

Solutions Tutorial 1

Events, Intersections, and Unions

A standard 52-card deck has 4 suits (Hearts, Diamonds, Clubs, Spades) and 13 ranks (2-10, J, Q, K, A).

1. **P(A): Probability of drawing a King** There are 4 Kings in the deck.

$$P(A) = \frac{\text{Number of Kings}}{\text{Total Cards}} = \frac{4}{52} = \frac{1}{13} \approx 0.0769$$

2. **P(B): Probability of drawing a Heart** There are 13 Hearts in the deck.

$$P(B) = \frac{\text{Number of Hearts}}{\text{Total Cards}} = \frac{13}{52} = \frac{1}{4} = 0.25$$

3. **P(A ∩ B): Probability of drawing a King AND a Heart** There is only one card that is both a King and a Heart (the King of Hearts).

$$P(A \cap B) = \frac{1}{52} \approx 0.0192$$

4. **P(A ∪ B): Probability of drawing a King OR a Heart** Using the formula for the union of events:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13} \approx 0.3077$$

Solution: Conditional Probability

1. $P(A|B)$: Probability of a King, given it's a Heart

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/52}{13/52} = \frac{1}{13}$$

Intuition: If we know the card is a Heart, our sample space is reduced to the 13 Hearts. Of those, only one is a King.

2. $P(B|A)$: Probability of a Heart, given it's a King

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{1/52}{4/52} = \frac{1}{4}$$

Intuition: If we know the card is a King, our sample space is reduced to the 4 Kings. Of those, only one is a Heart.

3. Independence Check Two events are independent if $P(A \cap B) = P(A) \times P(B)$.

- $P(A \cap B) = \frac{1}{52}$
- $P(A) \times P(B) = \frac{1}{13} \times \frac{1}{4} = \frac{1}{52}$ Since $P(A \cap B) = P(A) \times P(B)$, the events **are independent**.

Solution: Simulating an Experiment

```
import numpy as np

# 1. Create the deck
ranks = ['2', '3', '4', '5', '6', '7', '8', '9', '10', 'J', 'Q', 'K', 'A']
suits = ['H', 'D', 'C', 'S'] # Hearts, Diamonds, Clubs, Spades
deck = [rank + suit for suit in suits for rank in ranks]

# 2. Simulate 100,000 draws
n_simulations = 100_000
draws = np.random.choice(deck, size=n_simulations, replace=True)

# 3. Count events
count_A = np.sum(['K' in card for card in draws])
count_B = np.sum(['H' in card for card in draws])
count_A_intersect_B = np.sum(draws == 'KH')
count_A_union_B = np.sum(['K' in card or 'H' in card for card in draws])

# 4. Calculate and compare frequencies
print("Comparing Theoretical Probabilities with Empirical Frequencies:\n")

# Event A: King
prob_A_emp = count_A / n_simulations
print(f"P(King):          Theoretical={4/52:.4f}, Empirical={prob_A_emp:.4f}")

# Event B: Heart
prob_B_emp = count_B / n_simulations
print(f"P(Heart):          Theoretical={13/52:.4f}, Empirical={prob_B_emp:.4f}")

# Event A intersect B: King of Hearts
prob_A_intersect_B_emp = count_A_intersect_B / n_simulations
print(f"P(King   Heart): Theoretical={1/52:.4f}, Empirical={prob_A_intersect_B_emp:.4f}")

# Event A union B: King or Heart
prob_A_union_B_emp = count_A_union_B / n_simulations
print(f"P(King U Heart):  Theoretical={16/52:.4f}, Empirical={prob_A_union_B_emp:.4f}")
```

Solution: PMF, Expected Value, and Variance

1. **Verify PMF:** The function is a valid PMF because all probabilities are non-negative and they sum to 1:

$$0.1 + 0.5 + 0.3 + 0.1 = 1.0$$

2. **Expected Value $E[X]$:**

$$E[X] = \sum x \cdot P(X = x)$$

$$E[X] = (0 \times 0.1) + (1 \times 0.5) + (2 \times 0.3) + (3 \times 0.1)$$

$$E[X] = 0 + 0.5 + 0.6 + 0.3 = 1.4$$

The expected number of items per customer is 1.4.

3. **Variance $\text{Var}(X)$:**

$$\text{Var}(X) = \sum (x - \mu)^2 \cdot P(X = x)$$

$$\text{Var}(X) = (0 - 1.4)^2(0.1) + (1 - 1.4)^2(0.5) + (2 - 1.4)^2(0.3) + (3 - 1.4)^2(0.1)$$

$$\text{Var}(X) = (1.96)(0.1) + (0.16)(0.5) + (0.36)(0.3) + (2.56)(0.1)$$

$$\text{Var}(X) = 0.196 + 0.08 + 0.108 + 0.256 = 0.64$$

4. **Probability of buying more than one item:**

$$P(X > 1) = P(X = 2) + P(X = 3) = 0.3 + 0.1 = 0.4$$

Solution: The Bernoulli Distribution in Theory and Programming

(a) Theoretical Calculation

Given the probability of success (default) $p = 0.05$.

- **Expected Value:**

$$E[X] = p = 0.05$$

- **Variance:**

$$\text{Var}(X) = p(1 - p) = 0.05 \times (1 - 0.05) = 0.05 \times 0.95 = 0.0475$$

(b) Simulation and Comparison

```
import numpy as np

p_default = 0.05
n_simulations = 1_000_000

# Simulate 1,000,000 Bernoulli trials
simulations = np.random.choice([0, 1], size=n_simulations, p=[1-p_default, p_default])

# Calculate sample mean and variance
sample_mean = np.mean(simulations)
sample_var = np.var(simulations)

print("Bernoulli Distribution Comparison:\n")
print(f"Theoretical E[X]    = {0.05:.6f}")
print(f"Simulated mean       = {sample_mean:.6f}\n")
print(f"Theoretical Var(X)    = {0.0475:.6f}")
print(f"Simulated variance    = {sample_var:.6f}")
```

Solution: Standardization and Z-scores

Given $R \sim N(\mu = 0.12, \sigma = 0.20)$.

1. **Z-score for a return of 32% (0.32):**

$$Z = \frac{X - \mu}{\sigma} = \frac{0.32 - 0.12}{0.20} = \frac{0.20}{0.20} = 1.0$$

This Z-score signifies that a return of 32% is **exactly 1 standard deviation above the mean return**.

2. **Z-score for a return of -8% (-0.08):**

$$Z = \frac{-0.08 - 0.12}{0.20} = \frac{-0.20}{0.20} = -1.0$$

This means a return of -8% is **exactly 1 standard deviation below the mean return**.

3. **Portfolio return for a Z-score of 1.5:** Rearrange the formula: $X = \mu + Z\sigma$.

$$X = 0.12 + (1.5 \times 0.20) = 0.12 + 0.30 = 0.42$$

A Z-score of 1.5 corresponds to a portfolio return of **42%**.

Solution: Calculating Probabilities and Quantiles

```
from scipy.stats import norm

# Define the parameters
mu = 0.12
sigma = 0.20

# 1. Probability of a negative return ( $P(R < 0)$ )
prob_neg = norm.cdf(0, loc=mu, scale=sigma)
print(f"1. Probability of a negative return: {prob_neg:.4f} (or {prob_neg:.2%})")

# 2. Probability of return > 25% ( $P(R > 0.25)$ )
#  $P(R > 0.25) = 1 - P(R \leq 0.25)$ 
prob_gt_25 = 1 - norm.cdf(0.25, loc=mu, scale=sigma)
print(f"2. Probability of return > 25%: {prob_gt_25:.4f} (or {prob_gt_25:.2%})")

# 3. Probability of return between 0% and 15% ( $P(0 < R < 0.15)$ )
#  $P(0 < R < 0.15) = P(R < 0.15) - P(R < 0)$ 
prob_between = norm.cdf(0.15, loc=mu, scale=sigma) - norm.cdf(0, loc=mu, scale=sigma)
print(f"3. Probability of return between 0% and 15%: {prob_between:.4f} (or {prob_between:.2%})")

# 4. 5th percentile of the return distribution
# This is the value 'x' such that  $P(R < x) = 0.05$ 
p5 = norm.ppf(0.05, loc=mu, scale=sigma)
print(f"4. The 5th percentile return is: {p5:.4f} (or {p5:.2%})")
```

Solution: Linear Combinations of Normal Variables

(a) Theoretical Calculation

Given $R_A \sim N(0.10, 0.15^2)$ and $R_B \sim N(0.06, 0.10^2)$. The portfolio return is $R_P = 0.6R_A + 0.4R_B$.

- **Expected Value** $E[R_P]$:

$$E[R_P] = 0.6E[R_A] + 0.4E[R_B] = 0.6(0.10) + 0.4(0.06) = 0.06 + 0.024 = 0.084$$

- **Variance** $Var(R_P)$: Since the assets are independent, their covariance is 0.

$$Var(R_P) = 0.6^2Var(R_A) + 0.4^2Var(R_B) = 0.36(0.15^2) + 0.16(0.10^2)$$

$$Var(R_P) = 0.36(0.0225) + 0.16(0.01) = 0.0081 + 0.0016 = 0.0097$$

- **Full Distribution:** The portfolio return is also normally distributed:

$$R_P \sim N(0.084, 0.0097)$$

(b) Simulation

```
import numpy as np
import matplotlib.pyplot as plt

n_sim = 100_000
# Simulate returns for each asset
returns_A = np.random.normal(loc=0.10, scale=0.15, size=n_sim)
returns_B = np.random.normal(loc=0.06, scale=0.10, size=n_sim)

# Calculate the portfolio return for each simulation
returns_P = 0.6 * returns_A + 0.4 * returns_B

# Calculate sample mean and variance
mean_P_sim = np.mean(returns_P)
var_P_sim = np.var(returns_P)

print("Portfolio Simulation Comparison:\n")
print(f"Theoretical E[Rp] = {0.084:.4f}, Simulated mean = {mean_P_sim:.4f}")
print(f"Theoretical Var(Rp) = {0.0097:.4f}, Simulated variance = {var_P_sim:.4f}\n")

# Bonus: Plot histogram
plt.hist(returns_P, bins=50, density=True, alpha=0.7, label='Simulated Portfolio Returns')
plt.title('Histogram of Simulated Portfolio Returns')
```



```
plt.xlabel('Return')  
plt.ylabel('Density')  
plt.grid(alpha=0.3)  
plt.show()
```

Solution: Rules of Expectation and Variance

Given: $E[X] = 10, Var(X) = 4, E[Y] = 5, Var(Y) = 9, Cov(X, Y) = -2$.

1. $E[X + Y]$

$$E[X + Y] = E[X] + E[Y] = 10 + 5 = 15$$

2. $Var(X + Y)$

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y) = 4 + 9 + 2(-2) = 13 - 4 = 9$$

3. $E[3X - 2Y]$

$$E[3X - 2Y] = 3E[X] - 2E[Y] = 3(10) - 2(5) = 30 - 10 = 20$$

4. $Var(3X - 2Y)$

$$\begin{aligned} Var(3X - 2Y) &= 3^2 Var(X) + (-2)^2 Var(Y) + 2(3)(-2)Cov(X, Y) \\ &= 9(4) + 4(9) - 12(-2) = 36 + 36 + 24 = 96 \end{aligned}$$

5. $Var(X + Y)$ assuming independence If X and Y are independent, $Cov(X, Y) = 0$.

$$Var(X + Y) = Var(X) + Var(Y) = 4 + 9 = 13$$

This variance (13) is higher than the variance when the covariance was -2 (which was 9). The negative covariance implies the variables tend to move in opposite directions, which has a stabilizing (variance-reducing) effect on their sum.

Solution: Interpreting Conditional Expectation

(a) **Expected income for 12 years of education** We substitute $e = 12$ into the formula:

$$E[I|E = 12] = 15000 + 4000(12) = 15000 + 48000 = \$63,000$$

(b) **Expected income for 16 years of education** We substitute $e = 16$ into the formula:

$$E[I|E = 16] = 15000 + 4000(16) = 15000 + 64000 = \$79,000$$

(c) **Is $E[I|E]$ a number or a random variable?** $E[I|E]$ is a **random variable**.

Explanation:

- $E[I|E = e]$ (like in parts a and b) is a **single number**. It is the expected income for a *specific, fixed* level of education.
- $E[I|E]$ (without specifying the value of E) is a function of the random variable E . Since the value of E (the years of education for a randomly selected person) will vary, the value of $E[I|E]$ will also vary. Therefore, it is itself a random variable whose outcome depends on the outcome of E .