fast campus Copyright FASTCAMPUS Corp.All Rights Reserved

선수 지식 - 통계

이산확률분포

이산확률분포 | 딥러닝의 기초가 되는 확률 개념 알아보기

강사 나동빈

fast campus Copyright FASTCAMPUS Corp. All Rights Reserved

선수 지식 - 통계

이산확률분포

Copyright FASTCAMPUS Corp. All Rights Reserved

이산확률분포(Discrete Probability Distribution)

- 확률변수 X가 취할 수 있는 모든 값을 셀 수 있는 경우, 이를 **이산확률변수**라고 한다.
- 이때 **이산확률분포**는 이산확률변수의 확률 분포를 의미한다.
- 주사위를 던졌을 때 나올 수 있는 눈금(수)을 확률변수 X라고 하자.
- 확률변수 X는 {1, 2, 3, 4, 5, 6} 중 하나의 값을 가질 수 있다.

	X=1	X=2	X=3	X=4	X=5	X=6
P(X)	1/6	1/6	1/6	1/6	1/6	1/6

fast campus Copyright FASTCAMPUS Corp. All Rights Reserved

확률질량함수(Probability Mass Function, PMF)

- **확률질량함수**는 이산확률변수가 특정한 값을 가질 **확률을 출력하는 함수**다.
- **확률질량함수**는 이산확률분포를 표현하기 위해 사용하는 확률분포함수로 이해할 수 있다.
- 동전 2개를 동시에 던지는 시행에서 두 눈금의 합을 X라고 하자.
- 이때, X는 이산확률변수로, 확률질량함수 $f(x)$는 다음과 같이 정의할 수 있다.

$$f(0) = P(X = 0) = 1/4$$

$$f(1) = P(X = 1) = 1/2$$

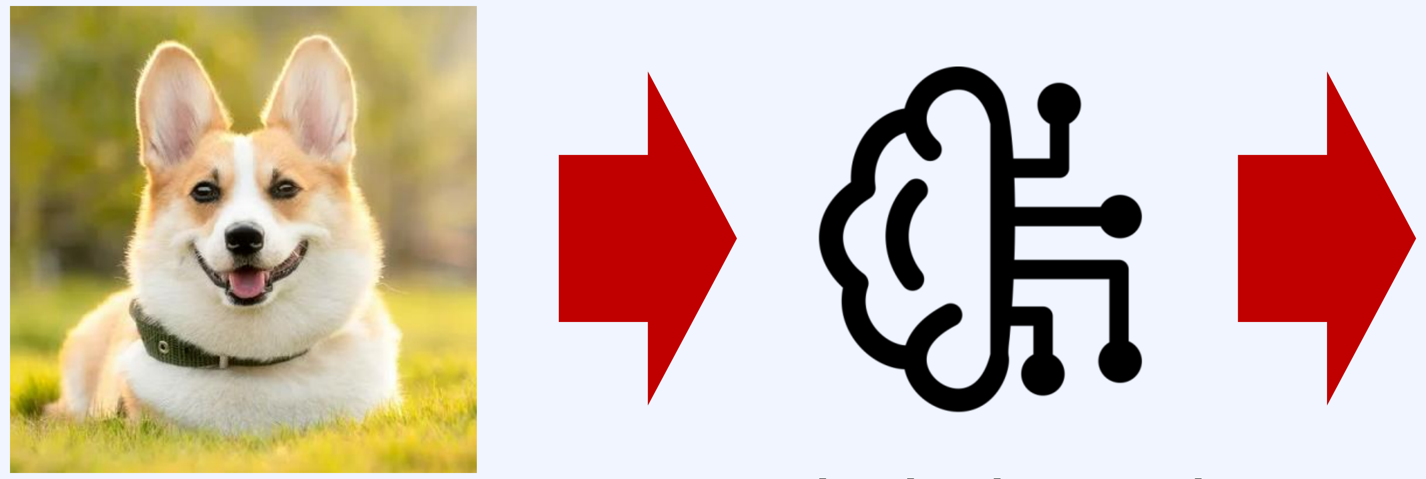
$$f(2) = P(X = 2) = 1/4$$

- 확률 변수 X에 대한 확률질량함수라는 의미로 $f_X(x)$라고 표기하기도 한다.

fast campus Copyright FASTCAMPUS Corp. All Rights Reserved

확률질량함수 예시

- 한 장의 이미지 x가 주어졌을 때, 분류 모델의 실행 결과가 다음과 같다고 해보자.
- $P(Y = \text{고양이}|X = x) = 15\%$
- $P(Y = \text{강아지}|X = x) = 55\%$
- $P(Y = \text{다람쥐}|X = x) = 30\%$



이미지 x 딥러닝 모델 추론 결과

[참고]

- $P(Y|X)$는 X 값이 주어졌을 때, 확률 변수 Y에 대한 **확률 분포**를 의미한다.
- **예시)** X가 이미지, Y가 클래스(class)라고 하면, 한 장의 이미지가 어떤 동물인지 예측하는 모델의 출력 결과로 이해할 수 있다.

fast campus Copyright FASTCAMPUS Corp. All Rights Reserved

베르누이 시행(Bernoulli Trial)

- 결과가 두 가지 중 하나로만 나오는 시행을 **베르누이 시행**이라고 한다.
- **예시 1)** 입학 시험 → 합격 혹은 불합격
- **예시 2)** 동전 던지기 → 앞면 혹은 뒷면
- **예시 3)** 꽝 혹은 당첨만 있는 복권

 Copyright FASTCAMPUS Corp. All Rights Reserved

베르누이 확률변수

- 베르누이 시행의 결과를 실수 0 혹은 1로 나타낸다.
- 확률 변수는 0 혹은 1의 값만 가질 수 있으므로, 이산확률변수다.

Copyright FASTCAMPUS Corp. All Rights Reserved

베르누이 확률분포

- 베르누이 확률변수의 분포를 베르누이 확률분포라고 한다.
- 확률변수 X가 베르누이 분포를 따른다고 표현하며, 수식으로는 다음과 같이 표현한다.

$$X \sim Bern(x; \mu)$$

[참고]

- 모수(parameter)는 세미콜론(;) 기호로 구분하여 표기한다.
- 베르누이 확률분포는 모수로 μ를 가지는데, 1이 나올 확률을 의미한다.

 Copyright FASTCAMPUS Corp. All Rights Reserved

베르누이 분포의 확률질량함수

선수 지식
통계
이산확률분포

- 베르누이 확률 분포의 확률질량함수는 다음과 같다.

$$Bern(x;\mu) = \begin{cases} \mu, & if\ x = 1 \\ 1-\mu, & if\ x = 0 \end{cases}$$

이는 간단히 아래와 같은 하나의 수식으로 표현할 수 있다.

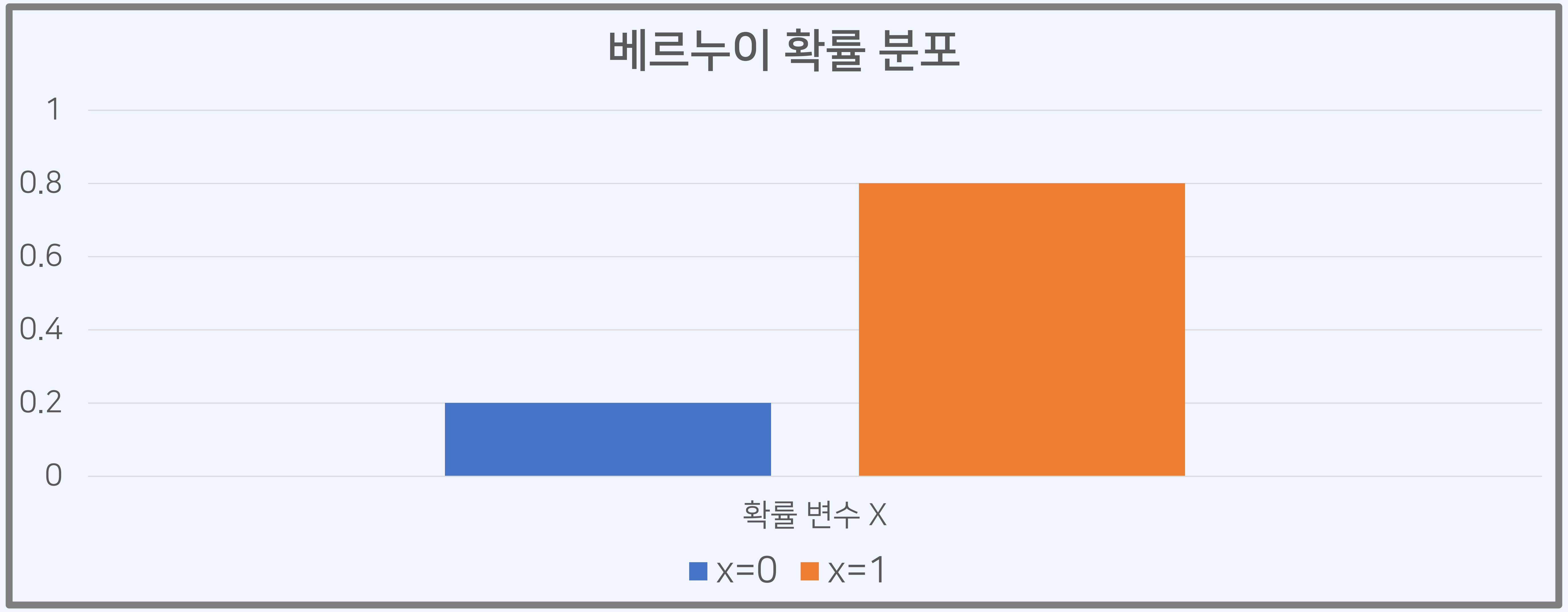
$$Bern(x;\mu) = \mu^x(1-\mu)^{1-x}$$

 Copyright FASTCAMPUS Corp. All Rights Reserved

베르누이 분포의 확률질량함수

- μ가 0.8인 베르누이 확률 분포는 다음과 같다.

$$Bern(x;\mu) = \begin{cases} \mu, & if\ x = 1 \\ 1-\mu, & if\ x = 0 \end{cases}$$



 Copyright FASTCAMPUS Corp. All Rights Reserved

이항 분포 개요

- 베르누이 시행을 N번 반복하는 경우가 있다.
- **예시)** 동전 던지기를 7회 시행할 수 있다.

 Copyright FASTCAMPUS Corp. All Rights Reserved

이항 분포란?

- 성공 확률이 μ인 베르누이 시행을 N번 반복한다.
- N번 중에서 성공한 횟수를 확률 변수 X라고 하자.
- X는 0부터 N까지의 정수 중 하나이다.
- 이러한 확률 변수를 이항 분포를 따른다고 한다.

$$X \sim Bin(x; N, \mu)$$

- 이항 분포는 모수(parameter)로 N과 μ를 가진다.
- **파라미터 1**: 시행 횟수 N
- **파라미터 2**: 한 번의 횟수에서 1이 나올 확률 μ

Copyright FASTCAMPUS Corp. All Rights Reserved

이항 분포란?

- 이항 분포 확률 변수 X의 확률 질량 함수는 다음과 같다.

$$X \sim Bin(x; N, \mu) = \binom{N}{x} \mu^x (1-\mu)^{N-x}$$

- 단, $\binom{N}{x}$는 N개에서 x개를 선택하는 조합(combination)의 수와 같다.
- 단, $N! = N \cdot (N-1) \cdots 2 \cdot 1$이다.

 Copyright FASTCAMPUS Corp. All Rights Reserved

이항 분포 공식

- 독립된 사건을 N번 반복 시행했을 때, 특정 사건이 x회 발생한다고 가정한다.
- **이항 분포**: 아래의 확률 값을 그대로 확률 질량 함수로 사용한다.

$$\underbrace{\binom{N}{x}} \times \underbrace{\mu^{x}} \times \underbrace{(1-\mu)^{N-x}}$$

: N번에서 x개를 고르는 조합의 수

: μ의 확률이 x번 적용

: $(1-\mu)$의 확률이 $(N-x)$번 적용

fast campus Copyright FASTCAMPUS Corp. All Rights Reserved

이항 분포 문제 예시 ① 고양이 분류 모델

[문제]

- 고양이 분류 딥러닝 모델 θ는 5개의 고양이 사진 중에 4개를 정확히 예측한다고 한다.
- 모델에 10개의 고양이 사진을 주었을 때, 7개를 정확히 예측할 확률은 얼마일까?

[해설]

- 예측 성공 확률: 80%, 예측 실패 확률: 20% → $p = 80\%$
- 10개 중에서 7개를 정확히 예측해야 하는 것이므로, 다음과 같다.
- $\binom{10}{7} p^7 \times (1 - p)^3 = 0.2013$

fast campus Copyright FASTCAMPUS Corp. All Rights Reserved

이항 분포 문제 예시 ② 가구 공장

[문제]

- 가구 공장에서 가구를 만들 때, 불량률이 10%라고 한다.
- 이 공장에서 만든 가구 10개를 확인했을 때, 불량품이 2개 이하로 나올 확률을 구하여라.
- 불량률 10% → $p = 10\%$

[해설]

- 불량품이 0개 나올 확률 + 불량품이 1개 나올 확률 + 불량품이 2개 나올 확률

$$= \binom{10}{0} p^0 \times (1-p)^{10} + \binom{10}{1} p^1 \times (1-p)^9 + \binom{10}{2} p^2 \times (1-p)^8$$

= 0.3487 + 0.3874 + 0.1937 = 0.9298

 Copyright FASTCAMPUS Corp. All Rights Reserved

포아송 분포

- **일정한 시간** 내 발생하는 사건의 **발생 횟수**에 대한 확률을 계산할 때 사용한다.
- 단위 시간에 어떤 사건이 발생할 기댓값이 λ일 때, 그 사건이 x회 일어날 확률을 구할 수 있다.
- 포아송 분포는 푸아송 분포라고 부르기도 한다.
- 포아송 분포의 확률 질량 함수는 다음과 같다.

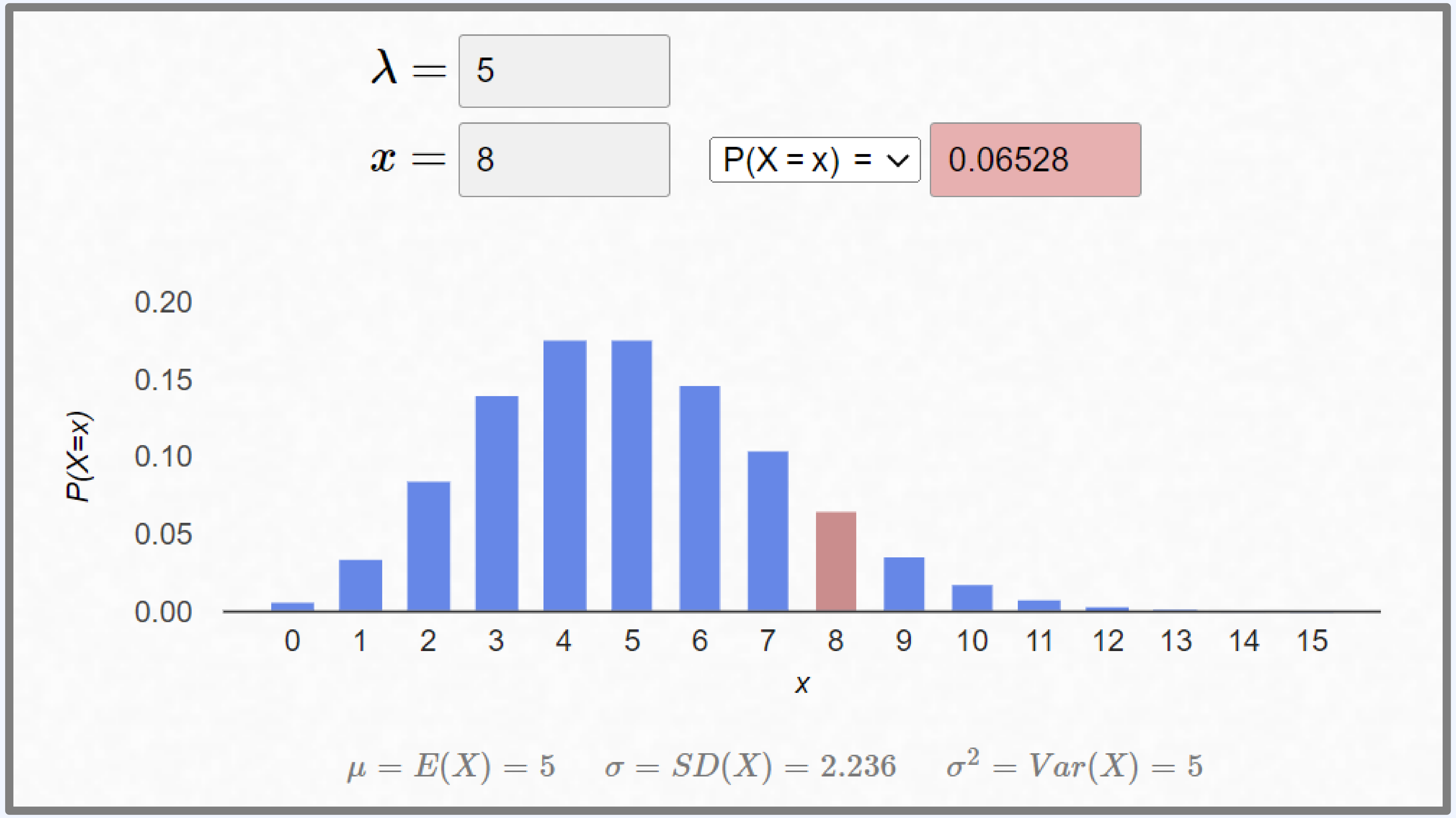
$$f(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

- 포아송 분포의 평균을 λ로 표기한다.
- e는 자연 상수를 의미한다. ($e = 2.718 \ldots$)

Copyright FASTCAMPUS Corp. All Rights Reserved

포아송 분포 그림 예시

- **단위 시간** 내 **평균 발생 횟수(λ)**가 5일 때, 그 사건이 8회 일어날 확률은?



$$\frac{e^{-5}5^8}{8!} = 0.06528$$

* https://homepage.divms.uiowa.edu/~mbognar/applets/pois.html

Copyright FASTCAMPUS Corp. All Rights Reserved

포아송 분포 문제 예시 – 스팸 메일

[문제]

- **하루**에 **평균적으로 5개**의 스팸 메일이 도착한다.

1) 오늘 **하루** 동안 스팸 메일이 **1개** 도착할 확률은 얼마일까?

2) 오늘 **하루** 동안 스팸 메일이 **5개** 도착할 확률을 얼마일까?

3) 오늘 **하루** 동안 스팸 메일이 **8개** 도착할 확률을 얼마일까?

fast campus Copyright FASTCAMPUS Corp.All Rights Reserved

포아송 분포 문제 예시 – 스팸 메일

- 단위 시간에 스팸 메일이 5개 도착한다. 따라서 평균 발생 횟수(λ)는 5다.

$$f(x) = \frac{e^{-5}5^{x}}{x!}$$

1. 스팸 메일이 1개 도착할 확률: $f(1) = \frac{e^{-5}5^{1}}{1!} = 0.0337$
2. 스팸 메일이 5개 도착할 확률: $f(5) = \frac{e^{-5}5^{5}}{5!} = 0.1755$
3. 스팸 메일이 8개 도착할 확률: $f(8) = \frac{e^{-5}5^{8}}{8!} = 0.0653$

fast campus Copyright FASTCAMPUS Corp. All Rights Reserved

간단히 수식 계산하는 방법

선수 지식
통계
이산확률분포

- 구글(Google) 검색 엔진에 수식을 입력하면, 계산된 결과를 얻을 수 있다.