

## 5장 확률 및 확률분포함수

### 5.1 확률의 정의

- 우리 주변에는 동전을 던진다든지 제품을 검사한다든지 유사한 일이 반복되는 경우가 많다. 이러한 일들의 가능한 결과는 ‘겉면’ 또는 ‘뒷면’이든지 ‘우량품’ 또는 ‘불량품’이 되는 것은 알지만 무슨 결과나 될지는 모른다. 이와 같이 비슷한 일이 반복되고 모든 가능한 결과들은 알지만 한 결과가 우연에 의하여 결정되는 실험이나 관찰을 **통계적 실험**이라 한다.
- 한 통계적 실험에서 모든 가능한 사건의 집합을 **표본공간(sample space)**이라 하고 이 표본공간의 한 부분집합을 **사건(event)**이라 한다. 표본공간의 원소수가 유한개 또는 셀수 있을 때 **이산형 표본공간**이라 하고, 원소수가 무한개일 때 **연속형 표본공간**이라 한다.
- **확률(probability)**이란 ‘한 사건이 일어날 가능성을 0 과 1 사이의 실수로 표시’하는 것인데, 한 사건이 발생할 가능성이 높으면 확률은 1 에 가까운 수로 표시하고, 반대로 발생할 가능성이 적으면 확률을 0 에 가까운 수로 표시한다. 어느 사건이 반드시 발생하면 확률은 1 로 정하고, 전혀 발생하지 않으면 확률은 0 으로 정한다. 구체적으로 ‘어느 사건의 확률을 0과 1사이의 어떤 값으로 정하느냐’하는 방법에는 여러 가지가 있는데, 여기서는 확률의 고전적 정의와 상대도수를 이용한 정의 두 가지를 소개한다.
- 확률의 고전적 정의는 표본공간의 모든 원소가 일어날 가능성이 같다고 하고 사건 A가 발생할 확률(P(A)로 표시)은 이산형 표본공간의 경우에 다음과 같이 정의한다.

$$P(A) = \frac{\text{사건 A 에 속하는 원소의 개수}}{\text{표본공간의 전체 원소의 개수}}$$

- 연속형 표본공간의 경우에 확률 P(A)는 다음과 같이 정의한다.

$$P(A) = \frac{\text{사건 A 에 속하는 원소에 대한 측도}}{\text{표본공간의 전체원소에 대한 측도}}$$

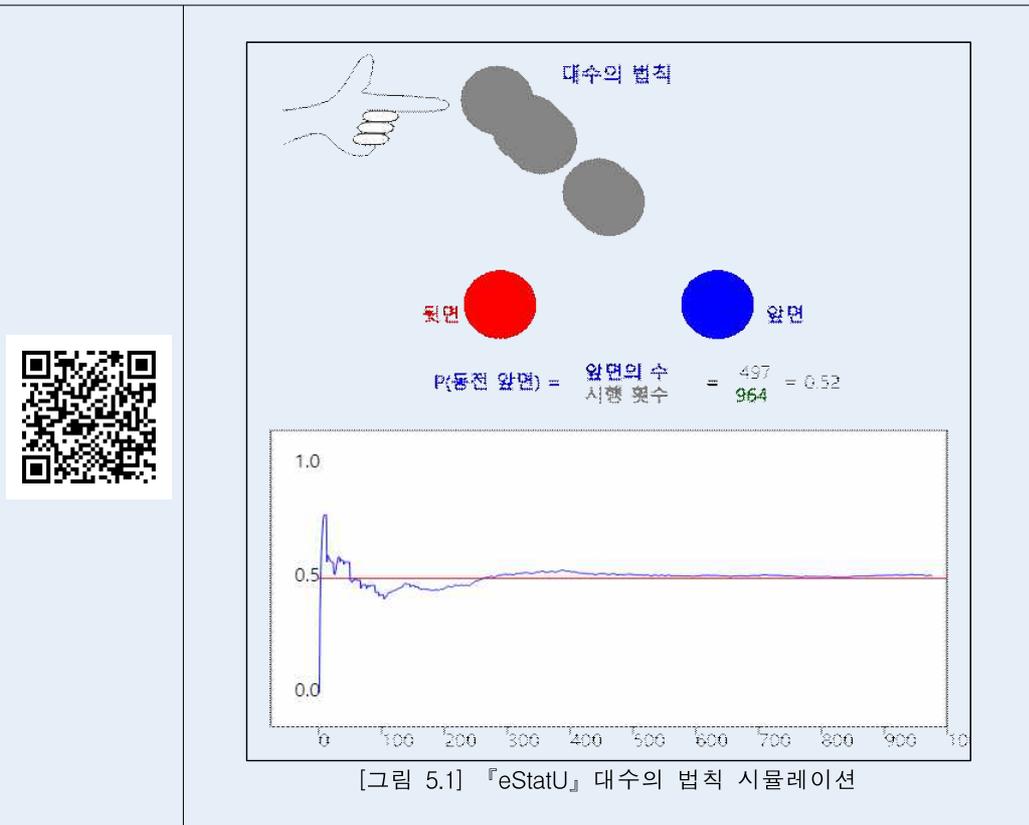
여기서 측도란 길이, 면적, 부피 등을 뜻한다.

- **확률의 고전적 정의**는 많은 현실문제 확률계산에 큰 문제점이 없다. 그러나 고전적 정의에서 표본공간의 모든 원소가 발생할 가능성이 같다는 가정은 성립이 안되는 경우가 있을 수 있다. 예를 들면, 공장에서 한 제품을 생산할 때 ‘우량품’과 ‘불량품’으로 표시한다면 표본공간은 {우량품, 불량품}이다. 하지만 우량품이 나올 가능성과 불량품이 나올 가능성은 같지 않아 표본공간의 각 원소가 발생할 가능성이 같다는 가정은 맞지 않는다. 이러한 문제를 해결하기 위한 것이 확률의

상대도수적 정의이다.

- **확률의 상대도수적 정의**는 사건 A가 발생할 확률(P(A)로 표시)을 같은 조건하에서 통계적 실험을 수없이 많이 반복시행하였을 때 사건 A가 발생하는 비율 즉, 상대도수이다.

[예 5.1] 『eStatU』 메뉴에서 ‘통계적 확률’을 선택하면 [그림 5.1]과 같은 시뮬레이션 창이 나타난다. 여기에서 동전을 던지는 횟수  $n$ 을 입력하고 [실행] 버튼을 누르면 동전의 걸이 나오는 확률이 동전을 많이 던졌을 때 1/2에 수렴하는 것을 대수의 법칙(law of large number)이라한다.



## 5.2 확률의 계산

### 순열과 조합

- $n$ 개의 사물 중  $r$ 개를 선택해 순서를 고려해 나열하는 방법의 수를 순열(permutation)이라 하고 다음과 같이 계산한다.

$${}_n P_r = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!}$$

그러므로  $n$ 개를 모두 나열하는 방법의 수는 다음과 같다.

$${}_n P_n = n(n-1)(n-2)\cdots 2 \cdot 1 = n!$$

참고:  $0! = 1$  로 정의한다.

- n개의 사물 중 r개를 순서를 고려치 않고 추출하는 방법의 수를 조합(combination)이라 하고 다음과 같이 계산한다.

$${}_n C_r = \frac{{}_n P_r}{r!} = \frac{n!}{r!(n-r)!}$$

[예 5.2] 『eStatU』 메뉴에서 ‘순열 조합’을 선택하면 [그림 5.2]와 같은 창이 나타난다. 여기에서 n과 r을 입력하면 여러 종류의 순열과 조합이 계산된다. n이 10보다 작고 r이 2인 경우에 [실행] 버튼을 누르면 모든 경우의 수에 대한 그림을 보여준다.

**순열 - 조합** 메뉴

실행  n =  1  30 r =  0  30

**[순열]**

n! =  r! =

${}_n P_r =$   (비중복)  $r \leq n$

${}_n \Pi_r =$   (중복)

**[조합]**

${}_n C_r =$   (비중복)  $r \leq n$

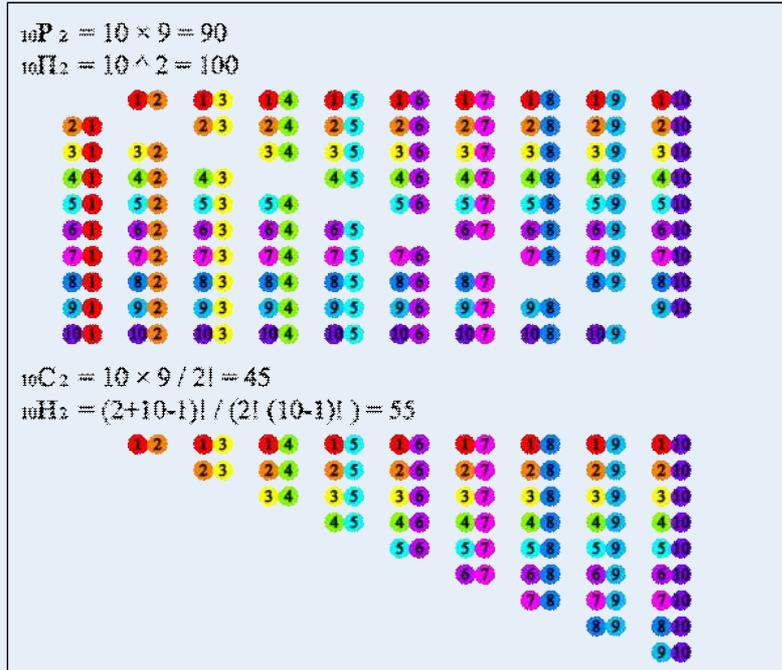
${}_n H_r =$   (중복)  ${}_{n+r-1} C_r$

**[같은 것이 있는 순열]**

n =  = p + q + r + s p =  q =  r =  s =

실행   $\frac{n!}{p! q! r! s!} =$

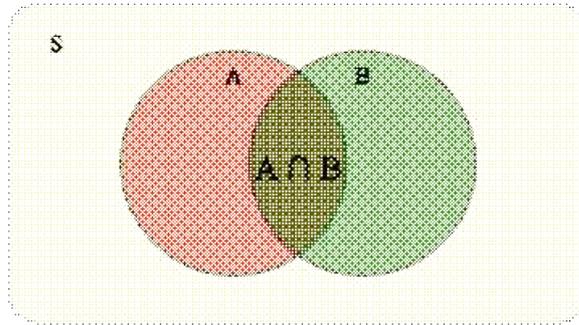
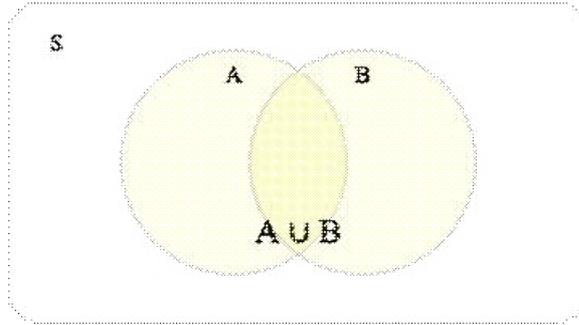
[그림 5.2] 순열과 조합의 계산



[그림 5.3] 10명중 2명을 뽑는 순열과 조합 경우의 수

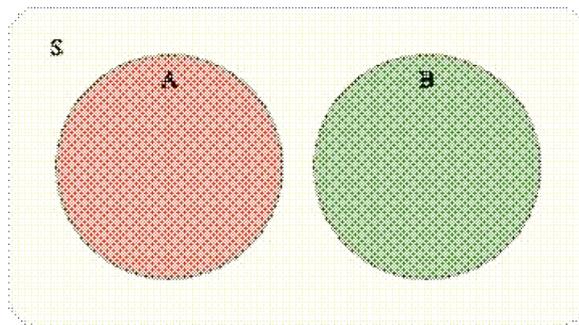
**확률의 덧셈법칙**

- 표본공간의 두 사건 A와 B에 대하여 A 또는 B가 일어나는 사건을  $(A \cup B)$ 로 나타내고, A와 B가 동시에 일어나는 사건을  $(A \cap B)$ 와 같이 나타낸다.



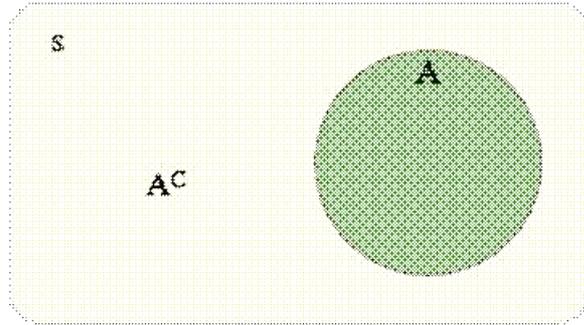
[그림 5.4] 사건  $(A \cup B)$ 와 사건  $(A \cap B)$

- 특히 사건 A와 B가 동시에 일어나지 않을 때, 즉  $A \cap B = \emptyset$  일 때, 이 두 사건을 서로 **배반사건**이라고 한다.



[그림 5.5] 서로 배반인 사건

- 어떤 사건 A에 대하여 사건 A가 일어나지 않는 사건을 A의 **여사건**이라고 하고, 이 사건을  $A^C$ 로 나타낸다. A와  $A^C$ 는 동시에 일어날 수 없으므로  $A \cap A^C = \emptyset$  이고 사건 A와  $A^C$ 는 서로 배반사건이다.



[그림 5.6] 사건 A의 여사건

- 확률의 덧셈법칙은 다음과 같다..

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

만일  $A \cap B = \emptyset$  일 때, 즉 A와 B가 서로 배반사건이면 확률은 다음과 같다.

$$P(A \cup B) = P(A) + P(B)$$

[예 5.3] 「eStatU」 메뉴에서 ‘확률의 덧셈정리’를 선택하면 [그림 5.7]과 같은 창이 나타난다. 여기에서  $P(A) = 0.6$ ,  $P(B) = 0.4$ ,  $P(A \cap B) = 0.2$ 를 입력한 후 [실행]버튼을 클릭하면 확률의 덧셈정리에 대한 그래프를 관찰 할 수 있다.  $P(A)$ ,  $P(B)$ ,  $P(A \cap B)$ 를 변화시키면서 덧셈 정리를 살펴볼 수 있다.

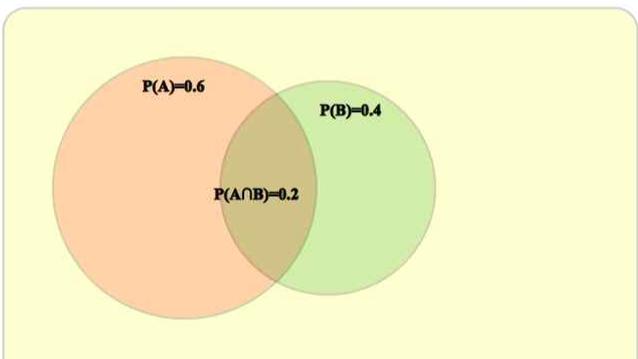


### 확률의 덧셈정리 메뉴

$P(A) = 0.60$  0  1  
 $P(B) = 0.40$  0  1  
 $P(A \cap B) = 0.20$  0.00  0.40

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.6 + 0.4 - 0.2 = 0.800$

실행



$P(A \cup B) = P(A) + P(B) - P(A \cap B)$   
 $= 0.6 + 0.4 - 0.2$   
 $= 0.800$

[그림 5.7] 확률의 덧셈정리

**확률의 곱셈법칙**

- 한 고등학교 학생 40명의 안경착용 여부를 조사하니 다음 표와 같다.

	안경착용( $B$ )	안경착용 안함( $N$ )	합계
남학생( $A$ )	$n(A \cap B) = 8$	$n(A \cap N) = 16$	$n(A) = 24$
여학생( $F$ )	$n(F \cap B) = 4$	$n(F \cap N) = 12$	$n(F) = 16$
합계	$n(B) = 12$	$n(N) = 28$	$n(S) = 40$

- 남학생 중에서 한 명을 임의로 뽑았을 때 이 학생이 안경을 착용했을 확률은 기호로  $P(B|A)$ 로 표시하며 남학생 사건  $A$ 가 일어났을 때 안경착용 사건  $B$ 의 **조건부 확률**이라 한다.

$$P(B|A) = \frac{n(A \cap B)}{n(A)} = \frac{8}{24}$$

- 일반적으로 사건  $A$ 가 일어났을 때 사건  $B$ 의 조건부 확률은 다음과 같다.

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

- 위의 식에서  $P(A)$ 를 양 변에 곱하면 다음과 같은 **확률의 곱셈법칙**이 된다.

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

만일 두 사건  $A$ 와  $B$ 에 대하여 사건  $A$ 가 일어나는 것이 사건  $B$ 가 일어날 확률에 영향을 주지 않을 때, 즉  $P(B|A) = P(B)$ 이면 사건  $A$ 와  $B$ 를 서로 **독립사건** (independent event) 이라고 한다. 이때는

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

로 쓸 수 있다.

[예 5.4] 『eStatU』 메뉴에서 ‘조건부확률’을 선택하면 [그림 5.8]과 같은 창이 나타난다. 여기에서 결합확률  $P(A_1 \cap B_1)$  를 조정하면 조건부확률을 관찰할 수 있고 그 밑에 [그림 5.9]와같은 선그래프와 각각의 행에 대한 조건부확률의 막대그래프를 관찰할 수 있다.

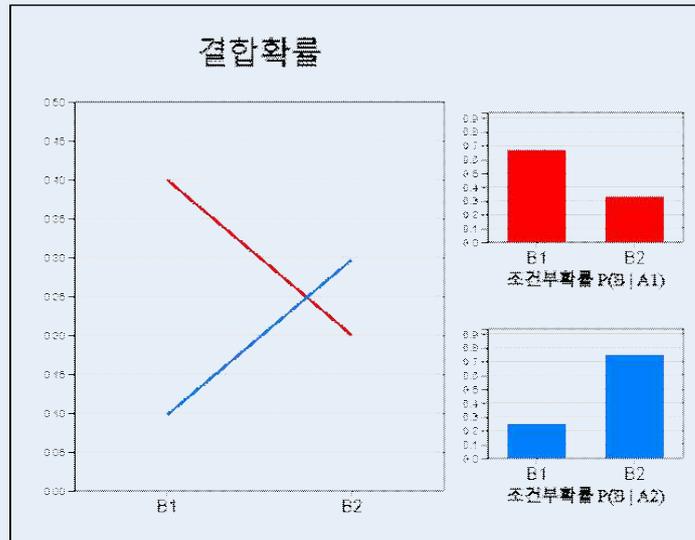
조건부확률
메뉴

결합확률	$B_1$	$B_2$	합계
$A_1$	$P(A_1 \cap B_1) = $ <input type="text" value="0.40"/> <small>0 ————— 0.50</small>	$P(A_1 \cap B_2) = 0.20$	$P(A_1) = $ <input type="text" value="0.60"/> <small>0 ————— 1</small>
$A_2$	$P(A_2 \cap B_1) = 0.10$	$P(A_2 \cap B_2) = 0.30$	$P(A_2) = 0.40$
합계	$P(B_1) = $ <input type="text" value="0.50"/> <small>0 ————— 1</small>	$P(B_2) = 0.50$	1.00

조건부확률	$B_1$	$B_2$	합계
$A_1$	$P(B_1 A_1) = \frac{P(A_1 \cap B_1)}{P(A_1)} = 0.67$	$P(B_2 A_1) = \frac{P(A_1 \cap B_2)}{P(A_1)} = 0.33$	1.00
$A_2$	$P(B_1 A_2) = \frac{P(A_2 \cap B_1)}{P(A_2)} = 0.25$	$P(B_2 A_2) = \frac{P(A_2 \cap B_2)}{P(A_2)} = 0.75$	1.00

[그림 5.8] 조건부확률



[그림 5.9] 결합확률에 대한 그래프와 조건부확률

### 5.3 이산형 확률변량

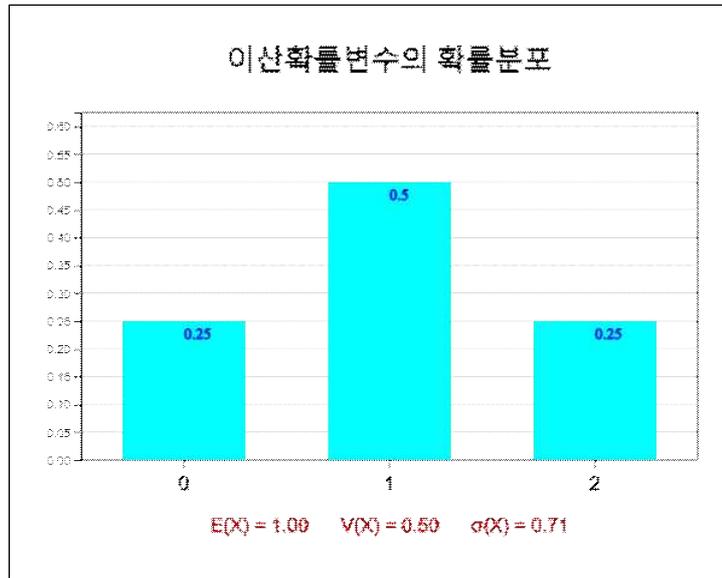
- 동전을 두 번 던졌을 때 표본공간의 여러 가지 사건에 대해 확률을 구할 수 있지만 그 중에서도 우리가 관심이 있는 것은 앞면이 나오는 횟수에 대한 사건들과 그 확률이다.  $X$ 를 ‘동전의 앞면이 나온 횟수’라고 하면  $X$ 의 값으로 가능한 수는 0, 1, 2가 된다. 즉  $X$ 는 표본공간  $S$ 의 각 원소에 다음과 같이 하나의 수를 대응한 것이다.

표본공간	$X$ 를 동전의 앞면이 나온 횟수
{TT}	0
{TH} {HT}	1
{HH}	2

- 이와 같이 어떤 시행에서 표본공간의 각 원소에 단 하나의 실수를 대응시킨 관계를 확률변수라고 한다. 확률변수는 대개 알파벳 대문자  $X, Y, Z$  등으로 표시하고, 확률변수의 값은 소문자  $x, y, z$  등으로 표시한다. 확률변수가 가질 수 있는 값이 유한개이거나 자연수와 같이 셀 수 있을 때 그 확률변수를 이산확률변수라고 하고, 임의의 실수 값을 가질 때 그 확률변수를 연속확률변수라고 한다. 예를 들어 동전을 두 개 던져 나타나는 겉면의 수  $X$ 는 이산확률변수이고, 피자를 주문해서 집에 도착할 때까지 걸리는 시간을  $Y$ 라 하면  $Y$ 는 임의의 양의 실수 값을 가질 수 있으므로 연속확률변수가 된다.
- 확률변수  $X$ 를 동전을 두 번 던졌을 때 나오는 앞면이 나온 횟수라고 할 때 각 값이 나오는 확률  $P(X=x)$ 은 다음 표와 같이 정리할 수 있다.

표본공간	$X =$ 동전의 앞면이 나온 횟수	$P(X=x)$
{TT}	0	
{TH} {HT}	1	
{HH}	2	

이 표와 같이 이산확률변수  $X$ 가 어느 한 값  $x$ 를 가질 확률  $P(X=x)$ 을 모두 정리한 것을 **확률분포함수**라고 한다. 위의 확률분포함수 [그림 5.10]과 같이 그래프로 나타낼 수 있다.



[그림 5.10] 이산형 확률분포함수

- 일반적으로 이산확률변수  $X$ 가 가질 수 있는 값이  $x_1, x_2, \dots, x_n$ 이고 각각의 값을 가질 확률이  $p_1, p_2, \dots, p_n$ 인 확률분포가 다음 표와 같다고 하자.

$X$	$x_1$	$x_2$	$\dots$	$x_n$	합계
$P(X = x)$	$p_1$	$p_2$	$\dots$	$p_n$	<b>1</b>

이 확률변수의 평균 또는 기댓값  $E(X) = \mu$ , 분산  $V(X) = \sigma^2$ , 표준편차 표준편차  $\sigma(X)$ 는 다음과 같이 구한다.

$$E(X) = \mu = \sum_{i=1}^n x_i p_i$$

$$V(X) = \sigma^2 = \sum_{i=1}^n (x_i - \mu)^2 p_i$$

$$\sigma(X) = \sqrt{V(X)}$$

- 확률변수  $X$ 에 대한 확률분포의 평균  $E(X)$ 는 모든 가능한 값  $x_1, x_2, \dots, x_n$ 에 대하여 확률  $p_1, p_2, \dots, p_n$ 를 가중값으로 한 무게중심으로 이해 할 수 있다. 그리고 분산  $V(X)$ 은 이 평균으로부터의 제곱거리에 대한 평균, 표준편차  $\sigma(X)$ 는 가능한 값들의 평균으로부터의 거리를 의미한다.
- 일반적으로 확률변수  $X$ 의 평균, 분산, 표준편차가 각각  $E(X) = \mu$ ,  $V(X) = \sigma^2$ ,  $\sigma(X) = \sigma$ 일 때  $Y = aX + b$ 의 평균, 분산, 표준편차는 다음과 같다.

$$E(Y) = aE(X) + b = a\mu + b$$

$$V(Y) = a^2 V(X) = a^2 \sigma^2$$

$$\sigma(Y) = |a| \sigma(X) = |a| \sigma$$

이와 같은  $aX + b$ 의 평균, 분산, 표준편차는  $X$ 가 이산확률변수이거나 연속확률변

수이거나 상관없이 성립한다.

[예 5.5] 『eStatU』 메뉴에서 '이산확률변량의 분포'를 선택하면 [그림 5.11]과 같은 자료입력 창이 나타난다. 여기에 확률분포를 그림과 같이 입력하고 [실행]버튼을 누르면 [그림 5.12]와 같은 확률분포 그래프가 나타난다. 확률분포의 평균과 분산 및 표준편차도 계산하여 준다.

**이산확률변량의 분포** 메뉴

	x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	x <sub>4</sub>	x <sub>5</sub>	x <sub>6</sub>	x <sub>7</sub>	x <sub>8</sub>	x <sub>9</sub>	합계
X	0	1	2	3	4					
P(X = x <sub>i</sub> )	0.1	0.2	0.3	0.3	0.1					1.000

평균 E(X) = 2.100
분산 V(X) = 1.290
표준편차 σ(X) = 1.136

[그림 5.11] 이산확률변량의 분포 자료입력



[그림 5.12] 『eStatU』를 이용한 이산확률분포 그래프

- 이산확률분포 중에서 현실에서 많이 이용되는 것이 이항분포이고, 그밖에 포아송 분포, 기하분포, 초기하분포 등이 있다.

### 이항분포

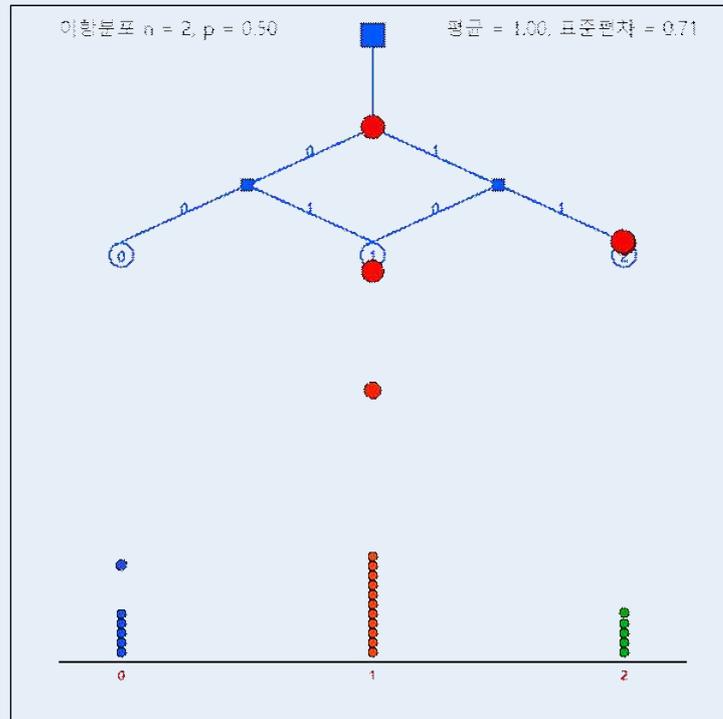
- '성공'인 사건의 확률이  $p$ 인 베르누이 실험을  $n$ 번 독립적으로 반복 시행하였을 때 확률변수  $X$ 를 '성공의 회수(X)'라 하면  $X$ 가  $x$ 일 확률분포는 다음과 같은 이항분포로  $B(n, p)$ 이다.

$$f(x) = {}_n C_x p^x (1-p)^{n-x} \quad , \quad (\text{단 } x = 0, 1, 2, \dots, n)$$

이항분포의 평균은  $E(X) = np$ 이고, 분산은  $V(X) = np(1-p)$ 이다.

[예 5.6] 과학관 등에 가면 공을 위에서 떨어뜨려 한 곳에 부딪치면 1/2 확률로 좌측(0점) 또는 우측(1점)으로 떨어지게 하는 기구가 있다. 떨어진 공은 다시 1/2 확률로 좌측 우측으로 떨어진다. 100개의 공을 떨어뜨렸을 때 전체 점수의 합계를 조사해본다.

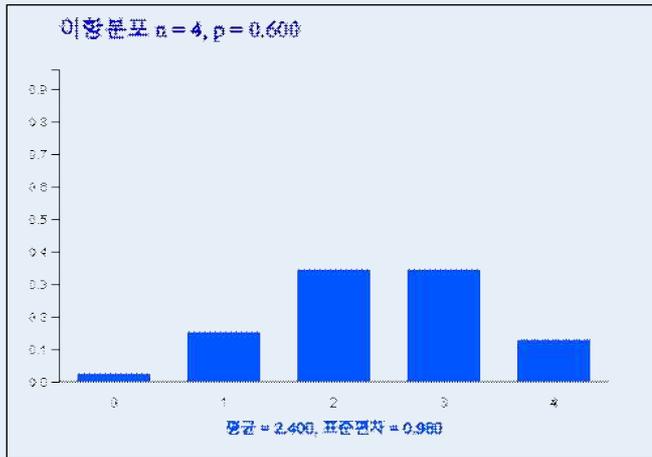
『eStatU』 메뉴에서 '이항분포실험'을 선택하면 [그림 5.13]과 같은 창이 나타난다. 여기에 몇 번 공이 부딪히게 하는 지  $n$ 과 우측으로 갈 확률  $p$ 를 입력하고 [실행]버튼을 누르면 공이 떨어지며 점수 합계가 이항분포가 되는 실험을 할 수 있다.



[그림 5.13] 『eStatU』의 이항분포 실험 시뮬레이션

[예 5.7] 프로야구팀 ‘호랑이’가 올해 시즌에 ‘곰’팀과 앞으로 더 치러야 할 게임 수는 네 게임이다. 호랑이팀이 매 게임 승리할 확률이 60%이다. 확률변수  $X =$  ‘호랑이가 승리하는 게임수’는 이항분포이다. 『eStatU』를 이용하여 이 확률분포를 구해 보자.

『eStatU』의 주메뉴에서 ‘이항분포’를 선택하고  $n = 4, p = 0.6$ 을 입력하고 [실행] 버튼을 누르면 [그림 5.14]와 같은 이항분포 그래프가 나타난다.



[그림 5.14]  $n = 4, p = 0.6$  이항분포 그래프

그래프 밑에는  $n$ 과  $p$  슬라이드바와 확률계산 박스가 있어 원하는 값을 넣고 [Enter] 키를 누르면 값이 계산된다.



실행  $n = 4$    $p = 0.5$

확률표시     정규분포근사

확률  $P(\text{ } \leq X \leq \text{ }) = \text{ }$   
 $P(X \leq \text{ }) = \text{ }$   
 $P(X \geq \text{ }) = \text{ }$

그래프 저장    참고사이트 : [Wikipedia](#) [Wolfram](#) [StatTrek](#)

[그림 5.15]  $n$ 과  $p$  슬라이드바와 확률계산 박스

그래프 오른쪽에는 이항분포에 대한 확률밀도함수 표가 나타난다. 이 표에는  $P(X = x)$  이외에도 누적확률  $P(X \leq x)$ 와  $P(X \geq x)$ 를 같이 보여 주어 여러 가지 확률계산을 쉽게 할 수 있다. 새로운  $n$ 과  $p$ 를 선택하고 [실행] 버튼을 누르면 이 값에 대한 이항분포표가 밑에 추가된다.

이항분포	$n = 4$	$p = 0.600$	
$x$	$P(X = x)$	$P(X \leq x)$	$P(X \geq x)$
0	0.0256	0.0256	1.0000
1	0.1536	0.1792	0.9744
2	0.3456	0.5248	0.8208
3	0.3456	0.8704	0.4752
4	0.1296	1.0000	0.1296

[그림 5.16] 이항분포표

**포아송분포**

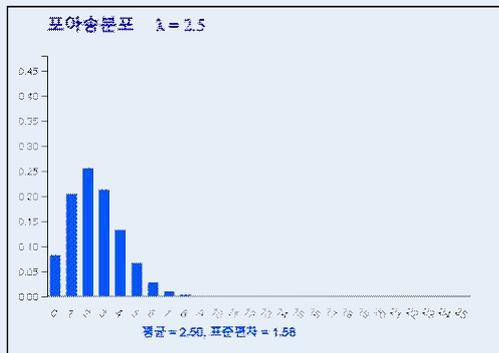
- 단위시간 또는 단위면적당 발생하는 한 사건(‘전화가 걸려옴’, ‘교통사고가 발생’, ‘기계가 고장’)의 수를 나타내는 확률변량을 **포아송 확률변량**(Poisson random variable)이라고 하며, 그 분포를 **포아송분포**(Poisson distribution) 라고 한다.

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

여기서  $\lambda$ 는 사건의 평균 발생횟수이다.  
포아송분포의 평균은  $E(X) = \lambda$ , 분산은  $V(X) = \lambda$  이다.

**[예 5.8]** 우리나라 남부지역에 한 해 동안 태풍이 지나가는 수는 평균  $\lambda=2.5$  회인 포아송 분포를 한다고 하자. 『eStatU』를 이용하여 포아송분포를 구해 보자.

『eStatU』의 주메뉴에서 ‘포아송분포’를 선택하고  $\lambda = 2.5$ 를 입력한 후 [실행] 버튼을 누르면 [그림 5.17]과 같은 포아송분포가 나타나고 그 오른쪽에 포아송분포표가 나타난다. 이 표에는  $P(X = x)$  이외에도 누적확률  $P(X \leq x)$ 와  $P(X \geq x)$ 를 같이 보여주어 여러 가지 확률계산을 쉽게 할 수 있다.



[그림 5.17]  $\lambda = 2.5$ 인 포아송분포

포아송분포	$\lambda = 2.5$		
x	P(X = x)	P(X ≤ x)	P(X ≥ x)
0	0.0821	0.0821	1.0000
1	0.2052	0.2873	0.9179
2	0.2568	0.5438	0.7127
3	0.2138	0.7576	0.4562
4	0.1336	0.8912	0.2424
5	0.0668	0.9580	0.1088
6	0.0278	0.9858	0.0420
7	0.0099	0.9958	0.0142
8	0.0031	0.9989	0.0042
9	0.0009	0.9997	0.0011
10	0.0002	0.9999	0.0003
11	0.0000	1.0000	0.0001
12	0.0000	1.0000	0.0000

[그림 5.18] 포아송분포표

그래프 밑에는  $\lambda$  슬라이드바와 확률계산 박스가 있어 원하는 값을 넣고 [Enter] 키를 누르면 값이 계산된다.

실행  $\lambda =$   0.1  10  확률표시

확률  $P(\text{ } \leq X \leq \text{ }) =$

$P(X \leq \text{ }) =$

$P(X \geq \text{ }) =$

[그림 5.19]  $\lambda$  슬라이드바와 확률계산 박스

### 기하분포

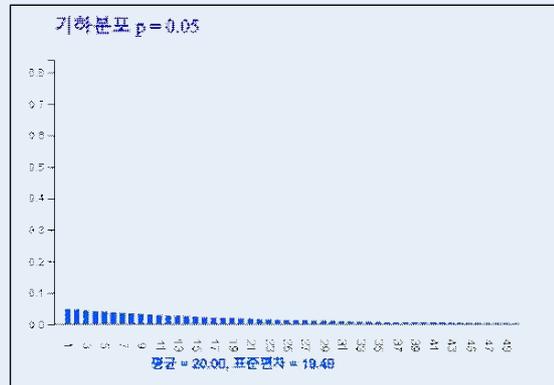
- 베르누이 시행에서 '성공'의 확률이  $p$  이고 '성공'이 나타날 때까지의 베르누이 시행회수를 변량  $X$ 라 할 때 이 분포를 기하분포(Geometric Distribution)라 한다.

$$f(x) = (1-p)^{x-1} p, \quad x = 1, 2, \dots$$

기하분포의 평균은  $E(X) = \frac{1}{p}$ , 분산은  $V(X) = \frac{1-p}{p^2}$  이다

[예 5.9] 한 공장에서 생산된 제품에서 불량률은 약 5%라고 한다. 불량품의 원인을 조사하기 위해 불량품을 찾을 때까지 계속 제품을 검사할 때 검사회수는 기하분포를 따른다. 『eStatU』를 이용하여 기하분포를 구해 보자.

『eStatU』의 주메뉴에서 '기하분포'를 선택하고  $p = 0.05$ 를 입력한 후 [실행] 버튼을 누르면 [그림 5.20]과 같은 기하분포가 나타나고 그 오른쪽에 기하분포표가 나타난다. 이 표에는  $P(X = x)$  이외에도 누적확률  $P(X \leq x)$ 와  $P(X \geq x)$ 를 같이 보여주어 여러 가지 확률계산을 쉽게 할 수 있다.



[그림 5.20] 『eStatU』의  $p = 0.05$ 인 기하분포

기하분포	$p = 0.05$		
x	$P(X = x)$	$P(X \leq x)$	$P(X \geq x)$
0	0.0000	0.0000	1.0000
1	0.0500	0.0500	1.0000
2	0.0475	0.0975	0.9500
3	0.0451	0.1426	0.9025
4	0.0429	0.1855	0.8574
5	0.0407	0.2262	0.8145
6	0.0387	0.2649	0.7738
7	0.0368	0.3017	0.7351
8	0.0349	0.3366	0.6983
9	0.0332	0.3698	0.6634
10	0.0315	0.4013	0.6302

[그림 5.21] 기하분포표

그래프 밑에는  $p$  슬라이드바와 확률계산 박스가 있어 원하는 값을 넣고 [Enter] 키를 누르면 값이 계산된다.

실행  $p = 0.05$     확률표시

확률  $P(\text{ } \leq X \leq \text{ }) = \text{ }$

$P(X \leq \text{ }) = \text{ }$

$P(X \geq \text{ }) = \text{ }$

[그림 5.22]  $p$  슬라이드바와 확률계산 박스

### 초기하분포

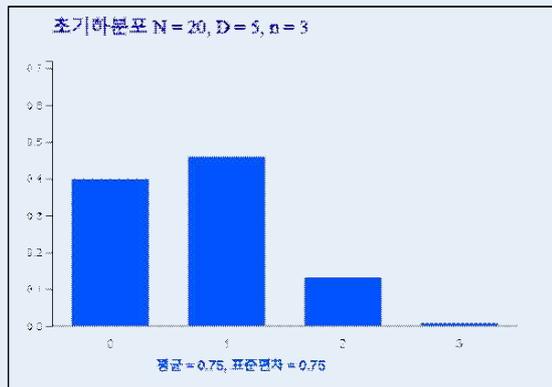
- 집단의 크기가 유한개인  $N$ 개이고 집단(속성이 '성공'인 것이  $D$ 개, 아닌 것이  $N-D$ 개)에서  $n$ 개를 추출할 때 '성공의 회수( $X$ )' 를 초기하 확률변량 (hypergeometric random variable)이라 하고, 그 분포를 초기하분포 (hypergeometric distribution)라 한다.

$$f(x) = \frac{D C_x N-D C_{n-x}}{N C_n}$$

$p = D/N$  라 할 때, 초기하분포의 평균은  $E(X) = np$ , 분산은  $V(X) = np(N-n)/(N-1)$  이다.

[예 5.10] 20개의 담배제품(우량품 15개, 불량품 5개)이 들어 있는 상자에서 3개를 추출하였을 때 이중 불량품의 개수는 초기하분포를 따른다. 「eStatU」를 이용하여 초기하분포를 구해 보자.

「eStatU」의 주메뉴에서 '초기하분포'를 선택하고  $N = 20, D = 5, n = 3$ 을 선택하고 '실행' 버튼을 클릭하면 [그림 5.23]과 같은 초기하분포가 나타나고 그 오른쪽에 초기하분포표가 나타난다. 이 표에는  $P(X = x)$  이외에도 누적확률  $P(X \leq x)$ 와  $P(X \geq x)$ 를 같이 보여주어 여러 가지 확률계산을 쉽게 할 수 있다.



[그림 5.23]  $N = 20, D = 5, n = 3$ 인 초기하분포

초기하분포	$N = 20$	$D = 5$	$n = 3$
$x$	$P(X = x)$	$P(X \leq x)$	$P(X \geq x)$
0	0.3991	0.3991	1.0000
1	0.4605	0.8596	0.6009
2	0.1316	0.9912	0.1404
3	0.0088	1.0000	0.0088

[그림 5.24] 초기하분포표

그래프 밑에는  $n$  슬라이드바와 확률계산 박스가 있어 원하는 값을 넣고 **Enter** 키를 누르면 값이 계산된다.

실행  $n =$     확률표시

확률  $P(\text{ } \leq X \leq \text{ }) =$

$P(X \leq \text{ }) =$

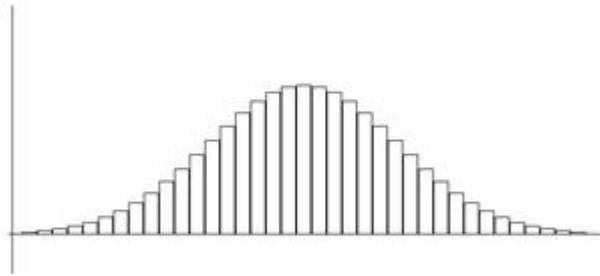
$P(X \geq \text{ }) =$

[그림 5.25]  $n$  슬라이드바와 확률계산 박스

## 5.4 연속형 확률변량

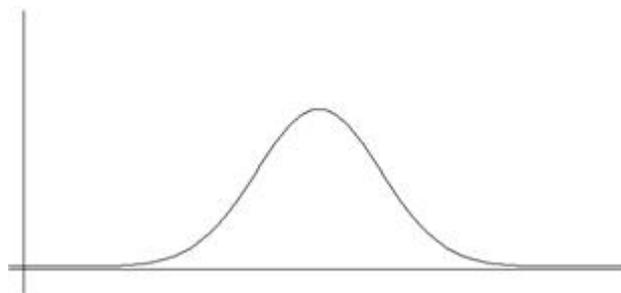
### 정규분포

- 한 피자 가게에 피자를 주문하여 집까지 배달되는 시간을 임의로 많이 조사하여 히스토그램을 그린 결과가 다음과 같다.



- 우리 주변의 연속확률변수의 자료 중에서는 위의 히스토그램과 같이 종을 엮어 놓은 모양으로 평균 근처에 자료가 많이 몰려있고, 평균에서 멀어질수록 자료 수가 적으며, 평균을 중심으로 대칭형인 형태가 많이 관찰된다. 이러한 형태의 모든 자료들에 대한 확률을 쉽게 구하기 위해 많은 수학자들이 이 분포 형태를 묘사할 수 있는 함수를 찾았다. 이 수학적 함수를 이용하면 굳이 도수분포표나 히스토그램을 그리지 않고 원하는 확률을 근사적으로 구할 수 있다.
- 드 브와브르(Abraham de Moivre(1667-1754))에 의해 이와 같은 함수를 처음 발견되었고, 그 후 독일의 수학자 가우스(Carl Friedrich Gauss(1777-1855))에 의해 물리학과 천문학 등에 폭 넓게 응용되었다. 이 함수를 **정규분포함수** 또는 **가우스분포함수**라고 부르는데 함수식과 그래프는 다음과 같다.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad -\infty < x < \infty$$

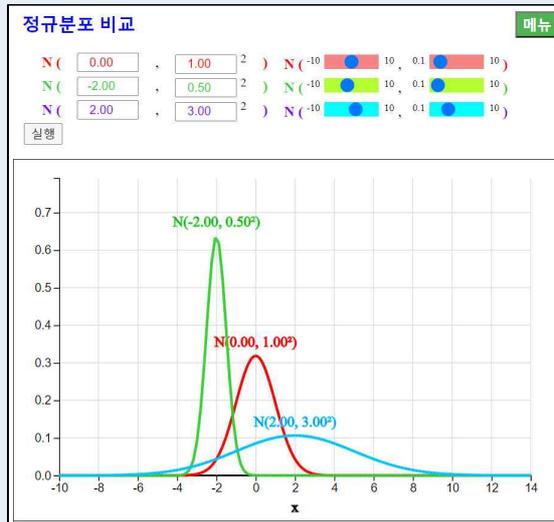


정규분포함수 그래프

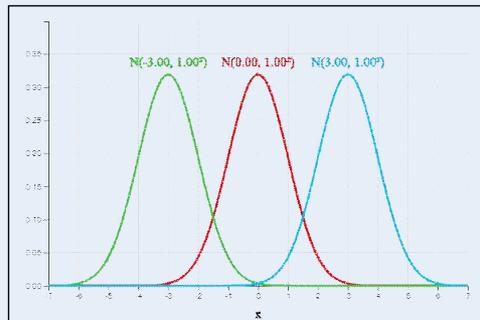
- 이 분포함수는 두개의 모수  $\mu$ 와  $\sigma$ 를 가지고 있는데, 각각 이 함수의 평균과 표준편차를 의미한다. 확률변량  $X$ 가 평균  $\mu$ , 표준편차  $\sigma$ 인 정규분포를 따를 때 기호로  $X \sim N(\mu, \sigma^2)$ 으로 표시하기도 한다.

[예 5.11] 『eStatU』를 이용하여 여러 가지 정규분포 함수의 모양을 비교해 보자.

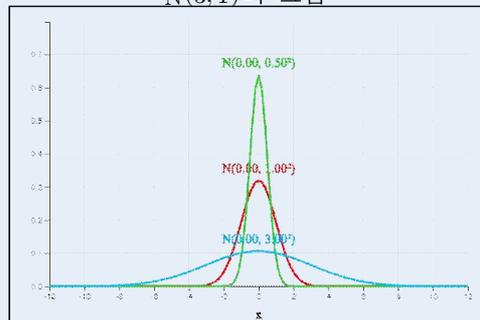
『eStatU』의 주메뉴에서 ‘정규분포 비교’를 선택하면 [그림 5.26]과 같은 화면이 나타난다. 세 정규분포  $N(-2, 0.5^2)$ ,  $N(0, 1)$ ,  $N(2, 3^2)$ 를 비교하기 위해 그림과 같이 평균 및 표준편차를 입력하고 [실행] 버튼을 누르면 세 정규분포의 그림이 나타난다. 평균을 중심으로 모두 대칭이고, 분산이 커지면 정규분포는 평평해지고, 분산이 작아지면 뾰족해 지는 것을 관찰할 수 있다. 이는 전체 넓이가 1이 되어야하기 때문이다. 슬라이드바를 이용하여 평균과 표준편차를 변화시키며 정규분포의 모양을 관찰할 수 있다.



[그림 5.26] 『eStatU』의 정규분포 비교



[그림 5.27] 세 정규분포  $N(-3, 1)$ ,  $N(0, 1)$ ,  $N(3, 1)$ 의 그림



[그림 5.28] 세 정규분포  $N(0, 0.5^2)$ ,  $N(0, 1)$ ,  $N(0, 3^2)$ 의 그림

정규분포에서의 확률계산

[예 5.12] 『eStatU』를 이용하여 정규분포의 확률계산을 하여 보자.

『eStatU』의 주메뉴에서 ‘정규분포’를 선택하면 [그림 5.29]와 같은 화면이 나타난다. 여기에서 어떠한 정규분포 확률변수의 구간  $[a, b]$ 에 대한  $P(a \leq X \leq b)$  확률계산과, 주어진 확률  $p$ 에 대한 백분위수(즉,  $P(X \leq x) = p$ 가 되는 백분위수  $x$ 를 쉽게 계산할 수 있다. 구간의 확률은  $\mu - 4\sigma$ 에서  $\mu + 4\sigma$ 까지 계산할 수 있다.  $X$ 가  $\mu - 4\sigma$ 보다 적든지  $\mu + 4\sigma$ 보다 큰 경우에 확률은 0.0000이 된다. [그림 5.30]은 표준정규분포의 백분위수표이다.



정규분포    평균  $\mu =$      표준편차  $\sigma =$      메뉴

숫자 입력 후 [실행] 또는 [Enter] 클릭

**확률**

$P(-1.960 \leq X \leq 1.960) = 0.9500$

$P(X \leq -1.645) = 0.0500$    

$P(X \geq 1.645) = 0.0500$

**백분위수**

$P(-1.960 \leq X \leq 1.960) = 0.9500$

$P(X \leq -1.645) = 0.0500$    

$P(X \geq 1.645) = 0.0500$

참고사이트 [Wikipedia](#) [Wolfram](#) [StatTrek](#) [KhanAcademy](#)

[그림 5.29] 『eStatU』에서 정규분포 확률계산

정규분포	$\mu = 0$	$\sigma = 1.000$								
$p$	$P(X \leq x) = p$	$p$	$P(X \leq x) = p$	$p$	$P(X \leq x) = p$	$p$	$P(X \leq x) = p$	$p$	$P(X \leq x) = p$	$p$
0.005	-3.576	0.305	-0.824	0.405	-0.240	0.605	0.266	0.805	0.860	
0.010	-3.226	0.210	-0.806	0.410	-0.235	0.610	0.279	0.810	0.878	
0.015	-3.170	0.215	-0.789	0.415	-0.215	0.615	0.292	0.815	0.896	
0.020	-3.054	0.220	-0.772	0.420	-0.202	0.620	0.305	0.820	0.915	
0.025	-2.960	0.225	-0.755	0.425	-0.189	0.625	0.319	0.825	0.935	
0.030	-2.881	0.230	-0.739	0.430	-0.176	0.630	0.332	0.830	0.954	
0.035	-2.812	0.235	-0.722	0.435	-0.164	0.635	0.345	0.835	0.974	
0.040	-2.751	0.240	-0.706	0.440	-0.151	0.640	0.358	0.840	0.994	
0.045	-2.695	0.245	-0.690	0.445	-0.138	0.645	0.372	0.845	1.015	
0.050	-2.645	0.250	-0.674	0.450	-0.126	0.650	0.385	0.850	1.036	
0.055	-2.598	0.255	-0.659	0.455	-0.113	0.655	0.399	0.855	1.058	
0.060	-2.555	0.260	-0.643	0.460	-0.100	0.660	0.412	0.860	1.080	
0.065	-2.514	0.265	-0.628	0.465	-0.088	0.665	0.426	0.865	1.103	
0.070	-2.476	0.270	-0.613	0.470	-0.075	0.670	0.440	0.870	1.128	
0.075	-2.440	0.275	-0.598	0.475	-0.063	0.675	0.454	0.875	1.150	
0.080	-2.405	0.280	-0.583	0.480	-0.050	0.680	0.468	0.880	1.175	
0.085	-2.372	0.285	-0.568	0.485	-0.038	0.685	0.482	0.885	1.200	
0.090	-2.341	0.290	-0.553	0.490	-0.025	0.690	0.496	0.890	1.227	
0.095	-2.311	0.295	-0.539	0.495	-0.013	0.695	0.510	0.895	1.254	
0.100	-2.282	0.300	-0.524	0.500	0.000	0.700	0.524	0.900	1.282	

[그림 5.30] 『eStatU』의 표준정규분포 백분위수표

## 지수분포

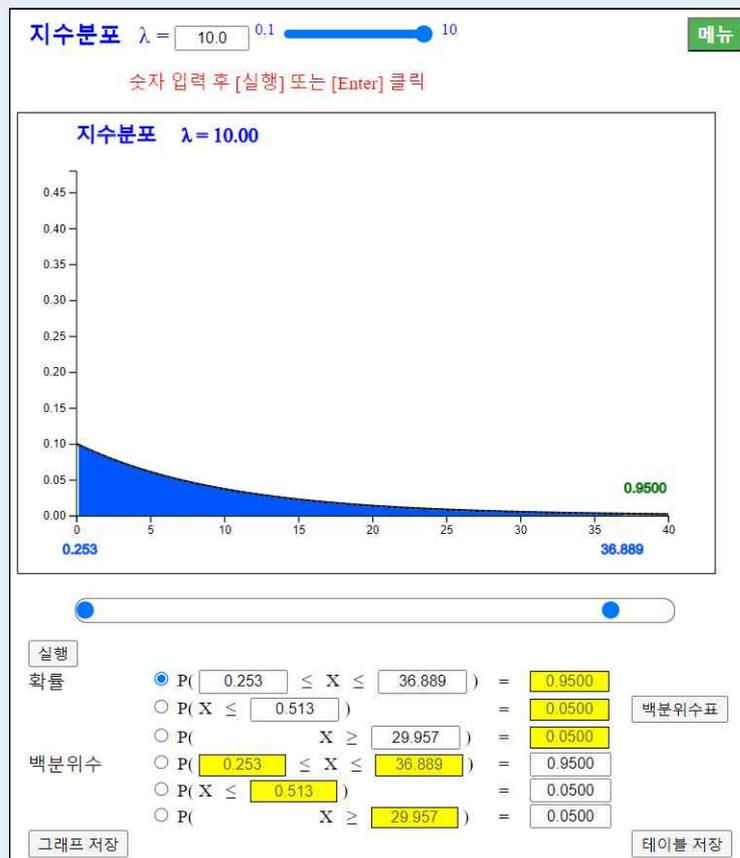
- 단위 시간당 발생하는 평균 사건수를  $\lambda$ 라 했을 때 확률변량  $X$ 를 발생하는 사건들 사이의 시간이라고 하면  $X$ 는 다음과 같은 지수분포(Exponential Distribution) 모형을 적용할 수 있다.  $\lambda$ 는 지수분포의 모수이고 지수분포함수 식은 다음과 같다.

$$f(x) = \lambda \exp(-\lambda x), \quad x > 0$$

$$\text{지수분포의 평균은 } E(X) = \frac{1}{\lambda}, \text{ 분산은 } V(X) = \frac{1}{\lambda^2} \text{ 이다.}$$

[예 5.13] 한 제품의 평균 수명은 10시간이며 지수분포를 따른다. 『eStatU』를 이용하여 이 지수분포의 확률을 구해 보자.

『eStatU』의 주메뉴에서 ‘지수분포’를 선택하면 [그림 5.31]과 같은 지수분포창이 나타난다. 여기서  $\lambda = 10$ 을 입력하고 [실행] 버튼을 누르면 지수분포 확률변수의 구간  $[a, b]$ 에 대한  $P(a \leq X \leq b)$  확률계산과, 주어진 확률  $p$ 에 대한 백분위수(즉,  $P(X \leq x) = p$ 가 되는 백분위수  $x$ )를 쉽게 계산할 수 있다. 그래프 밑의 ‘백분위수표’ 버튼을 누르면 지수분포의 백분위수표가 나타난다.



[그림 5.31] 『eStatU』의 지수분포 확률 계산