확률 및 통계

Assignment 2

2024.04.22.

과제 범위

- **Assignment #2 (~4/22)**
  - 연습 10. Problem 3.2
  - 연습 11. Problem 4.11
  - 연습 12. Problem 4.12
  - 연습 13. Problem 6.1
  - 연습 14. Problem 4.43
  - 연습 15. Problems 4.49
  - 실습 3. 히스토그램 생성 (NumPy)
  - 실습 4. Poisson 확률분포
  - 실습 5. Binomial 분포에 대한 중심극한정리

연습 10. Problems 3.2

Problems

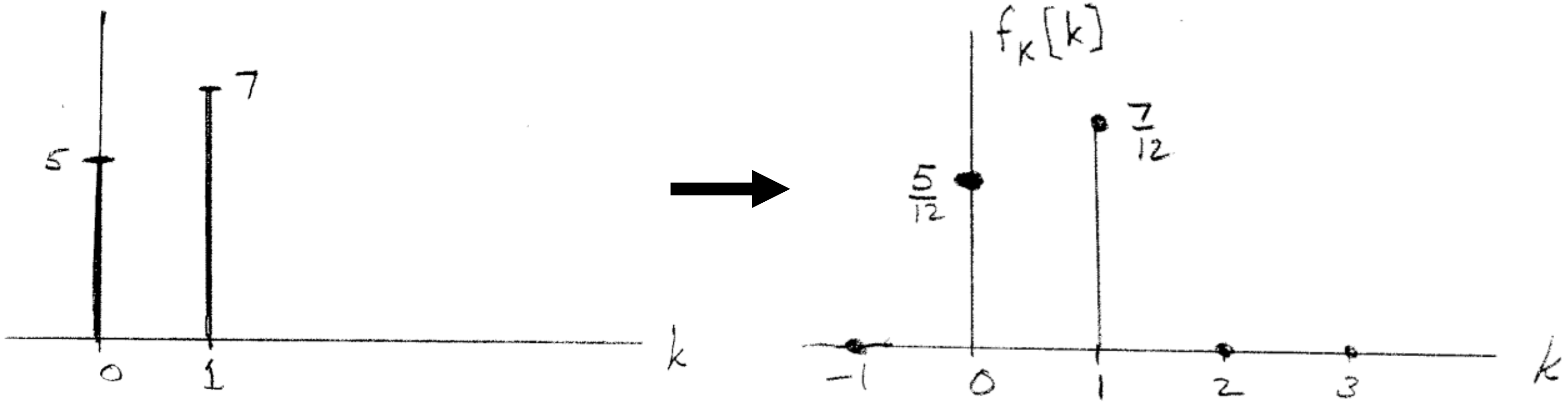
Discrete random variables

3.1 In a set of independent trials, a certain discrete random variable K takes on the values

$$1,\ 0,\ 0,\ 0,\ 1,\ 1,\ 0,\ 1,\ 1,\ 1,\ 0,\ 1$$

(a) Plot a histogram for the random variable.

(b) Normalize the histogram to provide an estimate of the PMF of K.



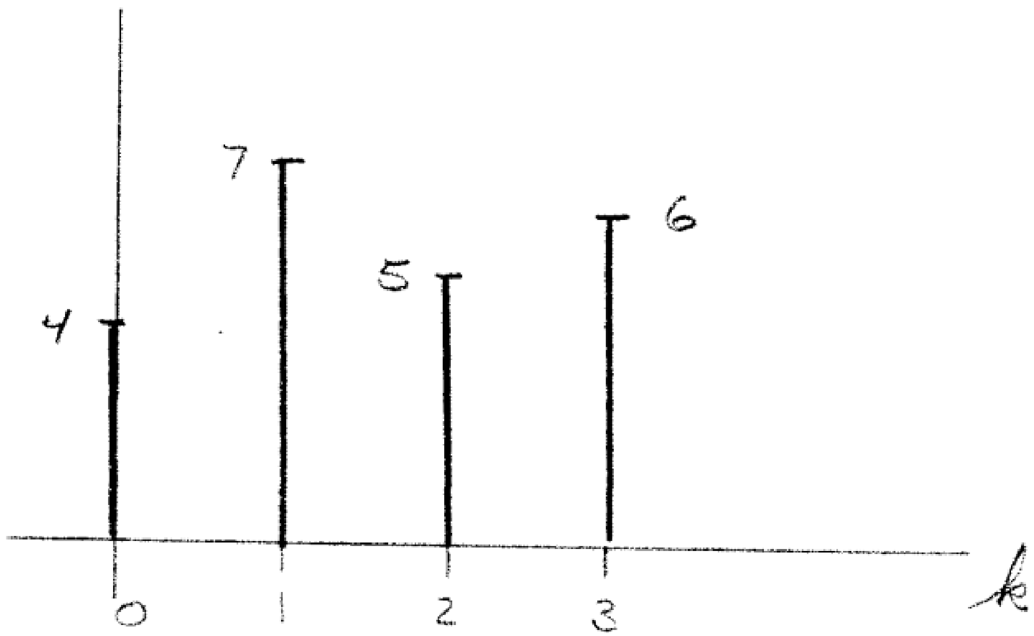
연습 10. Problems 3.2

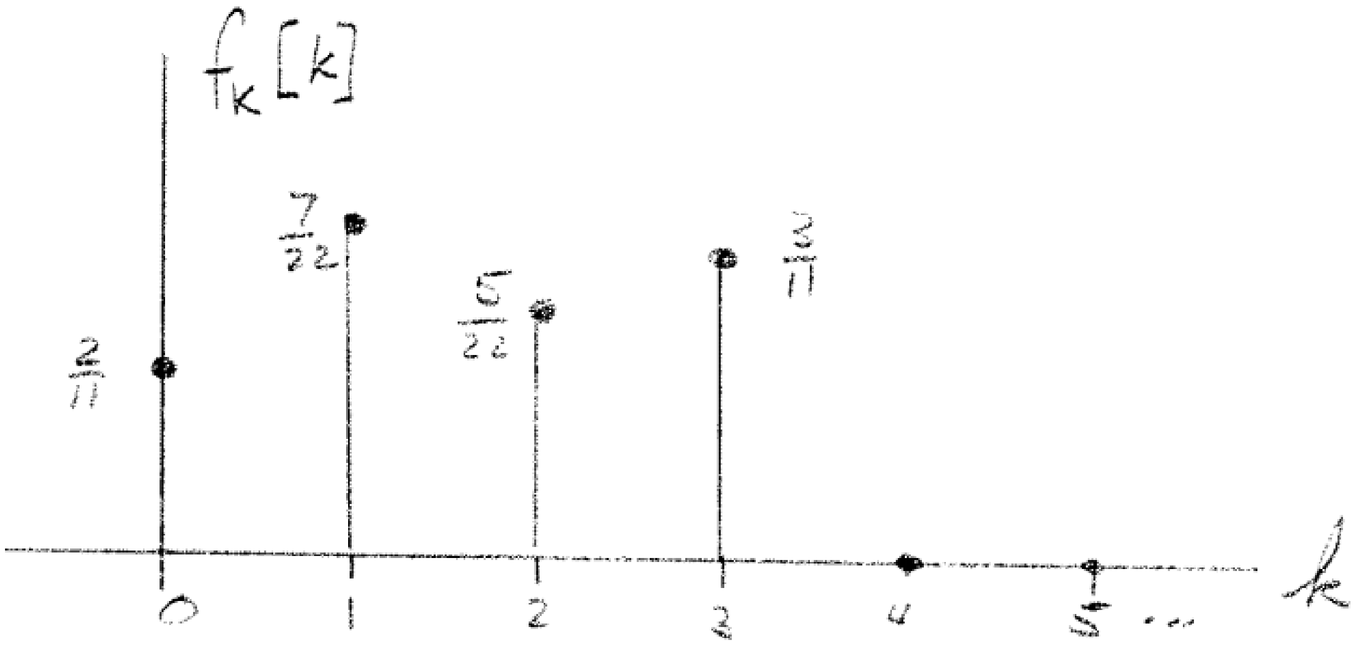
3.2 Repeat Problem 3.1 for the discrete random variable J that takes on the values

$$1,\ 3,\ 0,\ 2,\ 1,\ 2,\ 0,\ 3,\ 1,\ 1,\ 3,\ 1,\ 0,\ 1,\ 2,\ 3,\ 0,\ 2,\ 1,\ 3,\ 3,\ 2$$

(a) Plot a histogram for the random variable.

(b) Normalize the histogram to provide an estimate of the PMF of K.



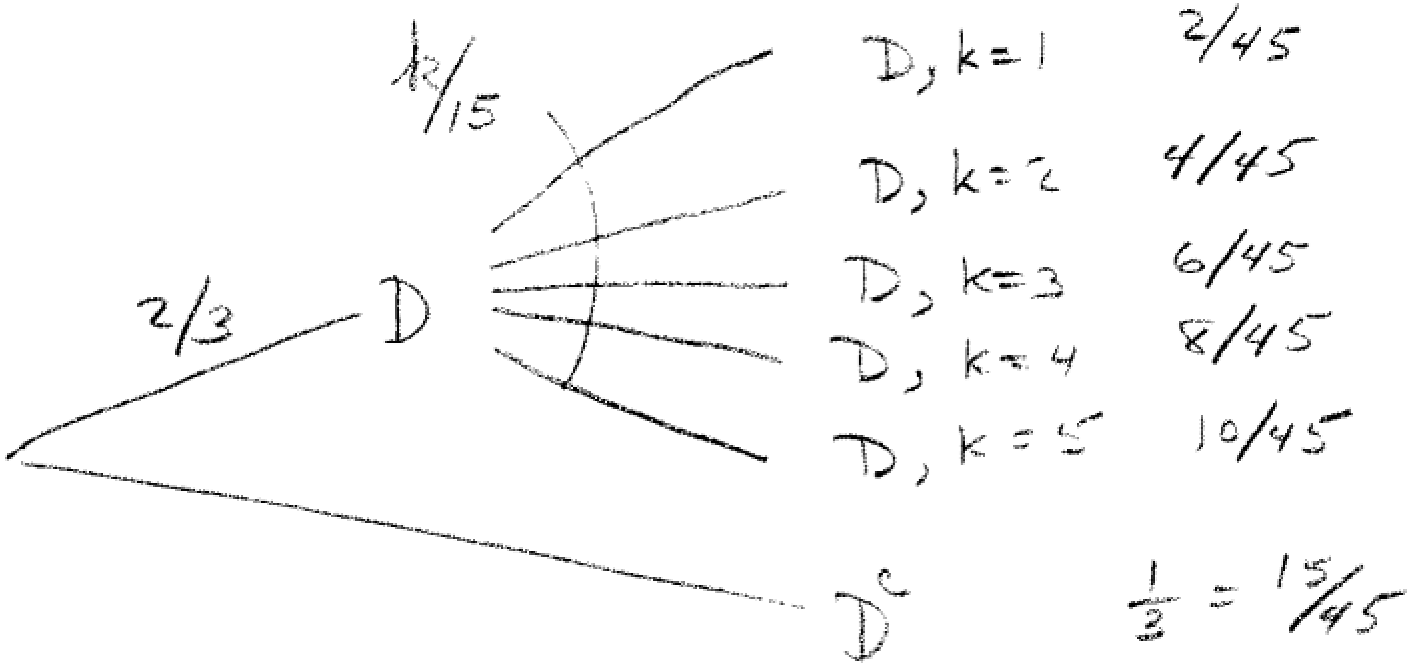


연습 11. Problems 4.11

4.11 Claudia is a student at the Technical University majoring in statistics. On any particular day of the week, her friend Rolf may ask her for a Saturday night date with probability 3/4. Claudia, however, is more attracted to Roberto de la Dolce, who is very handsome, drives an expensive Italian car, and really knows how to treat women! Roberto has other women he likes however, so the probability that he asks Claudia out in any particular week is only 2/3. Roberto is also very self-confident and does not plan his activities early in the week. Let D represent the event that Roberto asks Claudia for a Saturday night date. Then the day of the week on which he asks her (beginning on Monday) is a random variable K with PMF $\mathrm{f}_{K|\mathrm{D}}[k|\mathrm{D}] = k/15$ for $1 \leq k \leq 5$ and 0 otherwise. Claudia is aware of this formula and the probabilities.

연습 11. Problems 4.11

Original sample space for Claudia & Roberto



D is event; Roberto asks Claudia for a date

연습 11. Problems 4.11

Claudia needs to plan whether or not to accept if Rolf asks her for a date before Roberto; thus she decides to rate her emotional state (α) for the week on a scale of 0 to 10. A date with Roberto is actually way off the scale but she assigns it a 10. She further determines that a date with with Rolf is worth 4 points, and a Saturday night without a date is worth -5 points. She decides that if Rolf asks her out on the k^{th} day of the week she will compute the expected value of α given that she accepts and the expected value of α given that she does not accept. Then she will make a choice according to which expected value is larger.

(a) Make a plot of the probability that Roberto does *not* ask Claudia for a date given that he has not asked her by the end of the l^{th} day $1 \leq l \leq 5$.

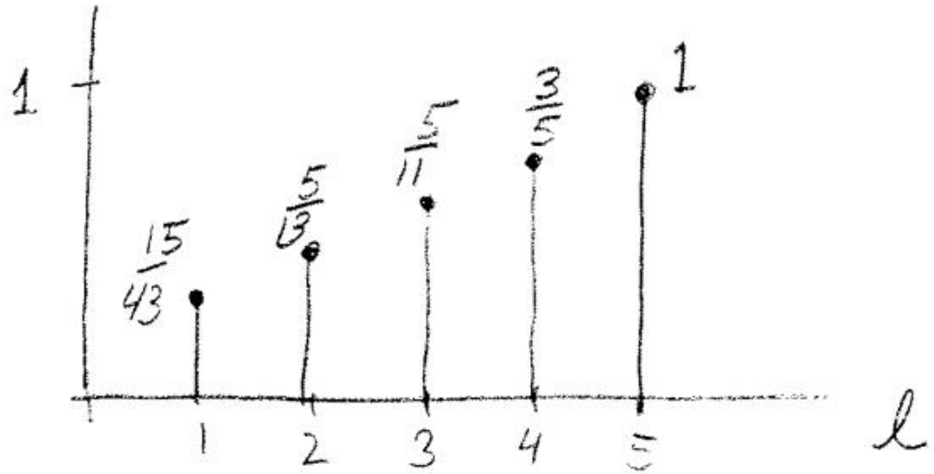
연습 11. Problems 4.11

(a) Define event

B_ℓ : Roberto has not asked Claudia by the end of the ℓ th day $1 \le \ell \le 5$

$$\Pr[B_\ell] = \frac{1}{3} + \sum_{k=\ell+1}^{5} \frac{2}{3}\cdot\frac{k}{15} = 1 - \sum_{k=1}^{\ell} \frac{2}{3}\frac{k}{15}$$

$$\Pr[D^c \mid B_\ell] = \frac{\Pr[D^c B_\ell]}{\Pr[B_\ell]}$$



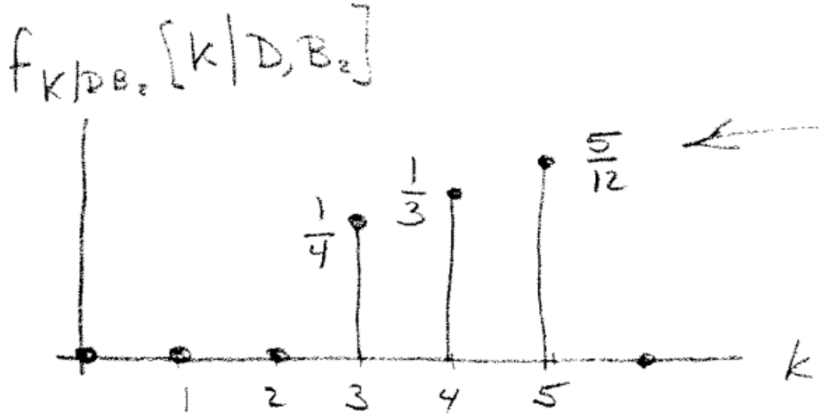
연습 11. Problems 4.11

(b) Sketch the conditional PMF for K given that Roberto asks Claudia out but has not done so by the end of the second day. Given this situation, what is the probability that Roberto asks her out

$$f[k|D,B_2] = \frac{f_{K,D,B_2}[k,D,B_2]}{\Pr[D,B_2]}$$

From the sample space, $\Pr[D,B_2] = \frac{6}{45} + \frac{8}{45} + \frac{10}{45}$

$$= \frac{24}{45} = \frac{8}{15}$$



$$\frac{10}{45} \Big/ \frac{8}{15} = \frac{5}{12}$$

연습 11. Problems 4.11

(b) Sketch the conditional PMF for K given that Roberto asks Claudia out but has not done so by the end of the second day. Given this situation, what is the probability that Roberto asks her out

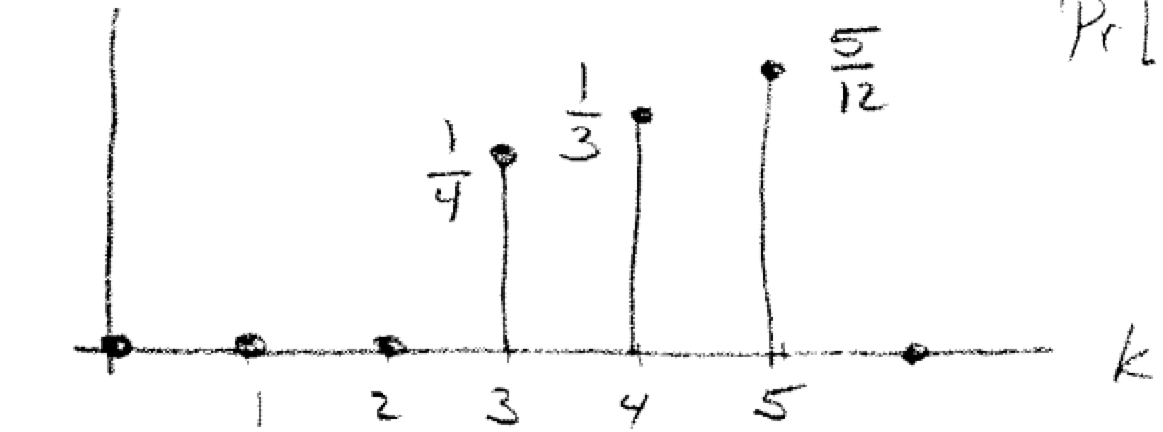
(i) on the third day ($k = 3$)?
(ii) on the fifth day ($k = 5$)?

$$P_r[K=k, D|B_2] = f_{K|DB_2}[k|DB_2] \cdot P_r[D|B_2]$$

$$P_r[K=3, D|B_2] = \frac{1}{4} \cdot \frac{8}{13} = \frac{2}{13}$$

$$P_r[K=5, D|B_2] = \frac{5}{\cancel{12}_{3}} \cdot \frac{\cancel{8}^{2}}{13} = \frac{10}{39}$$

$f_{K|DB_2}[k|D,B_2]$

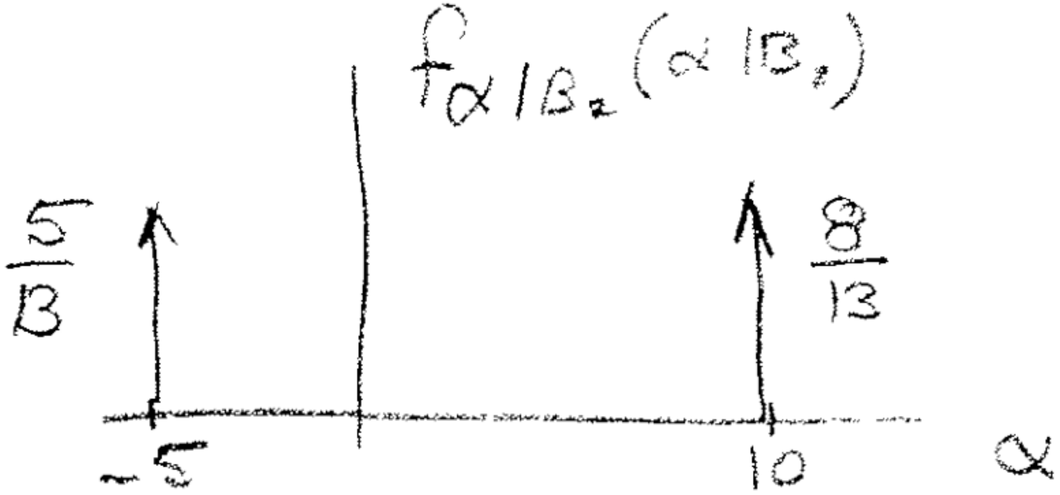


연습 11. Problems 4.11

(c) By the middle of the third day of the week Roberto has not asked Claudia for a date; but Rolf decides to ask her. Will she accept Rolf or not?

We have seen $\Pr[D^c/B_2] = \frac{5}{13}$

$\Pr[D/B_2] = \frac{8}{13}$



--> 확률변수 α에 대해 위 PMF를 그리고 기댓값을 계산

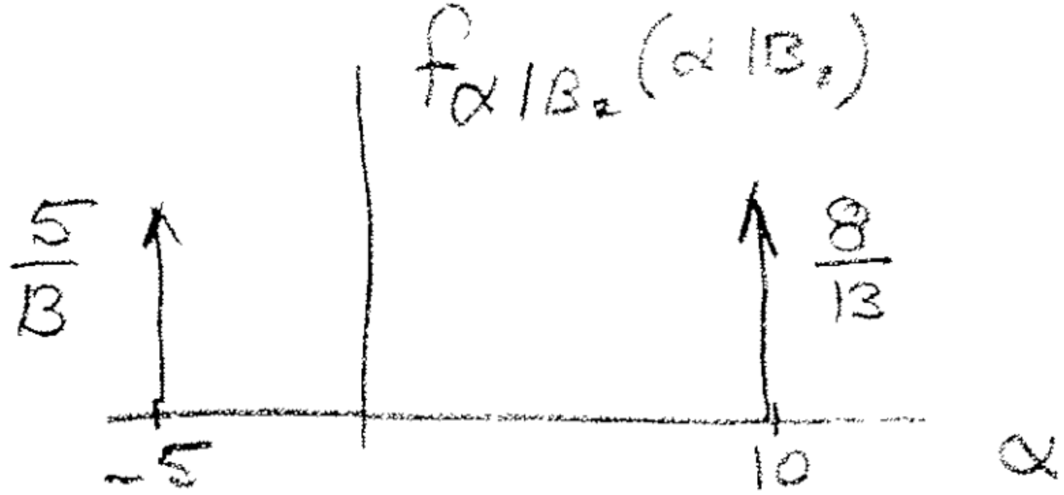
Claudia needs to plan whether or not to accept if Rolf asks her for a date before Roberto; thus she decides to rate her emotional state (α) for the week on a scale of 0 to 10. A date with Roberto is actually way off the scale but she assigns it a 10. She further determines that a date with with Rolf is worth 4 points, and a Saturday night without a date is worth −5 points.

연습 11. Problems 4.11

(c) By the middle of the third day of the week Roberto has not asked Claudia for a date; but Rolf decides to ask her. Will she accept Rolf or not?

We have seen $\Pr[D^c / B_2] = \frac{5}{13}$

$\Pr[D / B_2] = \frac{8}{13}$



$$E\{\alpha\} = (-5)\frac{5}{13} + 10\left(\frac{8}{13}\right) = \frac{-25+80}{13}$$

$\approx 4.231 \longrightarrow$ She will not accept Rolf (=4).

연습 11. Problems 4.11

(d) Rolf has not studied statistics (not even probability), and thinks that his chances for a date with Claudia will be better if he asks her earlier in the week. Is he right or wrong? $\longrightarrow$ He is wrong in his thinking.

(e) What is the optimal strategy for Rolf (i.e., when should he ask Claudia) in order to maximize the probability that Claudia will accept if he asks her for a Saturday night date?

After the 3rd full day of the week we have

$$\Pr[D^c / B_3] = \frac{5}{11} \qquad \Pr[D / B_3] = \frac{6}{11}$$

$$\therefore E\{\alpha\} = (-5)\frac{5}{11} + (10)\frac{6}{11} = \frac{60-25}{11} = 3.182$$

연습 12. Problems 4.12

Problems

4.12 The probability density function of a random variable X is given by

$$f_X(x) = \begin{cases} \frac{1}{2}x, & 0 \leq x \leq 1 \\ \frac{1}{6}(4-x), & 1 < x < c \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the constant c.

(b) Calculate the expectation: $E[3+2X]$.

연습 12. Problems 4.12

(a) Find the constant c.

$$\frac{1}{2}\int_0^1 x\,dx + \frac{1}{6}\int_1^c (4-x)\,dx = 1$$

$$\frac{1}{2}\cdot\frac{x^2}{2}\Big|_0^1 + \frac{1}{6}\left(4x-\frac{x^2}{2}\right)\Big|_1^c = 1$$

$$4c-\frac{c^2}{2}-4+\frac{1}{2} = \frac{9}{2}$$

$$c^2-8c+16=0$$

$$c=4$$

연습 12. Problems 4.12

(b) Calculate the expectation: $E[3+2X]$.

$$E[3+2X] = 3 + 2E[X] = \frac{9+2\cdot 5}{3} = \frac{19}{3}$$

$$E[X] = \frac{1}{2}\int_0^1 x^2\,dx + \frac{1}{6}\int_1^4 (4x - x^2)\,dx$$

$$= \frac{1}{2}\,\frac{x^3}{3}\Big|_0^1 + \frac{1}{6}\left(4\frac{x^2}{2} - \frac{x^3}{3}\right)\Big|_1^4$$

$$= \frac{1}{6} + \frac{1}{6}\left(32 - \frac{64}{3} - 2 + \frac{1}{3}\right)$$

$$= \frac{5}{3}$$

연습 13. Problems 6.1

Problems

확률 변수가 양수라는 것과
주어진 기댓값 정보 이외에
확률 분포에 대한 정보가 없음.

Inequalities[6]

6.1 The magnitude of the voltage V across a component in an electronic circuit has a mean value of 0.45 volts. Given only this information, find a bound on the probability that $V \geq 1.35$.

--> 마르코프 부등식을 통한 상한값 계산?

$$\Pr[V \geq 1.35] \leq \frac{0.45}{1.35} = \frac{1}{3}$$

연습 14. Problems 4.43

4.43 The probability density function of a Laplace random variable is given by

$$f_X(x) = \frac{\alpha}{2} e^{-\alpha|x-\mu|}, \quad -\infty < x < \infty.$$

(a) Find the mean m_X, second moment, and variance σ_X^2.

(b) Determine the following probabilities for $\alpha = 4$ and $\mu = 1$.

(i) $\Pr[|X - m_X| > 2\sigma_X]$

(ii) $\Pr[|X - m_X| < \sigma_X]$

연습 14. Problems 4.43

(a) Find the mean m_X, second moment, and variance σ_X^2.

$$f_X(x) = \frac{\alpha}{2} e^{-\alpha|x-\mu|}, \qquad -\infty < x < \infty$$

$$= \begin{cases} \frac{\alpha}{2} e^{\alpha(x-\mu)} & -\infty < x < \mu \\ \frac{\alpha}{2} e^{-\alpha(x-\mu)} & \mu \le x < \infty \end{cases}$$

$$m_X = \frac{\alpha}{2} \int_{-\infty}^{\mu} x\, e^{\alpha(x-\mu)} dx + \frac{\alpha}{2} \int_{\mu}^{\infty} x\, e^{-\alpha(x-\mu)} dx$$

$$= \frac{1}{2}\left[\mu - \frac{1}{\alpha}\right] + \frac{1}{2}\left[\mu + \frac{1}{\alpha}\right] = \mu$$

연습 14. Problems 4.43

(a) Find the mean m_X, second moment, and variance σ_X^2.

$$E[X^2] = \frac{\alpha}{2}\int_{-\infty}^{\mu} x^2 e^{\alpha(x-\mu)}\,dx + \frac{\alpha}{2}\int_{\mu}^{\infty} x^2 e^{-\alpha(x-\mu)}\,dx$$

$$= \frac{1}{2}\left[\mu^2 - \frac{2\mu}{\alpha} + \frac{2}{\alpha^2}\right] + \frac{1}{2}\left[\mu^2 + \frac{2\mu}{\alpha} + \frac{2}{\alpha^2}\right] = \mu^2 + \frac{2}{\alpha^2}$$

$$\sigma_X^2 = E[X^2] - \left(E[X]\right)^2$$

$$= \frac{2}{\alpha^2} + \mu^2 - \mu^2 = \frac{2}{\alpha^2}$$

연습 14. Problems 4.43

(b) Determine the following probabilities for $\alpha = 4$ and $\mu = 1$.

(i) $\Pr[|X - m_X| > 2\sigma_X]$

$$= \frac{\alpha}{2}\int_{-\infty}^{m_X-2\sigma_X} e^{\alpha(x-\mu)}\,dx + \frac{\alpha}{2}\int_{m_X+2\sigma_X}^{\infty} e^{-\alpha(x-\mu)}\,dx$$

$$= \frac{\alpha}{2}\cdot \left.\frac{e^{\alpha(x-\mu)}}{\alpha}\right|_{-\infty}^{m_X-2\sigma_X} + \frac{\alpha}{2}\cdot \left.\frac{e^{-\alpha(x-\mu)}}{-\alpha}\right|_{m_X+2\sigma_X}^{\infty}$$

$$= \frac{1}{2}\left[e^{\alpha(m_X-2\sigma_X-\mu)} + e^{-\alpha(m_X+2\sigma_X-\mu)}\right]$$

$$= e^{-2\alpha\sigma_X} = e^{-2\sqrt{2}} = 0.05911$$

연습 14. Problems 4.43

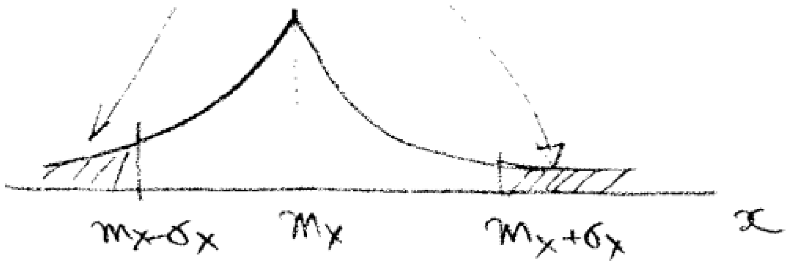
(ii) $\Pr[|X - m_X| < \sigma_X]$

$$\Pr[|X - m_X| < \sigma_X] = \Pr[-\sigma_X < X - m_X < \sigma_X]$$

$$= \Pr[X < m_X + \sigma_X] - \Pr[X < m_X - \sigma_X]$$

$$= 1 - \Pr[X > m_X + \sigma_X] - \Pr[X < m_X - \sigma_X]$$

$$= 1 - \frac{\alpha}{2}\int_{m_X+\sigma_X}^{\infty} e^{-\alpha(x-\mu)}\,dx - \frac{\alpha}{2}\int_{-\infty}^{m_X-\sigma_X} e^{\alpha(x-\mu)}\,dx$$



$$= 1 - \frac{\alpha}{2}\left.\frac{e^{-\alpha(x-\mu)}}{-\alpha}\right|_{m_X+\sigma_X}^{\infty} - \frac{\alpha}{2}\left.\frac{e^{\alpha(x-\mu)}}{\alpha}\right|_{-\infty}^{m_X-\sigma_X}$$

$$= 1 - e^{-\sqrt{2}}$$

$$= 0.7569$$

연습 15. Problems 4.49

4.49 The moment generating function of a continuous random variable X is given to be

$$M_X(s) = \frac{3}{1-2s}.$$

Find the second moment.

$$E[X^2] = \left.\frac{d^2 M_X(s)}{ds^2}\right|_{s=0}$$

$$\frac{d^2}{ds^2}\left[\frac{3}{1-2s}\right] = \frac{d}{ds}\left[\frac{6}{(1-2s)^2}\right]$$

$$= \frac{24}{(1-2s)^3}$$

$$E[X^2] = \left.\frac{24}{(1-2s)^3}\right|_{s=0}$$

$$= 24$$

실습 3. Python 프로그래밍

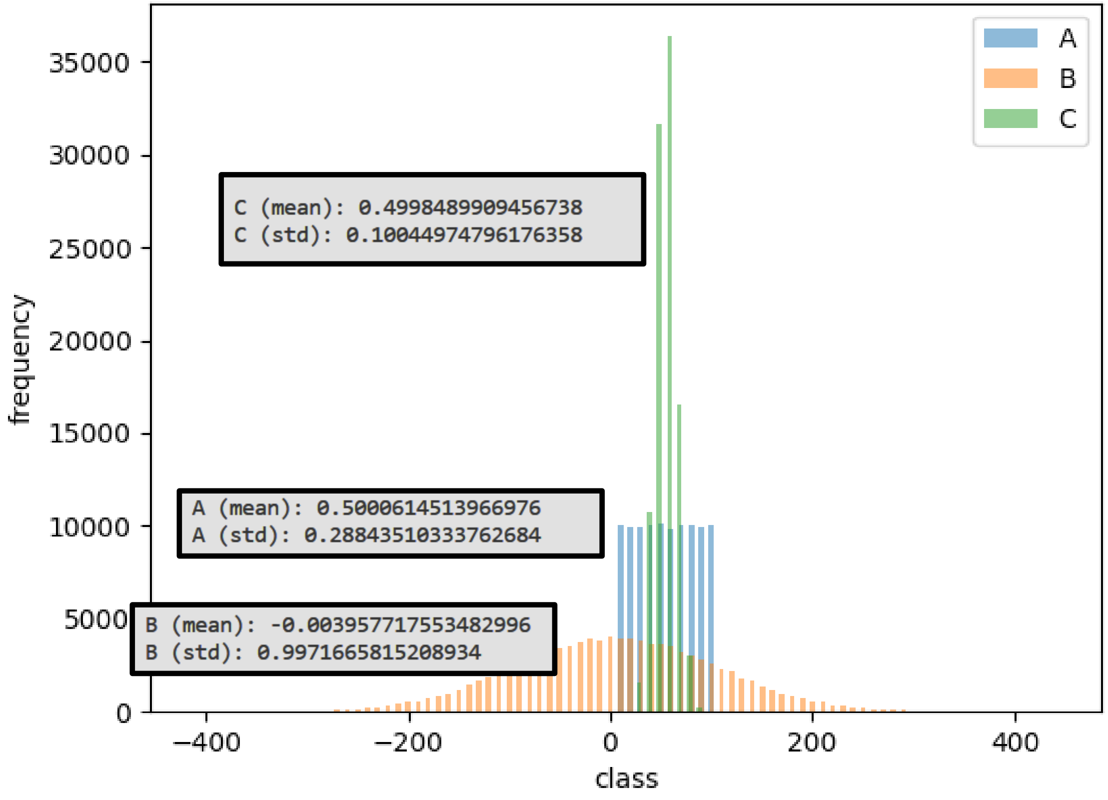
- **NumPy를 활용한 히스토그램 생성 과제**
  - np.random.rand, np.random.randn 함수를 이용한 과목별 성적 분포 히스토그램 작성
    - 과목 A, B, C의 점수 값을 100,000개 이상 무작위로 발생
    - 10점 간격의 구간별로 도수 분포를 파악하여, 막대 그래프로 표현 → 캡쳐
    - 과목 A, B, C의 평균(mean)과 표준편차(standard deviation) 확인 → 캡쳐

```python
sampleA, sampleB, sampleC = A*100, B*100, C*100
classA = np.arange(min(sampleA),max(sampleA)+1,10)
classB = np.arange(min(sampleB),max(sampleB)+1,10)
classC = np.arange(min(sampleC),max(sampleC)+1,10)

counts1, bins1 = np.histogram(sampleA, classA)
counts2, bins2 = np.histogram(sampleB, classB)
counts3, bins3 = np.histogram(sampleC, classC)

plt.bar(bins1[1:], counts1, alpha=0.5, width=5, label='A')
plt.bar(bins2[1:], counts2, alpha=0.5, width=5, label='B')
plt.bar(bins3[1:], counts3, alpha=0.5, width=5, label='C')

plt.legend()
plt.title("Ex.#2-1")
plt.xlabel("class")
plt.ylabel("frequency")
plt.savefig("histogramABC.png")
plt.show()
```



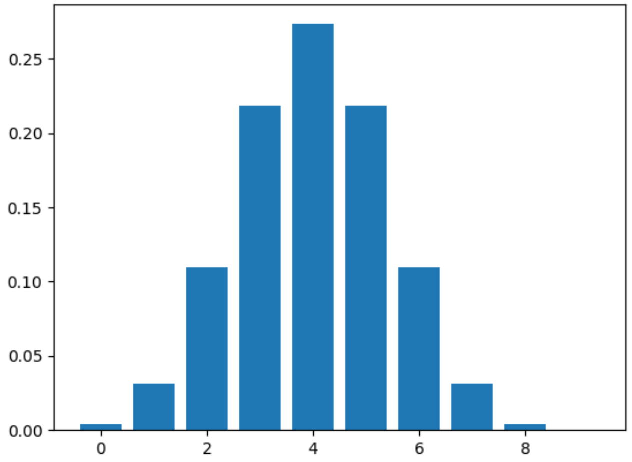
실습 4. Poisson 확률분포 실습

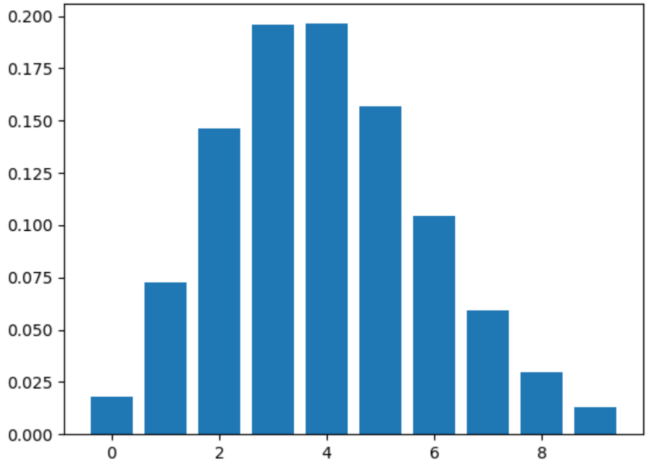
- **Poisson distribution 실습**
  - 낮은 성공률 p의 베르누이 실험을 매우 많은 N번 시행할 때 성공(1) 발생 횟수에 대한 확률변수 X
    - 성공 확률 0.5인 실험을 8번 수행할 때 성공 횟수에 대한 확률값 분포 (이항분포)?
    - 성공 확률 0.01인 실험을 400번 수행할 때 성공 횟수에 대한 확률값 분포 (이항분포)?
    - p x N = 4 비율이 일정하게 유지될 때, 성공 횟수에 대한 확률값 분포 (푸아송 분포)

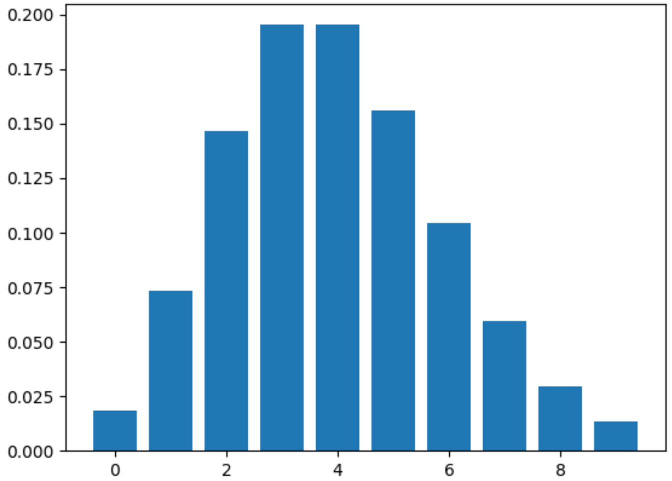
```
p = 1/2
x = np.arange(0,10)
y = st.binom.pmf(x, 8, p)
plt.bar(x, y)
plt.show()
```

```
p = 1/100
x = np.arange(0,10)
y = st.binom.pmf(x, 400, p)
plt.bar(x, y)
plt.show()
```

```
pn = 4
x = np.arange(0,10)
y = st.poisson.pmf(x, pn)
plt.bar(x, y)
plt.show()
```





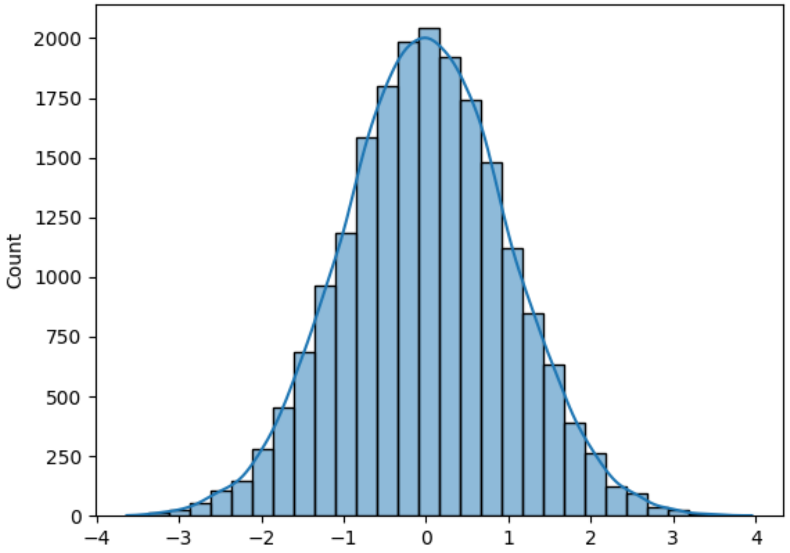


실습 5. Binomial 분포에 대한 중심극한정리 실습

- **중심 극한 정리 실습**
  - 서로 독립적이며 동일한(I.I.D) 이항 확률변수 $X_1, X_2, X_3, \ldots$ 의 합을 새로운 확률변수 Y로 가정
  - Y의 평균과 표준편차로 표준화 한 새로운 확률변수 Z
    - 동전 10개 던지기 게임의 이항분포들 N개의 합 $X_1 + X_2 + X_3, \ldots + X_N$ 에 대한 확률변수 Z
    - N이 커짐에 따라 확률변수 Z은 표준 정규분포에 근사됨을 확인

```
N = 20000 # from 30 to 20000
x_sums = np.zeros(20000)
for k in range(20000):
    x = st.binom.rvs(10,0.5,size=N)
    x_sums[k] = np.sum(x)
x_sum_mean = np.mean(x_sums)
x_sum_std = np.std(x_sums)
print(x_sum_mean, x_sum_std**2)
z = (x_sums-x_sum_mean)/x_sum_std
sns.histplot(z, bins=30, kde=True)
plt.show()
```

이항분포 합의 Z 분포




```
99998.84995 50008.55703499751
```

Y의 평균과 분산

표준 정규분포