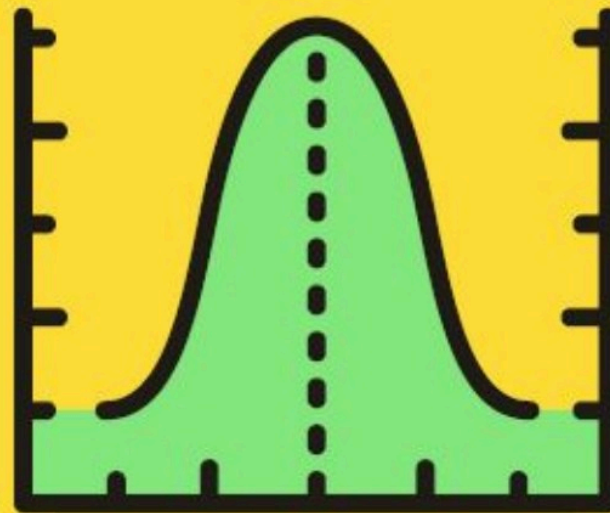


누구나 파이썬 통계분석



CHAPTER 06

대표적인 이산형 확률변수

CONTENTS

- 6.1 베르누이 분포
- 6.2 이항분포
- 6.3 기하분포
- 6.4 포아송 분포

6.0 준비

■ 모수적 기법

'모집단이 이와 같은 성질일 것이므로 이러한 형태를 가지는 확률분포일 것이다.'라는 가정을 하고, 그 뒤에 확률분포의 기댓값이나 분산을 결정하는 소수의 파라미터 추측

In [1]:

```
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats

%precision 3
%matplotlib inline
```

6.0 준비

In [2]:

그래프에서 선의 종류

```
linestyles = [ '-', '--', ':' ]
```

```
def E(X, g=lambda x: x):
```

```
    x_set, f = X
```

```
    return np.sum([g(x_k) * f(x_k) for x_k in x_set])
```

```
def V(X, g=lambda x: x):
```

```
    x_set, f = X
```

```
    mean = E(X, g)
```

```
    return np.sum([(g(x_k)-mean)**2 * f(x_k) for x_k in x_set])
```

```
def check_prob(X):
```

```
    x_set, f = X
```

```
    prob = np.array([f(x_k) for x_k in x_set])
```

```
    assert np.all(prob >= 0), 'minus probability'
```

```
    prob_sum = np.round(np.sum(prob), 6)
```

```
    assert prob_sum == 1, f'sum of probability{prob_sum}'
```

```
    print(f'expected value {E(X):.4}')
```

```
    print(f'variance{(V(X)):.4}')
```

```
def plot_prob(X):
```

```
    x_set, f = X
```

```
    prob = np.array([f(x_k) for x_k in x_set])
```

```
    fig = plt.figure(figsize=(10, 6))
```

```
    ax = fig.add_subplot(111)
```

```
    ax.bar(x_set, prob, label='prob')
```

```
    ax.vlines(E(X), 0, 1, label='mean')
```

```
    ax.set_xticks(np.append(x_set, E(X)))
```

```
    ax.set_ylim(0, prob.max()*1.2)
```

```
    ax.legend()
```

```
    plt.show()
```

6.1 베르누이 분포

- 확률변수가 취할 수 있는 값이 0과 1밖에 없는 분포
- 베르누이 분포를 따르는 확률변수의 시행이 베르누이 시행
 - 1은 성공, 0은 실패
- 1이 나오는 확률은 p , 0이 나오는 확률은 $1 - p$
- 파라미터 p 인 베르누이 분포는 $\text{Bern}(p)$

베르누이 분포의 확률함수

$$f(x) = \begin{cases} p^x (1-p)^{(1-x)} & (x \in \{0, 1\}) \\ 0 & (\text{otherwise}) \end{cases}$$

6.1 베르누이 분포

- 동전을 던져 앞면이 나올 확률

$$P(X = 1) = \left(\frac{1}{2}\right)^1 \times \left(1 - \frac{1}{2}\right)^{(1-1)} = \frac{1}{2}$$

- 주사위를 한 번 굴려 6이 나오지 않을 확률

$$P(X = 0) = \left(\frac{1}{6}\right)^0 \times \left(1 - \frac{1}{6}\right)^{(1-0)} = \frac{5}{6}$$

6.1 베르누이 분포

베르누이 분포의 기댓값과 분산

$X \sim \text{Bern}(p)$ 라고 할 때

$$E(X) = p, \quad V(X) = p(1 - p)$$

In [3] :

```
def Bern(p):  
    x_set = np.array([0, 1])  
    def f(x):  
        if x in x_set:  
            return p ** x * (1-p) ** (1-x)  
        else:  
            return 0  
    return x_set, f
```

– $X \sim \text{Bern}(0.3)$

In [4] :

```
p = 0.3  
X = Bern(p)
```

6.1 베르누이 분포

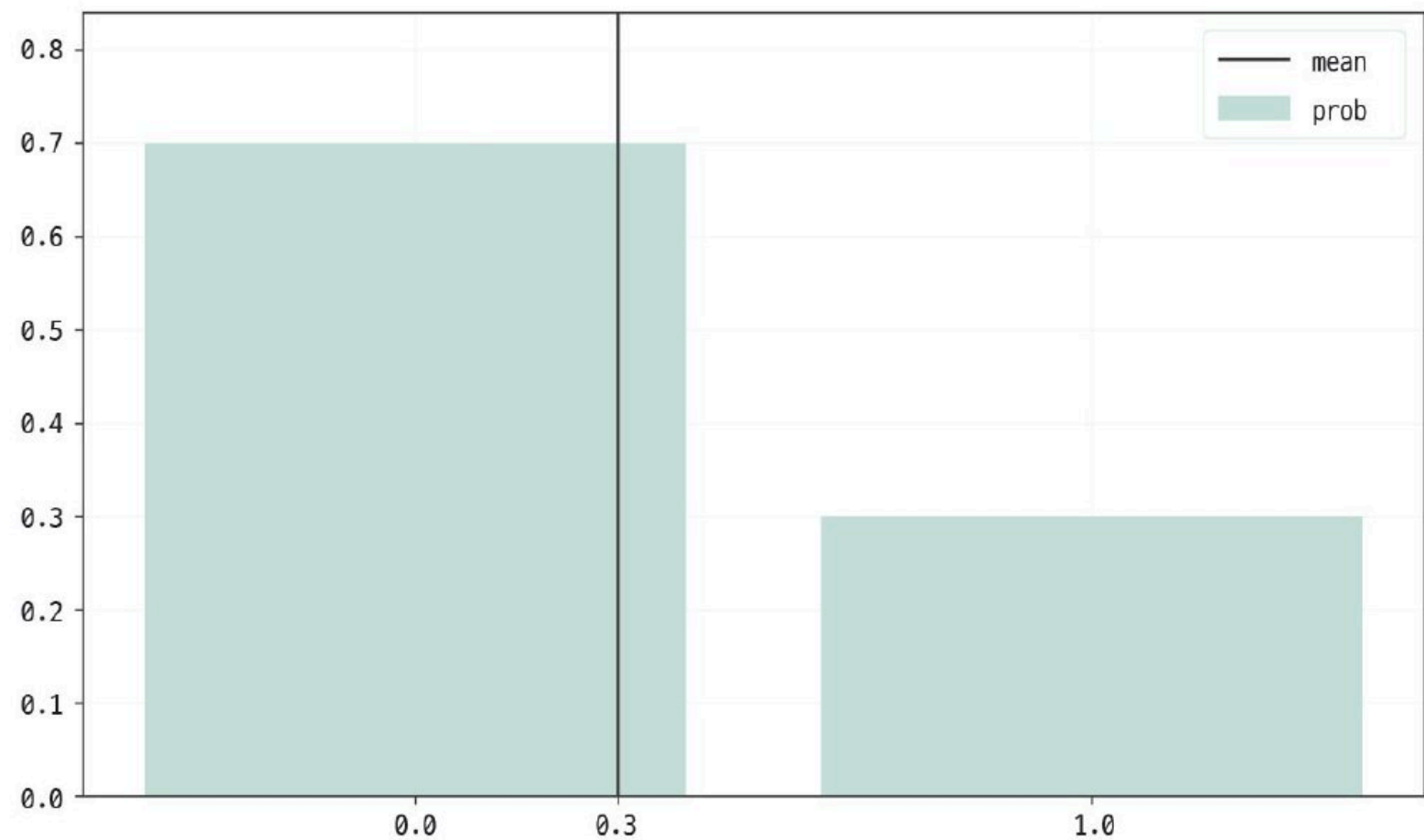
- 기댓값은 0.3, 분산은 $0.3 \times 0.7 = 0.21$

In [5]:

```
check_prob(X)
```

Out [5]:

```
expected value 0.3  
variance 0.21
```



[그림 6-1] 베르누이 분포

6.1 베르누이 분포

scipy.stats 의 bernoulli 함수는 베르누이 분포를 따르는 확률변수를 생성한다 ($p = 0.3$)

```
In [7]: rv = stats.bernoulli(p)
```

```
In [8]: rv.pmf(0), rv.pmf(1)
```

```
Out[8]: (0.700, 0.300)
```

```
In [9]: rv.pmf([0, 1])
```

```
Out[9]: array([0.7, 0.3])
```

```
In [10]: rv.cdf([0, 1])
```

```
Out[10]: array([0.7, 1. ])
```

1 pmf 메서드는 확률함수 계산

2 cdf 메서드는 누적분포함수 계산

6.1 베르누이 분포

- mean, var 메서드로 기댓값, 분산 계산

In [11]:

```
rv.mean(), rv.var()
```

Out [11]:

```
(0.300, 0.210)
```

[표 6-1] 베르누이 분포의 정리

파라미터	p
취할 수 있는 값	$\{0, 1\}$
확률함수	$p^x(1-p)^{(1-x)}$
기댓값	p
분산	$p(1-p)$
scipy.stats	<code>bernoulli(p)</code>

$X \sim \text{Bern}(0.3)$ 에서 $P(X = 0)$ 은?

성공이면 1, 실패면 0 을 취하는 베르누이 분포입니다. 성공 확률은 0.3 입니다.

- ① 0
- ② 0.3
- ③ 0.21
- ④ 0.7

$X \sim \text{Bern}(0.3)$ 에서 $P(X = 0)$ 은?

4번 — 0.7

$X = 0$ 은 실패입니다. 실패 확률은 $1 - p = 1 - 0.3 = 0.7$ 입니다. 0.3 은 $X = 1$ 일 확률이고, 0.21 은 분산 $p(1 - p)$ 입니다.

6.2 이항분포

- 성공 확률이 p 인 베르누이 시행을 n 번 했을 때의 성공 횟수가 따르는 분포
- 파라미터 n, p 인 이항분포는 $\text{Bin}(n, p)$ ${}_nC_x = \frac{n!}{x!(n-x)!}$

이항분포의 확률함수

$$f(x) = \begin{cases} {}_nC_x p^x (1-p)^{(n-x)} & (x \in \{0, 1, \dots, n\}) \\ 0 & (\text{otherwise}) \end{cases}$$

6.2 이항분포

- 동전을 10번 던져 앞면이 3번 나올 확률

$$P(X = 3) = {}_{10}C_3 \left(\frac{1}{2}\right)^3 \left(1 - \frac{1}{2}\right)^{(10-3)} = \frac{15}{128}$$

- 주사위를 4번 굴려 6이 한 번도 나오지 않을 확률

$$P(X = 0) = {}_4C_0 \left(\frac{1}{6}\right)^0 \left(1 - \frac{1}{6}\right)^{(4-0)} = \frac{625}{1296}$$

6.2 이항 분포

이항분포의 기댓값과 분산

$X \sim \text{Bin}(n, p)$ 라고 할 때

$$E(X) = np, \quad V(X) = np(1 - p)$$

In [12] :

```
from scipy.special import comb

def Bin(n, p):
    x_set = np.arange(n+1)
    def f(x):
        if x in x_set:
            return comb(n, x) * p**x * (1-p)**(n-x)
        else:
            return 0
    return x_set, f
```

– $X \sim \text{Bin}(10, 0.3)$

In [13] :

```
n = 10
p = 0.3
X = Bin(n, p)
```

- 기댓값은 $10 \times 0.3 = 3$, 분산은 $10 \times 0.3 \times 0.7 = 2.1$

In [14] :

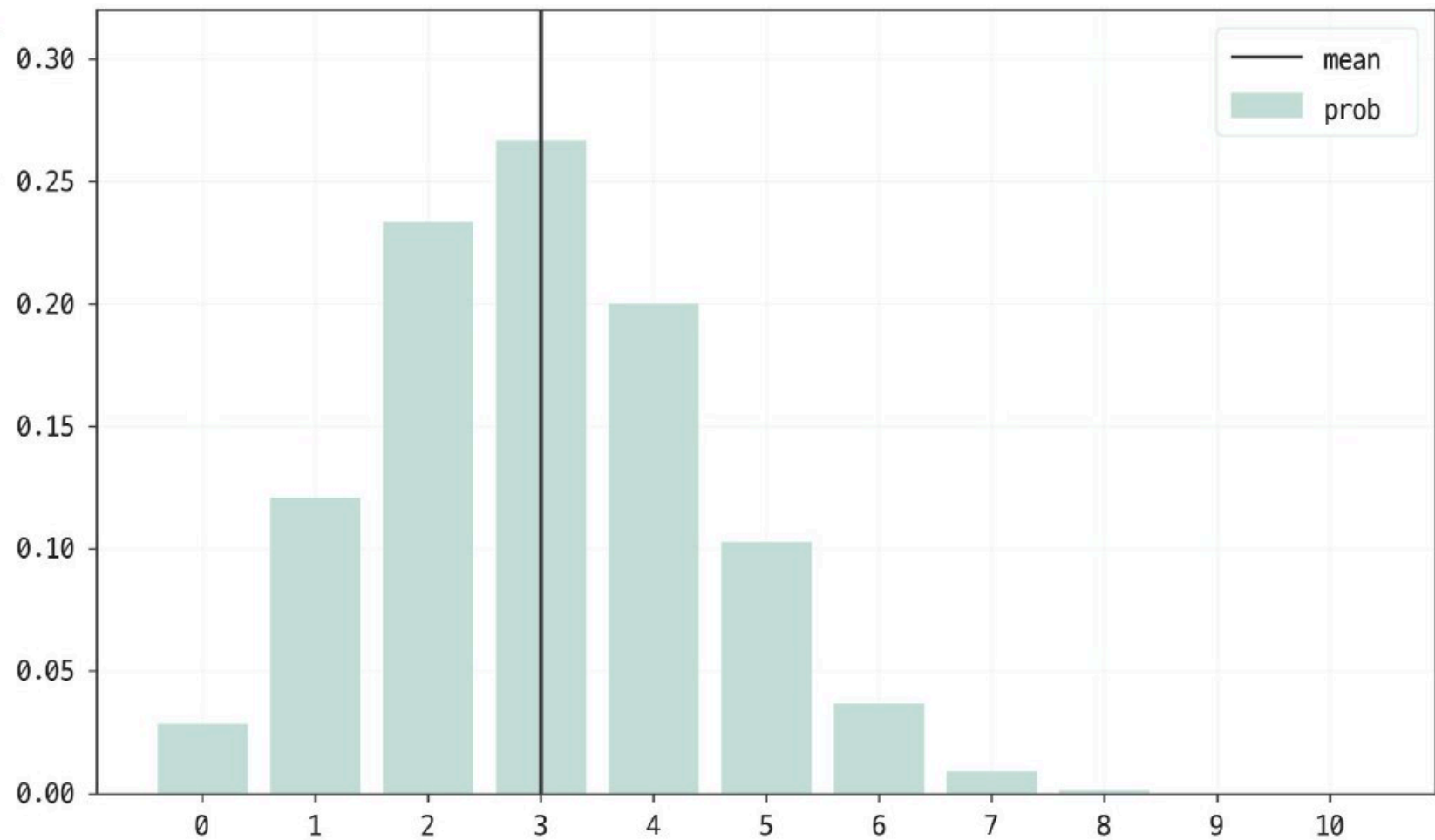
```
check_prob(X)
```

Out [14] :

```
expected value 3.0  
variance 2.1
```

In [15] :

```
plot_prob(X)
```



[그림 6-2] 이항분포

6.2 이항 분포

- scipy.stats의 binom 함수는 이항분포를 따르는 확률변수를 생성

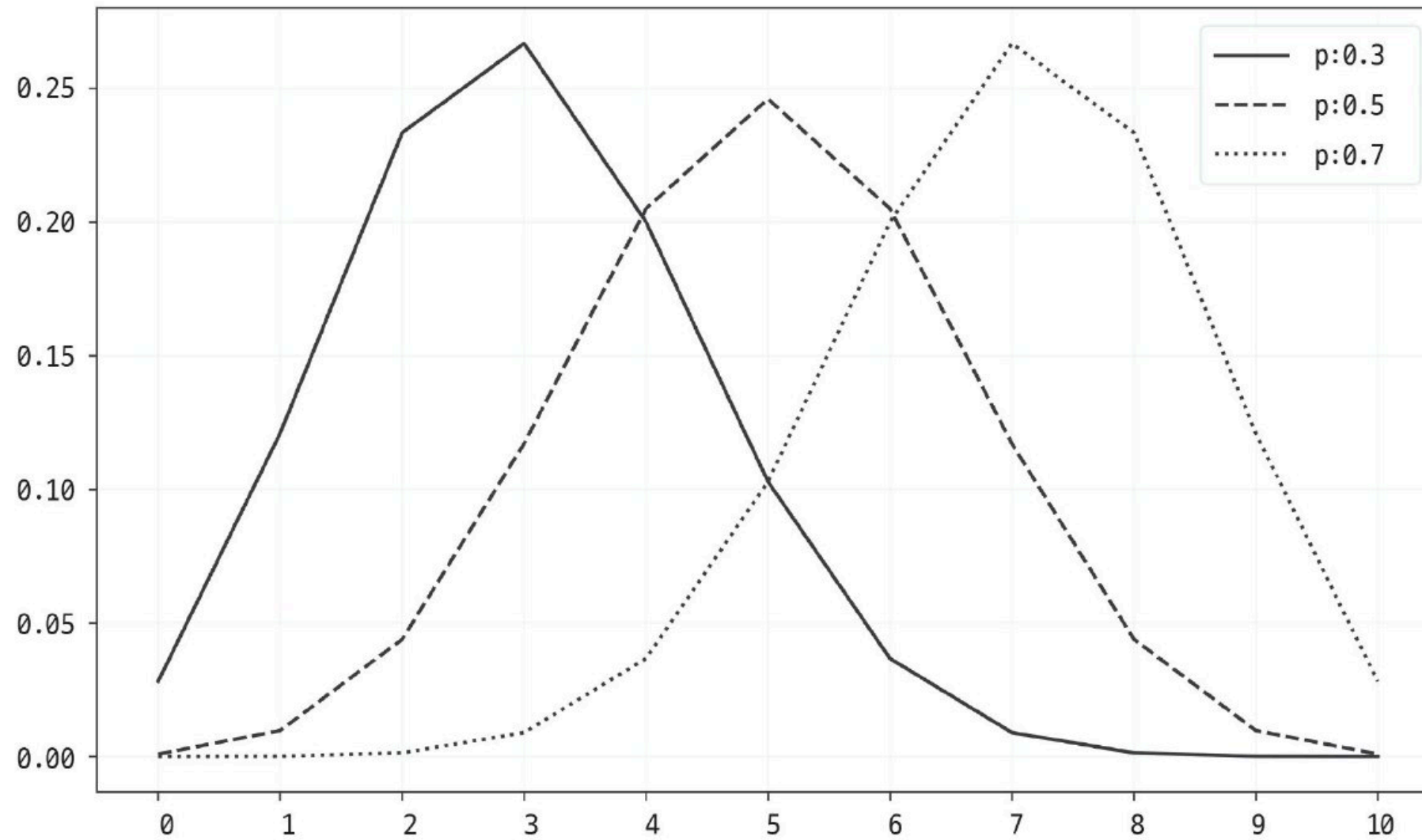
In [16]:

```
fig = plt.figure(figsize=(10, 6))
ax = fig.add_subplot(111)

x_set = np.arange(n+1)
for p, ls in zip([0.3, 0.5, 0.7], linestyle):
    rv = stats.binom(n, p)
    ax.plot(x_set, rv.pmf(x_set),
            label=f' p:{p} ', ls=ls, color=' gray ')
ax.set_xticks(x_set)
ax.legend()

plt.show()
```

6.2 이항 분포



[그림 6-3] 다양한 이항분포 [SAMPLE CODE](#)

6.2 이항 분포

[표 6-2] 이항분포의 정리

파라미터	n, p
취할 수 있는 값	$\{0, 1, \dots, n\}$
확률함수	${}_nC_x p^x (1-p)^{(1-x)}$
기댓값	np
분산	$np(1-p)$
scipy.stats	<code>binom(n, p)</code>

다음 중 이항분포를 따르는 것은?

확률변수 X 를 이렇게 정했습니다. 이항분포 $\text{Bin}(n, p)$ 로 나타낼 수 있는 것은?

- ① 동전을 앞면이 나올 때까지 던진 횟수
- ② 손님이 서로 독립적으로 아무 때나 들어오는 가게의 한 시간 손님 수
- ③ 공정한 동전을 10번 던져 앞면이 나온 횟수
- ④ 주사위를 한 번 굴려 나온 눈의 수

다음 중 이항분포를 따르는 것은?

3번 — $\text{Bin}(10, 1/2)$

이항분포는 성공 확률이 p 인 시행을 서로 독립으로 정해진 n 번 했을 때 성공한 횟수입니다. 1번은 던지는 횟수가 정해져 있지 않으므로 기하분포, 2번은 정해진 시행 횟수 없이 단위 시간 동안 독립적으로 아무 때나 일어난 횟수이므로 포아송 분포로 봅니다. 4번은 성공·실패로 가르는 시행이 아니라 1 부터 6 까지 고르게 나오는 값이라 이항분포가 아닙니다.

6.3 기하분포

- 베르누이 시행에서 처음 성공할 때까지 반복한 시행 횟수가 따르는 분포
- 파라미터 p 인 기하분포는 $Ge(p)$

기하분포의 확률함수

$$f(x) = \begin{cases} (1-p)^{(x-1)}p & (x \in \{1, 2, 3, \dots\}) \\ 0 & (otherwise) \end{cases}$$

6.3 기하분포

- 동전을 다섯 번째 던져 처음으로 앞면이 나오는 확률

$$P(X = 5) = \left(1 - \frac{1}{2}\right)^4 \times \frac{1}{2} = \frac{1}{32}$$

- 주사위를 세 번째 굴려 처음으로 6이 나오는 확률

$$P(X = 3) = \left(1 - \frac{1}{6}\right)^2 \times \frac{1}{6} = \frac{25}{216}$$

6.3 기하분포

기하분포의 기댓값과 분산

$X \sim Ge(p)$ 라고 할 때

$$E(X) = \frac{1}{p}, \quad V(X) = \frac{(1-p)}{p^2}$$

In [17]:

```
def Ge(p):  
    x_set = np.arange(1, 30)  
    def f(x):  
        if x in x_set:  
            return p * (1-p) ** (x-1)  
        else:  
            return 0  
    return x_set, f
```

– $X \sim Ge(1/2)$

In [18]:

```
p = 0.5  
X = Ge(p)
```

6.3 기하분포

- 기댓값은 $\frac{1}{1/2} = 2$, 분산은 $\frac{1-1/2}{(1/2)^2} = 2$

In [19] :

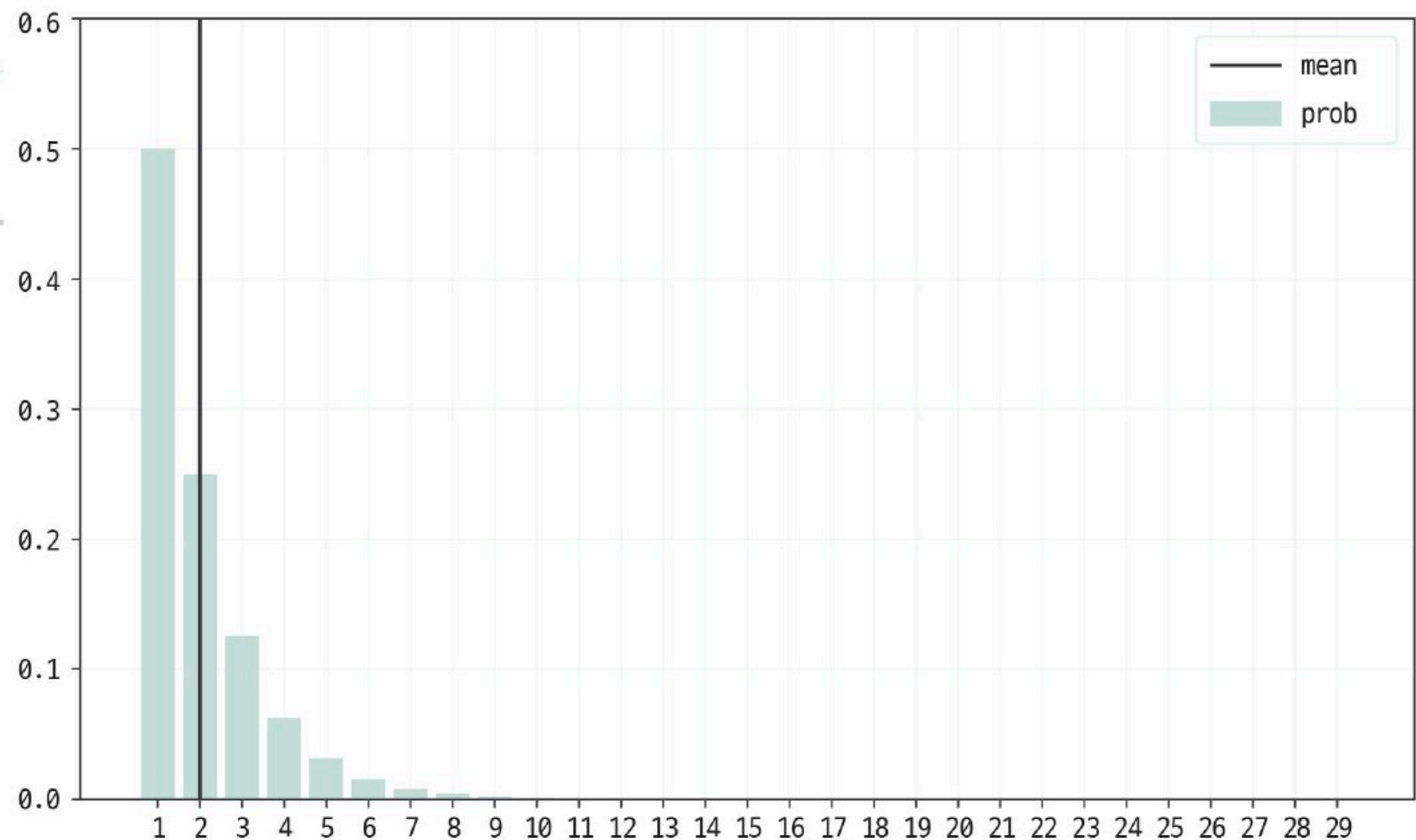
```
check_prob(X)
```

Out [19] :

```
expected value 2.0  
variance 2.0
```

In [20] :

```
plot_prob(X)
```



[그림 6-4] 기하분포

6.3 기하분포

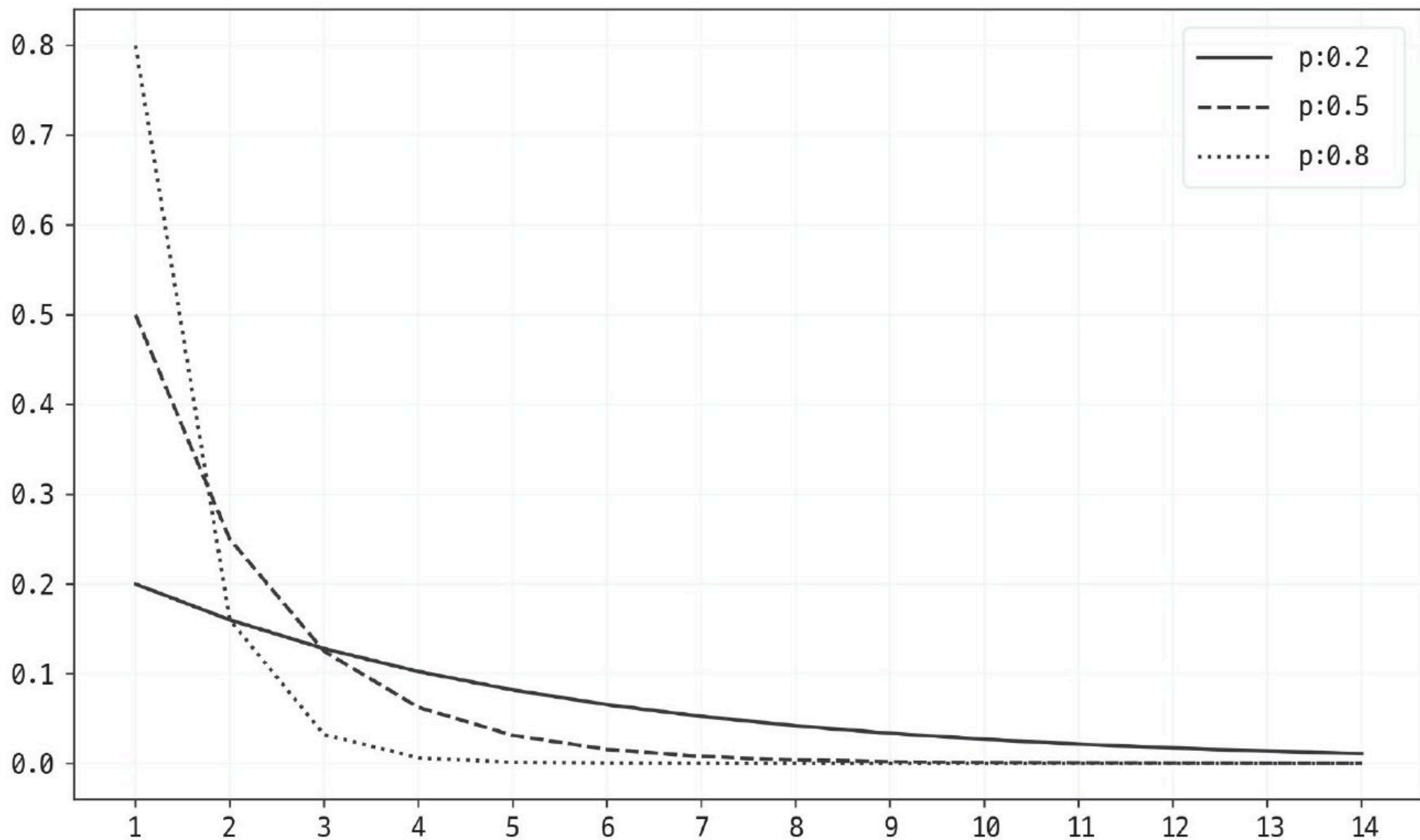
In [21] :

```
fig = plt.figure(figsize=(10, 6))
ax = fig.add_subplot(111)

x_set = np.arange(1, 15)
for p, ls in zip([0.2, 0.5, 0.8], linestyle):
    rv = stats.geom(p)
    ax.plot(x_set, rv.pmf(x_set),
            label=f ' p:{p} ', ls=ls, color= ' gray ' )
ax.set_xticks(x_set)
ax.legend()

plt.show()
```

6.3 기하분포



[그림 6-5] 다양한 기하분포 [SAMPLE CODE](#)

6.3 기하분포

[표 6-3] 기하분포의 정리

파라미터	p
취할 수 있는 값	$\{1, 2, 3, \dots\}$
확률함수	$(1 - p)^{(x-1)}p$
기댓값	$\frac{1}{p}$
분산	$\frac{(1 - p)}{p^2}$
scipy.stats	geom(p)

$X \sim \text{Ge}(1/2)$ 에서 $P(X = 3)$ 은?

앞면이 나올 확률이 $1/2$ 인 동전을, 처음 앞면이 나올 때까지 던집니다. 세 번째에 처음 앞면이 나올 확률은?

① $1/8$

② $1/4$

③ $3/8$

④ $1/2$

$X \sim \text{Ge}(1/2)$ 에서 $P(X = 3)$ 은?

기하분포의 확률함수 $(1-p)^{x-1} p$

$$P(X = 3) = (1 - p)^2 p = \left(\frac{1}{2}\right)^2 \times \frac{1}{2} = \frac{1}{8}$$

1번 — 1/8

세 번째에 처음 성공하려면 앞의 두 번은 실패하고 세 번째에 성공해야 합니다. 실패 확률을 두 번, 성공 확률을 한 번 곱합니다.

6.4 포아송 분포

- 임의의 사건이 단위 시간당 발생하는 건수가 따르는 확률분포
- 파라미터 λ 인 포아송 분포는 $\text{Poi}(\lambda)$

포아송 분포의 확률함수

$$f(x) = \begin{cases} \frac{\lambda^x}{x!} \cdot e^{-\lambda} & (x \in \{0, 1, 2, \dots\}) \\ 0 & (otherwise) \end{cases}$$

- $\text{Poi}(\lambda)$ 는 단위 시간당 평균 λ 번 발생하는 임의의 사건이 단위 시간에 발생하는 건수가 따르는 분포

6.4 포아송 분포

- 하루에 평균 2건의 교통사고가 발생(Poi(2))하는 지역에서 하루에 교통사고가 한 건도 일어나지 않을 확률

$$P(X=0) = \frac{2^0}{0!} \cdot e^{-2} \simeq 0.135$$

- **한 시간당** 평균 10번 액세스(Poi(10))하는 사이트에서 한 시간에 15건 액세스 확률

$$P(X=15) = \frac{10^{15}}{15!} \cdot e^{-10} \simeq 0.035$$

6.4 포아송 분포

포아송 분포의 기댓값과 분산

$X \sim Poi(\lambda)$ 라고 할 때

$$E(X) = \lambda, \quad V(X) = \lambda$$

In [22] :

```
from scipy.special import factorial

def Poi(lam):
    x_set = np.arange(20)
    def f(x):
        if x in x_set:
            return np.power(lam, x) / factorial(x) * np.exp(-lam)
        else:
            return 0
    return x_set, f
```

– $X \sim Poi(3)$

In [23] :

```
lam = 3
X = Poi(lam)
```

- 기댓값과 분산은 모두 3

In [24] :

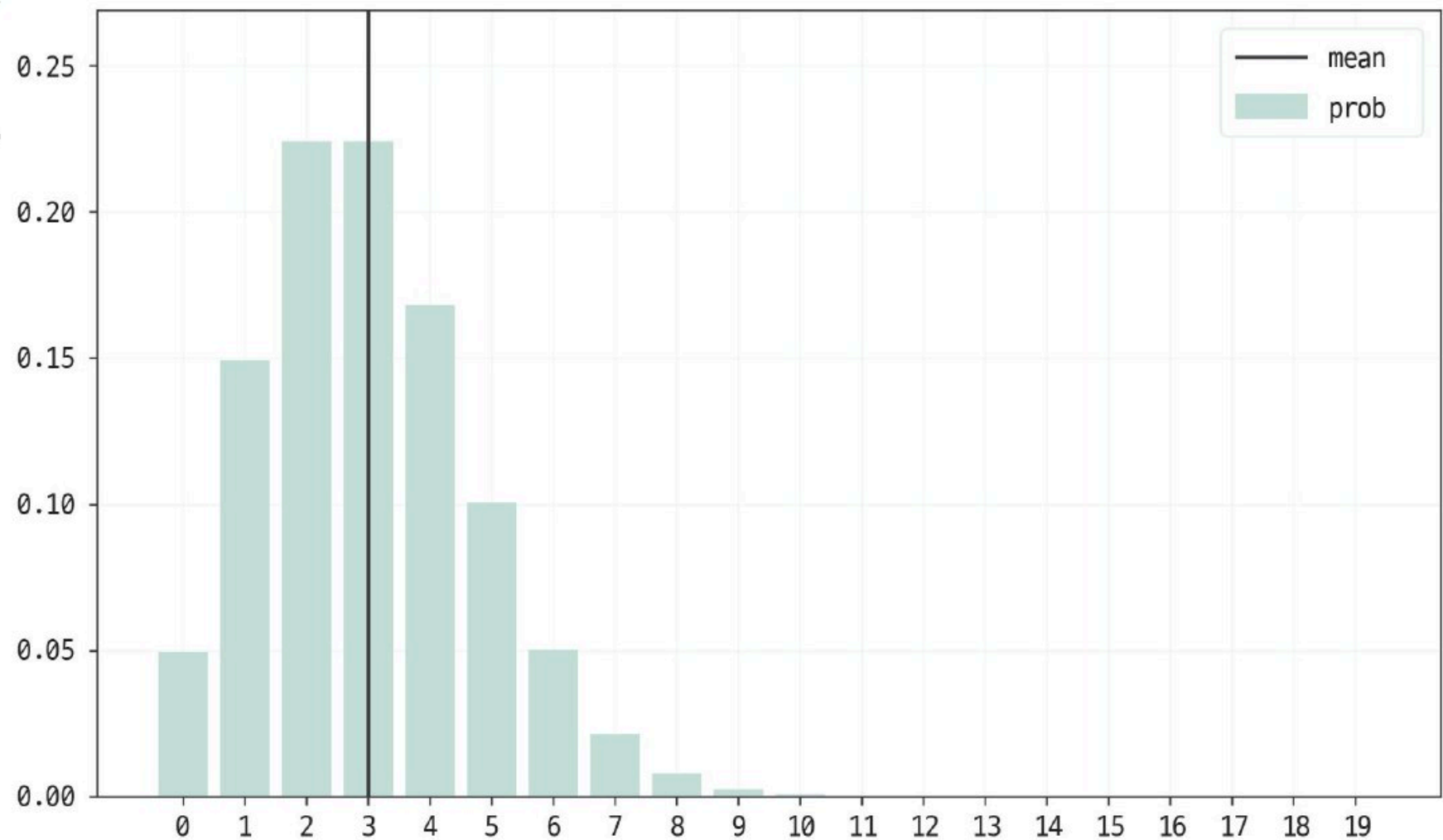
```
check_prob(X)
```

Out [24] :

```
expected value 3.0  
variance 3.0
```

In [25] :

```
plot_prob(X)
```



[그림 6-6] 포아송 분포

6.4 포아송 분포

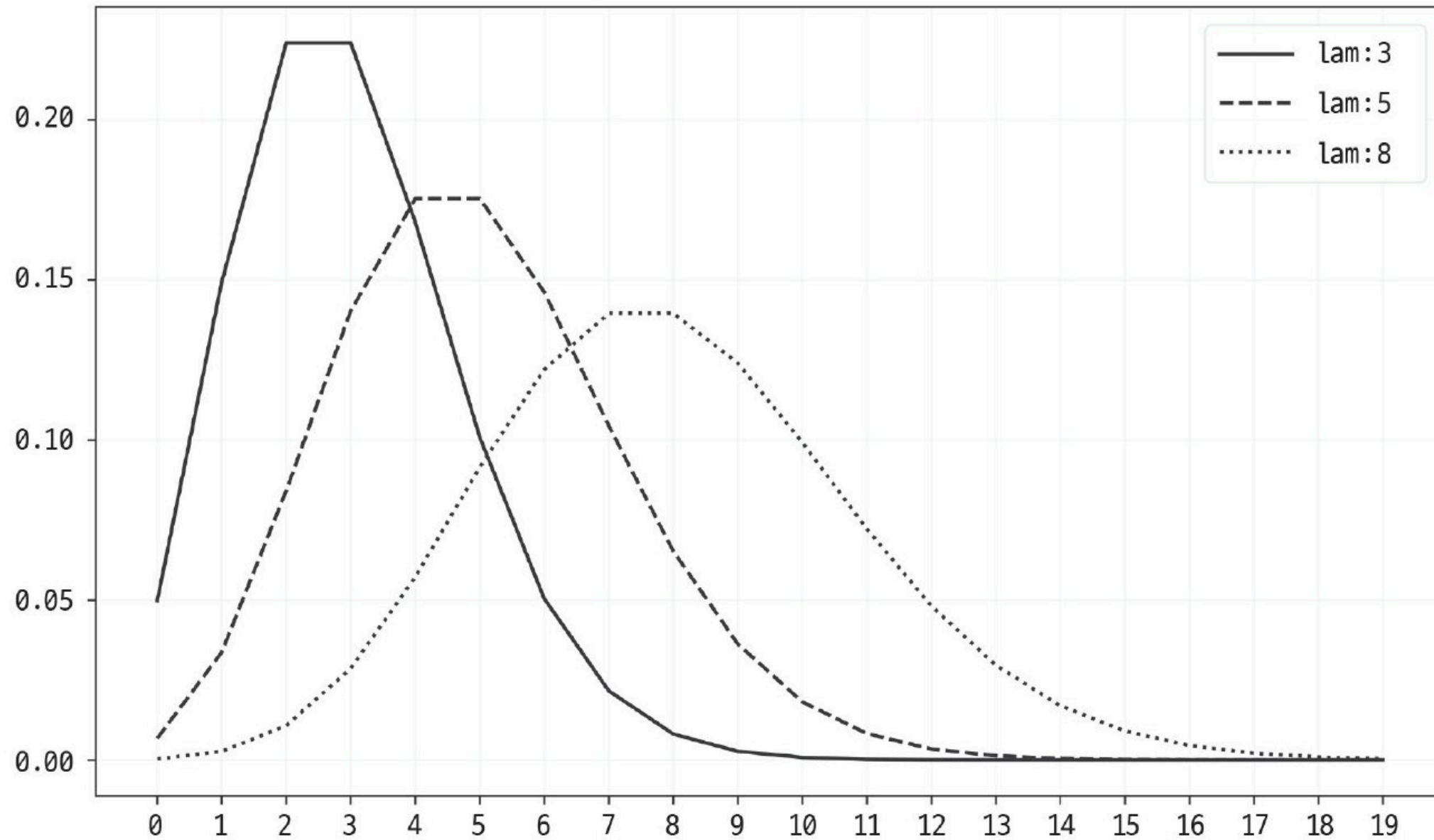
In [26] :

```
fig = plt.figure(figsize=(10, 6))
ax = fig.add_subplot(111)

x_set = np.arange(20)
for lam, ls in zip([3, 5, 8], linestyle):
    rv = stats.poisson(lam)
    ax.plot(x_set, rv.pmf(x_set),
            label=f' lam:{lam} ', ls=ls, color=' gray ' )
ax.set_xticks(x_set)
ax.legend()

plt.show()
```

6.4 포아송 분포



[그림 6-7] 다양한 포아송 분포 [SAMPLE CODE](#)

6.4 포아송 분포

[표 6-4] 포아송 분포의 정리

파라미터	λ
취할 수 있는 값	$\{0, 1, 2, \dots\}$
확률함수	$\frac{\lambda^x}{x!} \cdot e^{-\lambda}$
기댓값	λ
분산	λ
scipy.stats	<code>poisson(λ)</code>

평균 3번 일어나는 사건이 한 번도 안 일어날 확률은?

단위 시간에 평균 3번 일어나는 사건 $X \sim \text{Poi}(3)$ 에서 $P(X = 0)$ 을 구하세요.

- ① 0
- ② e^{-3} (약 0.05)
- ③ $3e^{-3}$ (약 0.15)
- ④ $1/3$

평균 3번 일어나는 사건이 한 번도 안 일어날 확률은?

포아송 분포의 확률함수에 $x = 0$

$$P(X = 0) = \frac{3^0}{0!} e^{-3} = e^{-3} \approx 0.05$$

2번 — e^{-3}

$x = 0$ 을 넣으면 λ^0 과 $0!$ 이 모두 1 이라 $e^{(-\lambda)}$ 만 남습니다. 평균이 3 이어도 한 번도 안 일어날 확률은 0 이 아닙니다. $3e^{-3}$ 은 $P(X = 1)$ 입니다.

확인 문제 — 어느 분포인지 고르고 계산하기

시험에 이런 형태로 나올 수 있는 문제입니다. 문제를 띄우면 답 장으로 넘기기 전에 먼저 손으로 풀어 보세요.

5 다음 상황은 어느 분포인가?

6 $\text{Bin}(10, 0.3)$ 의 기댓값과 분산은?

7 주사위를 4번 굴려 6 이 한 번도 안 나올 확률은?

8 6 이 처음 나올 때까지 평균 몇 번 굴리나?

9 포아송 분포의 분산이 4 이면 기댓값은?

다음 상황은 어느 분포인가?

'불량률이 2% 인 공장에서, 처음으로 불량품이 나올 때까지 검사한 제품 수'

- ① 베르누이 분포
- ② 이항분포
- ③ 기하분포
- ④ 포아송 분포

다음 상황은 어느 분포인가?

3번 — 기하분포 $Ge(0.02)$

'처음 성공(여기서는 불량품 발견)할 때까지의 시행 횟수'가 기하분포입니다. 검사 수 n 이 정해져 있고 그 안의 불량품 개수를 물었다면 이항분포입니다.

Bin(10, 0.3) 의 기댓값과 분산은?

성공 확률 0.3 인 시행을 10번 합니다.

Bin(10, 0.3) 의 기댓값과 분산은?

이항분포 공식

$$E = np = 10 \times 0.3 = 3, \quad V = np(1 - p) = 10 \times 0.3 \times 0.7 = 2.1$$

$$E(X) = 3, V(X) = 2.1$$

10번 중 평균 3번 성공합니다. 분산은 np 에 실패 확률 0.7 을 한 번 더 곱합니다.

주사위를 4번 굴려 6 이 한 번도 안 나올 확률은?

$X \sim \text{Bin}(4, 1/6)$ 에서 $P(X = 0)$ 을 구하세요.

주사위를 4번 굴려 6 이 한 번도 안 나올 확률은?

풀이

$$P(X = 0) = \binom{4}{0} \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^4 = \frac{625}{1296}$$

$$(5/6)^4 = 625/1296 \approx 0.48$$

$x = 0$ 이면 조합은 1, p^0 도 1 이라 실패 확률만 네 번 곱하면 됩니다. 네 번이나 굴려도 6 이 안 나올 확률이 절반 가까이 됩니다.

6 이 처음 나올 때까지 평균 몇 번 굴리나?

공정한 주사위를 6 이 나올 때까지 굴립니다.

6 이 처음 나올 때까지 평균 몇 번 굴리나?

6번 — $Ge(1/6)$ 의 기댓값 $1/p$

성공 확률이 $1/6$ 이면 평균 6번째에 처음 성공합니다. 확률이 작을수록 오래 걸린다는 직관과 맞습니다.

포아송 분포의 분산이 4 이면 기댓값은?

- ① 2
- ② 4
- ③ 16
- ④ 알 수 없다

포아송 분포의 분산이 4 이면 기댓값은?

2번 — 4

포아송 분포는 기댓값과 분산이 모두 λ 입니다. 분산이 4 이면 $\lambda = 4$ 이고 기댓값도 4 입니다. 표준편차(2)와 헷갈리지 마세요.

계산 연습 — scipy 로 바로 구하기

분포를 정하면 계산은 한 줄입니다. 문제를 읽고 어떤 코드를 쓸지, 값이 얼마인지 번갈아 고릅니다.

분포	만드는 코드	정확히 K 일 확률	기댓값
이항 Bin(n, p)	stats.binom(n, p)	.pmf(k)	np
기하 Ge(p)	stats.geom(p)	.pmf(k)	1/p
포아송 Poi(λ)	stats.poisson(λ)	.pmf(k)	λ

기댓값·분산은 .mean(), .var() 로도 구합니다. k 이하일 확률은 .cdf(k) 입니다.

동전을 8번 던져 앞면이 정확히 4번 나올 확률 — 코드는?

공정한 동전을 8번 던집니다. 앞면이 정확히 4번 나올 확률을 구하는 코드를 고르세요.

- ① `stats.binom(8, 0.5).pmf(4)`
- ② `stats.binom(4, 0.5).pmf(8)`
- ③ `stats.geom(0.5).pmf(4)`
- ④ `stats.binom(8, 0.5).cdf(4)`

동전을 8번 던져 앞면이 정확히 4번 나올 확률 — 코드는?

1번 — `stats.binom(8, 0.5).pmf(4)`

던지는 횟수 8 이 정해져 있고 앞면(성공) 횟수를 세니 이항분포 $\text{Bin}(8, 0.5)$ 입니다. `binom` 에는 n, p 순서로 넣습니다. `cdf(4)` 는 '4번 이하' 의 확률이라 다릅니다.

`stats.binom(8, 0.5).pmf(4)` 의 값은?

방금 고른 코드를 실행하면 얼마가 나올까요?

- ① 0.5
- ② 0.273
- ③ 0.125
- ④ 0.637

stats.binom(8, 0.5).pmf(4) 의 값은?

2번 — 0.273

${}_8C_4 \times (1/2)^8 = 70/256 \approx 0.273$ 입니다. 8번 중 딱 절반이 나올 확률이 0.5 가 아니라는 점이 직관과 다릅니다. 0.637 은 cdf(4), 곧 4번 이하일 확률입니다.

주사위를 두 번째에 처음 6 이 나올 확률 — 코드는?

주사위를 굴립니다. 첫 번째는 6 이 아니고, 두 번째에 처음으로 6 이 나올 확률을 구하는 코드를 고르세요.

- ① `stats.binom(2, 1/6).pmf(1)`
- ② `stats.geom(2).pmf(1/6)`
- ③ `stats.geom(1/6).pmf(2)`
- ④ `stats.poisson(1/6).pmf(2)`

주사위를 두 번째에 처음 6 이 나올 확률 — 코드는?

3번 — `stats.geom(1/6).pmf(2)`

'처음 성공할 때까지의 횟수' 는 기하분포입니다. `geom` 에는 성공 확률 p 를 넣고, `pmf` 에 몇 번째인지를 넣습니다. 1번은 '두 번 중 한 번' 이라 첫 번째에 6 이 나온 경우까지 셉니다.

stats.geom(1/6).pmf(2) 의 값은?

실패 한 번, 그다음 성공 한 번의 확률입니다.

- ① 0.167
- ② 0.028
- ③ 0.278
- ④ 0.139

stats.geom(1/6).pmf(2) 의 값은?

4번 — 0.139

$(5/6) \times (1/6) = 5/36 \approx 0.139$ 입니다. 0.028 은 $(1/6)^2$ 으로 두 번 다 6 이 나온 확률, 0.278 은 이항분포로 셋을 때 $(2 \times 5/36)$ 입니다.

한 시간 평균 4명 오는 가게에 정확히 2명 — 코드는?

손님은 서로 상관없이 한 시간에 평균 4명 옵니다. 한 시간 동안 정확히 2명 올 확률을 구하는 코드를 고르세요.

- ① `stats.poisson(2).pmf(4)`
- ② `stats.poisson(4).pmf(2)`
- ③ `stats.binom(4, 0.5).pmf(2)`
- ④ `stats.poisson(4).cdf(2)`

한 시간 평균 4명 오는 가게에 정확히 2명 — 코드는?

2번 — `stats.poisson(4).pmf(2)`

단위 시간당 평균 횟수가 주어졌으니 포아송분포입니다. poisson 에는 평균 $\lambda = 4$ 를, pmf 에는 알고 싶은 횟수 2 를 넣습니다. 1번은 둘을 바꿔 넣었고, 4번은 '2명 이하' 입니다.

stats.poisson(4).pmf(2) 의 값은?

$4^2 / 2! \times e^{-4}$ 입니다.

- ① 0.147
- ② 0.238
- ③ 0.5
- ④ 0.018

stats.poisson(4).pmf(2) 의 값은?

1번 — 0.147

$16/2 \times e^{-4} = 8e^{-4} \approx 0.147$ 입니다. 0.238 은 cdf(2), 곧 0·1·2명을 모두 더한 값이고, 0.018 은 한 명도 안 올 확률 e^{-4} 입니다.

Bin(20, 0.1) 의 기댓값은?

불량률 10% 인 제품 20개 중 불량품 수입니다. `stats.binom(20, 0.1).mean()` 의 값을 고르세요.

- ① 0.1
- ② 1.8
- ③ 2
- ④ 20

Bin(20, 0.1) 의 기댓값은?

3번 — 2

이항분포의 기댓값은 $np = 20 \times 0.1 = 2$ 입니다. 1.8 은 분산 $np(1 - p) = 20 \times 0.1 \times 0.9$ 입니다.

Ge(0.25)의 기댓값은?

성공률 25%인 뽑기를 처음 성공할 때까지 합니다. `stats.geom(0.25).mean()`의 값을 고르세요.

① 0.25

② 3

③ 4

④ 12

Ge(0.25)의 기댓값은?

3번 — 4

기하분포의 기댓값은 $1/p = 1/0.25 = 4$ 입니다. 평균 네 번째에 처음 성공합니다. 12는 분산 $(1-p)/p^2$ 입니다.

pmf 와 cdf — `stats.binom(3, 0.5).cdf(1)` 은?

동전을 3번 던진 앞면 수 X 에 대해 $rv = \text{stats.binom}(3, 0.5)$ 입니다. $rv.cdf(1)$ 의 값을 고르세요.

- ① 0.125
- ② 0.375
- ③ 0.5
- ④ 0.875

pmf 와 cdf — `stats.binom(3, 0.5).cdf(1)` 은?

3번 — 0.5

`cdf(1)` 은 $P(X \leq 1) = P(X = 0) + P(X = 1) = 1/8 + 3/8 = 0.5$ 입니다. 0.375 는 `pmf(1)`, 곧 정확히 1번일 확률입니다. '이하' 는 `cdf`, '정확히' 는 `pmf` 입니다.

OX 문제 10개를 찍어 정확히 7개 맞힐 확률은?

분포를 정하고 코드 한 줄로 구해 보세요.

- ① 0.117
- ② 0.7
- ③ 0.172
- ④ 0.945

OX 문제 10개를 찍어 정확히 7개 맞힐 확률은?

1번 — 0.117

문제 수 10 이 정해져 있고 맞힌 개수를 세니 $\text{Bin}(10, 0.5)$ 입니다. $\text{stats.binom}(10, 0.5).\text{pmf}(7) = {}_{10}C_7 / 2^{10} = 120/1024 \approx 0.117$ 입니다. 0.945 는 $\text{cdf}(7)$, 곧 7개 이하일 확률입니다.

Q&A