sum_{i=1}^n X_i$ 의

MGF는 $M_{\sum X_i}(t) = M_{X_1}(t)M_{X_2}(t)\dots M_{X_n}(t) = [M_X(t)]^n$ 이다.

THEOREM (정리)

(X_1, X_2, \dots, X_n) 을 $N(\mu, \sigma^2)$ 으로부터 크기 확률표본이라 하자. $\sum_{i=1}^n a_i X_i$ 의 분포함수는

다음과 같다. $Normal(\mu \sum_{i=1}^n a_i, \sigma^2 \sum_{i=1}^n a_i^2)$.

$a_1 = a_2 = \dots = a_n = 1$ 이면 합의 분포, $a_1 = a_2 = \dots = a_n = 1/n$ 이면 평균의 분포이다.

**EXAMPLE 6.12**

일정 단위 시간에 은행 서비스 창구에 도착한 고객의 수는 λ 인 포아송(Poisson) 분포를 따른다고 하자. 이런 경우 고객이 오는 사이 시간의 분포는 $\beta = \frac{1}{\lambda}$ 인 지수분포를 따른다.

즉 사이 시간 Y 의 확률분포함수는 $f_X(x) = \frac{1}{\beta} e^{-x/\beta}$, $0 < x$ 이다. 고객 n 명이 서로 독립적

으로 은행 서비스 데스크에 왔다고 하자. 각 사이 시간을 X_1, X_2, \dots, X_n 이라 할 때

$U = \sum X_i$ 의 확률분포함수는 $(\alpha = n, \beta)$ 인 감마분포(Gamma)를 따름을 보이시오.

◎ 지수분포(β)로부터 확률표본 합은 감마분포($\alpha = n, \beta$)를 따른다.

$$\textcircled{1} 0 < U = \sum X_i < \infty.$$

$$\textcircled{2} M_U(t) = E(e^{tu}) = \sum E(e^{tX_i}) = (1 - \beta t)^{-n}$$

$$\textcircled{3} U \sim \text{Gamma}(n, \beta).$$



HOMEWORK #17-3

Suppose Z_1, Z_2, \dots, Z_n is a random sample of size n from $N(0,1)$. Show that the pdf of

$$\sum Z_i^2 \text{ is } \chi^2(\lambda = n).$$

⊙ 표준정규분포를 따르는 확률표본들의 제곱 합은 자유도 $\lambda = n$ 인 χ^2 분포 따른다.



HOMEWORK #17-4

Suppose $X_1, X_2 \sim \text{Normal}(0, \sigma^2)$ and are independent. Show that $U = X_1 + X_2$ and $V = X_1 - X_2$ are independent and the marginal pdf of them is $\text{Normal}(0, 2\sigma^2)$ respectively.



HOMEWORK #17-5

Let $X \sim \text{Gamma}(\alpha = n/2, \beta)$. Show $U = 2X / \beta \sim \chi^2(\lambda = n)$.

⊙ 감마분포와 카이제곱(χ^2) 분포의 관계

6.4 순서통계량 (Order Statistics)

(X_1, X_2, \dots, X_n) 은 확률밀도함수 $f(x)$ 으로부터의 확률표본(random sample)이라 하자. 표본을 크기 순으로 정렬하고 $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ 표현하자. 이를 순서 통계량이라 한다.

• 최소값(Minimum value): $X_{(1)} = \min(X_1, X_2, \dots, X_n)$

• 최대값(Maximum value): $X_{(n)} = \max(X_1, X_2, \dots, X_n)$

• 범위(Range): $R = X_{(n)} - X_{(1)}$

• 중앙값(Median): $X_{(n+1)/2}$ (n 이 홀수), $\frac{X_{n/2} + X_{n/2+1}}{2}$ (n 이 짝수)

**EXAMPLE 6.13**

크기 n 인 확률표본 (X_1, X_2, \dots, X_n) 으로부터의 $X_{(n)}$ 과 $X_{(1)}$ 의 확률밀도함수를 구하시오.

우선 $X_{(n)}$ 의 확률밀도함수를 구해보자.

$$F_{X_{(n)}}(x_n) = P(X_{(n)} \leq x_n) = P(X_1 \leq x_n, X_2 \leq x_n, \dots, X_n \leq x_n) = [F_X(x_n)]^n$$

$$\text{그러므로 } f_{X_{(n)}}(x) = F'_{X_{(n)}}(x) = n[F_X(x)]^{n-1} f_X(x)$$

$$\text{같은 방법으로 } F_{X_{(1)}}(x) = P(X_{(1)} \leq x) = 1 - P(X_1 > x, X_2 > x, \dots, X_n > x) = 1 - [1 - F_X(x)]^n$$

$$f_{X_{(1)}}(x) = F'_{X_{(1)}}(x) = n[1 - F_X(x)]^{n-1} f_X(x)$$

**EXAMPLE 6.14**

(X_1, X_2) 은 $f(x)$ 으로부터 확률표본이라 하자. $X_{(1)}$ 와 $X_{(n)}$ 의 결합밀도함수를 구하시오.

$$\begin{aligned} F_{X_{(1)}, X_{(2)}}(x_1, x_2) &= P(X_{(1)} \leq x_1, X_{(2)} \leq x_2) = P((X_1 \leq x_1, X_2 \leq x_2) \cup (X_2 \leq x_1, X_1 \leq x_2)) \\ &= P(X_1 \leq x_1, X_2 \leq x_2) + P(X_2 \leq x_1, X_1 \leq x_2) - P(X_1 \leq x_1, X_2 \leq x_1) \\ &= F_{X_1, X_2}(x_1, x_2) + F_{X_1, X_2}(x_2, x_1) - F_{X_1, X_2}(x_1, x_1) \\ &= F_X(x_1)F_X(x_2) + F_X(x_2)F_X(x_1) - F_X(x_1)F_X(x_1) \\ &= 2F_X(x_1)F_X(x_2) - F_X(x_1)F_X(x_1) \end{aligned}$$

먼저 x_2 에 대해 미분하고 x_1 에 대해 미분하면 결합밀도함수를 얻는다.

$$f_{(1),(2)}(x_1, x_2) = 2f(x_1)f(x_2), x_{(1)} \leq x_{(2)}$$

같은 방법으로 다음 관계를 얻는다.

$$f_{(1),(2), \dots, (n)}(x_1, x_1, \dots, x_n) = n! f(x_1)f(x_2) \dots f(x_n), x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$$

**EXAMPLE 6.15**

성분 수명 확률변수 X 의 확률밀도함수는 $f_X(x) = \frac{1}{100}e^{-x/100}, 0 < x$. 시스템은 두 개의 성분을 직렬 연결하였다. 성분은 서로 독립적으로 작동한다. 시스템 수명의 확률밀도함수를 구하고 평균을 구하시오.

The life-time of the system is $X_{(1)} = \min(X_1, X_2)$.

Since $F_X(x) = 1 - e^{-x/100}$, $f_{(1)}(x_1) = \frac{1}{50}e^{-x_1/50}, 0 < x_{(1)}$. $E(X_{(1)}) = 50$.

**HOMEWORK #18-1**

성분 수명 확률변수 X 의 확률밀도함수는 $f_X(x) = \frac{1}{100}e^{-x/100}, 0 < x$. 시스템은 두 개의 성분을 병렬 연결하였다. 성분은 서로 독립적으로 작동한다. 시스템 수명의 확률밀도함수를 구하고 평균을 구하시오.

THEOREM (정리)

크기 n 인 확률표본으로부터 순서통계량 $X_{(k)}, 1 \leq k \leq n$ 의 확률밀도함수는

$$f_{(k)}(x_k) = \frac{n!}{(k-1)!(n-k)!} [F(x_k)]^{k-1} [1 - F(x_k)]^{n-k} f(x_k), -\infty < x_{(k)} < \infty$$

PROOF $P(x \leq X_{(k)} \leq x + \Delta x) \approx f_{(k)}(x_k) \Delta x$ 을 이용하자.

$$P(x \leq X_{(k)} \leq x + \Delta x) \approx P[(k-1) \text{ from } < x_{(k)}, 1 \text{ near } x_{(k)}, (n-k) \text{ from } > x_{(k)}]$$

$$P(x \leq X_{(k)} \leq x + \Delta x) \approx \binom{n}{k-1, 1, n-k} [F(x_k)]^{k-1} f(x_k) [1 - F(x_k)]^{n-k} \quad \boxed{\text{Q.E.D.}}$$

같은 방법으로 $X_{(i)}$ and $X_{(j)}$ for $i < j$ 의 결합밀도함수는

$$f_{(i),(j)}(x_i, x_j) = \frac{n!}{(i-1)!(j-i-1)!(n-k)!} [F(x_i)]^{i-1} [F(x_j) - F(x_i)]^{j-i-1} [1 - F(x_k)]^{n-k} f(x_i) f(x_j)$$

**EXAMPLE 2.16**

(X_1, X_2, \dots, X_n) 은 $f_X(x) = 1, 0 \leq x \leq 1$ 으로부터의 확률표본이다.

순서통계량 $X_{(2)}$ 의 주변밀도함수로 구하시오. 순서통계량 $X_{(2)}$ and $X_{(4)}$ 의 결합밀도함수를 구하시오.

Since $F_X(x) = x$,

$$f_{(2)}(x_2) = 20x_2(1-x_2), 0 \leq x_{(2)} \leq 1.$$

$$f_{(2),(4)}(x_2, x_4) = 5!x_2(x_4-x_2)(1-x_2), 0 \leq x_{(2)} \leq x_{(4)} \leq 1$$

**HOMEWORK #18-2**

Suppose that (X_1, X_2, \dots, X_n) is a random sample from $f_X(x) = 1, 0 \leq x \leq 1$.

Find the pdf of $R = X_{(n)} - X_{(1)}$.

**HOMEWORK #18-3**

Suppose that (X_1, X_2, \dots, X_n) is a random sample from $f_X(x) = \frac{1}{\beta} e^{-x/\beta}, 0 < x$.

① Find the pdf of $X_{(1)}$. It will be $\exp(\beta/n)$.

② If $n=5$ and $\beta=2$, find $P(X_{(1)} \leq 3.6)$.

6.5 이산형/연속형 확률밀도함수 (PDF of discrete and continuous random variables)**6.5.1 이산형**

① Bernoulli Experiment(베르누이 실험)

- 실험이 서로 독립이다. (결과에 영향을 미치지 않는다)
- 실험의 결과는 성공과 실패(binary) 두 개만 있다.
- 각 실험의 성공 확률은 p 로 일정하다.

② $X \sim \text{Bernoulli}(p)$ 성공확률 p 인 베르누이(Bernoulli) 분포 (평균, 분산) = (p, pq)

③ $X \sim \text{Binomial}(n, p)$ 모수 (n, p) 인 이항(Binomial) 분포 (평균, 분산) = (np, npq)

- 베르누이 실험을 n 번 했을 때 성공의 회수 $X = \sum X_i$
- 베르누이 분포를 따르는 크기 n 인 확률표본의 합은 이항분포를 따른다.

④ $X \sim Geometric(p)$ 모수 p 인 기하(Geometric)분포 (평균, 분산) $= (1/p, q/p^2)$

- 베르누이 실험을 n 번 했을 때 첫번째 성공까지의 실험 회수 X

⑤ $X \sim NB(r, p)$ 모수 (r, p) 인 음이항(Negative Binomial)분포 (평균, 분산) $= (r/p, rq/p^2)$

- 베르누이 실험을 n 번 했을 때 r 번째 성공까지의 실험 회수 X
- 기하분포를 따르는 표본 크기 r 인 확률표본의 합은 음이항분포를 따른다.

⑥ $X \sim Hypergeometric(N, M, n)$ 모수 (N, M, K) 인 초기하(Hyper-Geometric)분포

•(평균, 분산) $= (n \frac{M}{N}, \frac{nM}{N} \frac{(N-M)(n-n)}{N(N-1)})$

- 전체 N (2부분 $M, N-M$ 나누어져 있음)개로부터 n 개 뽑을 때 M 부분에서 뽑힌 개수

- 만약 N 이 무한대로 커지면 초기하 분포는 $Binomial(n, p = \frac{M}{N})$ 분포에 근사한다.

⑦ $X \sim Poisson(\lambda)$ 평균 λ 인 포아송(Poisson)분포, 분산 $= \lambda$

- 일정 단위(시간, 공간)에 발생하는 사건(event)의 수를 의미한다.
- 발생한 사건이 하나 이상 들어가지 않게 단위를 나누자. 그러면 그 구간 안에는 사건이 발생하지 않거나 단 하나 발생하게 된다. 그러면 각 구간의 베르누이 시행이 된다. 구간의 수를 n 이라 하면 사건 발생 회수는 이항분포를 따른다.

• $\lim_{n \rightarrow \infty} \binom{n}{x} p^x q^{n-x}, np = \lambda$ 이라 놓고 $\lim_{n \rightarrow \infty} (1 - \frac{\lambda}{n})^n = e^{-\lambda}$ 사실을 이용.



EXAMPLE 6.17

$Bernoulli(p)$ 으로부터 얻은 표본 크기 n 인 확률표본의 합은 $Binomial(n, p)$ 을 따른다.



EXAMPLE 6.18

$X_1 \sim Binomial(n_1, p)$ (모수 (n_1, p) 인 이항분포), $X_2 \sim Binomial(n_2, p)$ 이고 서로 독립인 경우 $X_1 + X_2$ 도 이항분포($n_1 + n_2, p$)임을 보이시오.



EXAMPLE 6.19

$X_i \sim Geometric(p)$ (모수 p 인 기하분포)인 경우 확률표본 합의 분포는 모수 (n, p) 인 음이항(Negative binomial)분포($NB(n, p)$)임을 보이시오.



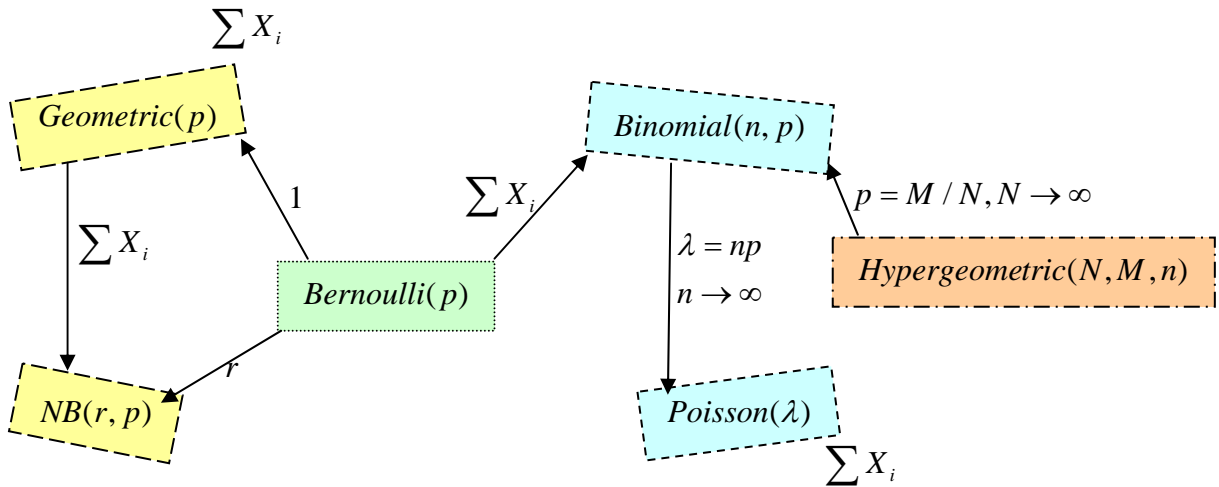
EXAMPLE 6.20

$X_i \sim \text{Poisson}(\lambda_i)$ 이면 확률표본 합 $\sum X_i$ 이 $\text{Poisson}(\sum \lambda_i)$ 임을 보이시오. ⊙ 포아송 분포 가법성



EXAMPLE 6.21

$X_i \sim \text{Geometric}(p)$ 일 경우 $P(X > a + b | X > a) = P(X > b)$ 임을 보이시오. ⊙ Memoryless property



6.5.2 Relation between discrete and continuous

이항분포의 정규분포 근사

- 표본의 크기가 커지면 $X \sim \text{Binomial}(n, p)$ 는 $\text{Normal}(np, npq)$ 에 근사한다.

포아송 분포와 지수분포의 관계

W 를 모수 λ 인 포아송 분포를 따르는 사건인 정확하게 1번 발생할 때까지 걸리는 시간이라면

$$F(w) = P(W \leq w) = 1 - P(W > w) = 1 - P(X = 0 | X \sim \text{Poisson}(\lambda w)) = 1 - \frac{(\lambda w)^0 e^{-\lambda w}}{0!} = 1 - e^{-\lambda w} \quad \text{이다.}$$

이를 미분하면 $f(w) = \lambda e^{-\lambda w} \sim \text{Exponential}(\beta = 1/\lambda)$ 이다.

6.5.3 Continuous

① $X \sim \text{Normal}(\mu, \sigma^2)$ 평균 μ , 분산 σ^2 인 정규(normal)분포

- 일반적으로 측정 오차, 측정형 변수에 대한 모집단 확률분포함수로 가정

• $\mu = 0, \sigma^2 = 1$ 인 경우 이를 표준정규분포(standard normal)이라 한다.

② $X \sim Exp(\beta)$ 평균 β 인 지수(exponential)분포, 분산 = β^2

• 평균 λ 인 포아송 분포를 따르는 사건 사이의 시간, $\lambda = \frac{1}{\beta}$

• 일반적으로 시간에 대한 분포로 사용하나 memoryless property 때문에 수명에 대해서는 자주 사용하지 않는다.



EXAMPLE 6.22

$X \sim Exp(\beta)$ 일 경우 $P(X > a + b | X > a) = P(X > b)$ 임을 보이시오. ⊙ Memoryless property

③ $X \sim Gamma(\alpha, \beta)$ 모수 ($\alpha = shape, \beta = scale$) 인 감마(gamma)분포

• (평균, 분산) = $(\alpha\beta, \alpha\beta^2)$

• 평균(수명) β 인 지수분포로부터의 표본 크기 n 인 확률표본 합(전체 수명)의 분포는 모수 (α, β) 감마분포를 따른다.

• $X \sim Gamma(\alpha = \lambda/2, \beta = 2)$ 은 $\chi^2(\lambda)$ 분포를 따른다.



EXAMPLE 6.23

$X_i \sim (iid)Exp(\beta)$ 일 경우 $\sum_{i=1}^n X_i$ 의 확률분포함수를 구하시오.

④ $X \sim Beta(\alpha, \beta)$ 모수 (α, β) 인 베타(beta)분포, (평균, 분산) = $(\frac{\alpha}{\alpha + \beta}, \frac{\alpha\beta}{(\alpha + \beta)(\alpha + \beta + 1)})$

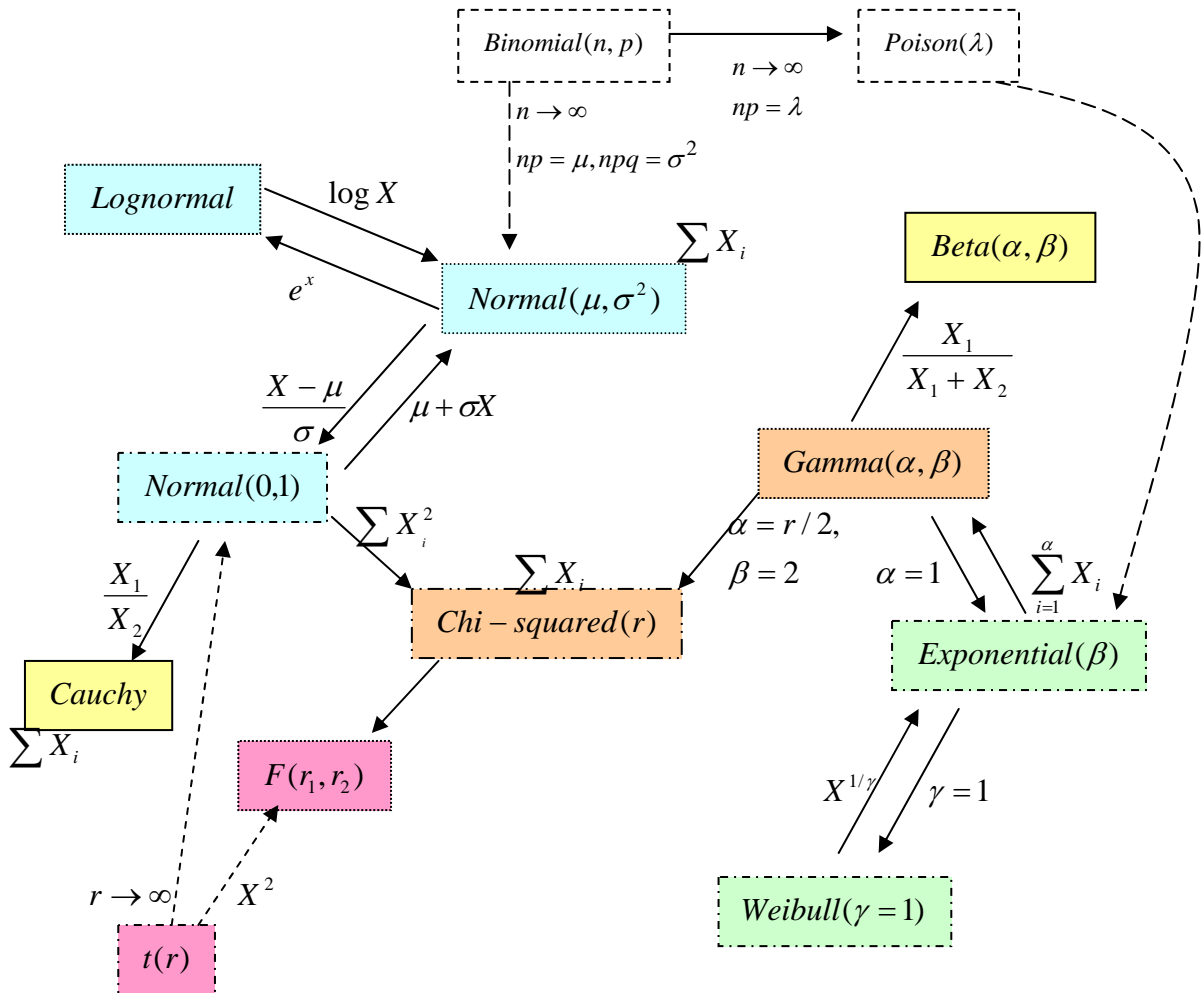
X_1, X_2 서로독립이고 $Gamma(\alpha, \beta)$ 분포를 따를 경우 $\frac{X_1}{X_1 + X_2}$ (수명이 동일한 2개의 제품 중 하나의 수명 비율)는 $Beta(\alpha, \beta)$ 을 따른다.

⑤ $X \sim t(r)$ 모수(자유도) r 인 Student t-분포, (평균, 분산) = $(0, \frac{r}{r-2})$

$W \sim N(0,1), V \sim \chi^2(r)$ 이고 서로 독립일 때 $\frac{W}{\sqrt{V/r}} \sim t(r)$ 이다.

⑥ $X \sim F(r_1, r_2)$ 자유도 r_1, r_2 인 F-분포, (평균, 분산) = $(\frac{r_2}{r_2 - 2}, complicated)$

$W \sim \chi^2(r_1), V \sim \chi^2(r_2)$ 이고 서로 독립일 때 $\frac{W/r_1}{V/r_2} \sim F(r_1, r_2)$ 이다.



EXAMPLE 6.24

$X_i \sim Gamma(\alpha_i, \beta)$ (모수 (α, β) 인 감마분포)인 경우 확률표본 합 $\sum X_i$ 이 감마분포 $Gamma(\sum \alpha_i, \beta)$ 임을 보이시오. ⊙ 감마분포의 가법성



EXAMPLE 6.25

$X_i \sim \chi^2(\lambda_i)$ (모수 λ_i 인 카이제곱 분포)인 경우 확률표본 합 $\sum X_i$ 도 카이제곱 분포 $\chi^2(\sum \lambda_i)$ 임을 보이시오. ⊙ 카이제곱 분포의 가법성