

Session 8: Problem Solving with Probability (with Solutions)

Four steps of math problem solving:

1. **Describe** what is desired and what is given in succinct and precise language.
2. **Identify** all relevant concepts and formula that you know about. (This is a brainstorming exercise so you should try to draw as many connections as you can.)
3. **Plan** a pathway from what is given to what is desired. (You can either start backward from what is desired or go forward from what is given.)
4. **Execute** the plan above to solve the problem and compute the final answer.

Q1 (Weather Prediction)

In Oblako County, any day can be either sunny or cloudy. If a day is sunny, the following day will be sunny with probability 0.6. If a day is cloudy, the following day will be cloudy with probability 0.7. Suppose it is cloudy on Monday, what is the probability that it will be sunny on Wednesday?

Describe what is desired and what is given:

Desired: Probability of sunny on wednesday

Given: Cloudy on Monday, conditional probability of tomorrow's weather given today's weather.

Identify all relevant concepts and formula:

Conditional probability.

$$P(A|B) = P(A \text{ and } B) / P(B)$$

If A and B are mutually exclusive, $P(A \text{ or } B) = P(A) + P(B)$.

Plan a pathway to solving the problem: Use the given conditional probabilities to compute probability that Tuesday's weather is sunny. Then do the same thing to compute probability that Wednesday's weather is sunny. The key is

$$P(\text{Wed. Sunny}) = P(\text{Wed. Sunny and Tue. Sunny}) + P(\text{Wed. Sunny and Tue. Cloudy}),$$

and

$$P(\text{Wed. Sunny and Tue. Sunny}) = 0.6P(\text{Tue. Sunny}),$$

$$P(\text{Wed. Sunny and Tue. Cloudy}) = (1 - 0.7)P(\text{Tue. Cloudy}),$$

Execute the plan to solve the problem:

Let $T_{\text{sunny}}, T_{\text{cloudy}}$ be the event that Tuesday is sunny and cloudy respectively. Similarly define W_{sunny} for Wednesday. We have

$$P(T_{\text{sunny}}) = 1 - 0.7 = 0.3,$$

$$P(T_{\text{cloudy}}) = 0.7,$$

$$P(T_{\text{sunny}} \text{ and } W_{\text{sunny}}) = P(W_{\text{sunny}}|T_{\text{sunny}})P(T_{\text{sunny}}) = (0.6)(0.3) = 0.18,$$

$$P(T_{\text{cloudy}} \text{ and } W_{\text{sunny}}) = P(W_{\text{sunny}}|T_{\text{cloudy}})P(T_{\text{cloudy}}) = (0.3)(0.7) = 0.21,$$

$$P(W_{\text{sunny}}) = 0.18 + 0.21 = 0.39.$$

Q2 (Television Marketing)

An athletic footwear company is attempting to estimate the sales that will result from a television advertisement campaign of its new athletic shoe. The contribution to earnings from each pair of shoes sold is 40 dollars. For simplicity, assume that the exact # of viewers are as given below. Suppose that the probability that a television viewer will watch the advertisement (as opposed to turn his/her attention elsewhere) is 0.40, independent of the others. Furthermore, suppose that 1% of viewers who watch the advertisement on a local television channel will buy a pair of shoes, independent of the others. The advertisement will take one minute of airtime, and the company can buy advertising time in one of the slots as shown below.

Time Slot	Cost of Advertisement (per minute)	# of Viewers
Morning	120,000	1,000,000
Afternoon	200,000	1,300,000
Prime Time	400,000	3,200,000
Late Evening	150,000	800,000

(a) Which time slot would maximize the company's expected profit? (Profit is total earnings from advertisement minus cost of advertisement.)

Describe:

Desired: the slot that maximizes expected profit.

Given: the cost of each slot, the # of viewers of each slot, the probability that each viewer watches the ad, the probability that a viewer who watches the ad purchases, the earning from selling each pair of shoes.

Identify:

Profit of slot = expected earnings - cost of slot

Expected earnings is equal to price multiplied by expected number of sales.

The probability each viewer purchases = probability of watching the ad \times probability of one who sees the ad purchases = $(0.40)(0.01) = 0.004$.

The total number of people who purchases is related to the binomial distribution, n is the # of viewers and p is the probability of purchasing above. The expected value of this is np .

Plan:

First compute the expected number of people who purchase based on the binomial distribution formula for expected values, then compute the profit by multiplying the price and subtracting the cost.

Execute:

Since each viewer purchases with probability $0.4 \times 0.01 = 0.004$, the expected # of purchases is equal to this times the # of viewers.

We complete the above table below:

Time Slot	Cost	# of Viewers	Expected # of purchases	Expected Profit
Morning	120,000	1,000,000	4,000	$(4000)(40) - 120000 = 40000$
Afternoon	200,000	1,300,000	5,200	$(5200)(40) - 200000 = 8000$
Prime Time	400,000	3,200,000	12,800	$(12800)(40) - 400000 = 112000$
Late Evening	150,000	800,000	3,200	$(3200)(40) - 150000 = -22000$

Hence, the best slot is the prime time, with expected profit of 112,000.

(b) For the best slot, what is the estimated expected value and standard deviation in profit, assuming that the # of viewers for each time slot is deterministically known as in the table?

Describe:

Desired: standard deviation of profit (expected value is already obtained above)

Given: a binomial distribution for the # of purchases ($n = 3200000$, $p = 0.004$), price of each product, cost.

Identify:

The standard deviation of a random variable X satisfies the formula $SD(aX) = aSD(X)$, and $SD(X - a) = SD(X)$ for any constant a .

The standard deviation of a binomial random variable with parameters n and p is $\sqrt{np(1-p)}$.

Plan:

First compute the standard deviation of the # of purchases using the above formula, then multiply the result by 40 to obtain the standard deviation in the earnings. The standard deviation in profit is the same because cost is constant.

Execute:

The standard deviation in purchases is

$$40\sqrt{(3200000)(0.004)(0.996)} = 4516.42$$

```
[2]: import math
     40*math.sqrt(3.2*1e6*0.004*0.996)
```

```
4516.423363680602
```

Q3 (Pricing with Market Segmentation)

Blaise owns a store selling a certain product. His market research team categorizes potential customers into two segments, A and B. They estimate that on average, 30% of customers are of segment A, and 70% customers are of segment B. (However, the actual proportion of your customers each day who are of segment A may vary from day to day, as there are random fluctuations.) They further estimate that the maximum willingness to pay of a segment A customer is normally distributed with mean 150 and standard deviation 30, while the maximum willingness to pay of a segment B customer is normally distributed with mean 120 and standard deviation 40. Suppose Blaise prices his product at 160 dollars and that he has more than enough inventory.

(a) Calculate the probability that a customer from each of the two segment purchases the product.

Describe:

Desired: 1) probability a segment A customer purchases; 2) probability a segment B customer purchases.

Given: the relative proportion of segment A and segment B customers, the price, the distribution of willingness to pay for each segment.

Identify:

Purchase probability is the probability that the willingness to pay is at least equal to the price. This is related to the CDF of the willingness to pay. For example, if X is the amount a segment A customer is willing to pay, then the purchase probability is $P(X \geq 160) = 1 - P(X \leq 160) = 1 - F(160)$, where F is the Normal CDF with $\mu = 150$ and $\sigma = 30$.

Plan: Use Python to obtain the CDF F of valuation for each segment, and probability of purchasing is $1 - F(160)$.

Execute: Let X and Y be random variables representing the willingness to pay of a segment A and segment B customer respectively. $X \sim \text{Normal}(150, 30)$ and $Y \sim \text{Normal}(120, 40)$. Let the

CDFs be F_A and F_B respectively. Let b_A and b_B be the purchasing probability of each segment. Then

$$b_A = 1 - F_A(160) = .369,$$

$$b_B = 1 - F_B(160) = .159.$$

```
[3]: from scipy.stats import norm
      b_A=1-norm(150,30).cdf(160)
      b_B=1-norm(120,40).cdf(160)
      b_A,b_B
```

```
(0.36944134018176367, 0.15865525393145707)
```

(b) Calculate the probability that a randomly chosen customer who purchases the product is from segment A.

Describe:

Desired: probability that a customer is from segment A, conditional on having purchased.

Given: the overall proportion of segment A customers, and the probability of purchasing conditional on the segment.

Identify:

This is a conditional probability question which one can solve using the joint probability table or Bayes' rule (see DMD readings).

$$P(A) = 0.3$$

$$P(B) = 0.7$$

$$P(buy|A) = .369$$

$$P(buy|B) = .159$$

$$P(A|buy) = ?$$

Plan:

Use Bayes' rule to solve.

Execute: Let A and B denote the event that a chosen customer is from segment A and segment B respectively. Let buy denote the event of purchasing. By Bayes' rule

$$P(A|buy) = \frac{P(buy|A)P(A)}{P(buy|A)P(A) + P(buy|B)P(B)} = 0.499.$$

```
[4]: p_a_buy=(b_A*0.3)/(b_A*0.3+b_B*0.7)
      p_a_buy
```

```
0.49949011988745085
```

(c) The market research team estimates that each segment A customer who purchase the product would later return it with 20% probability, and each segment B customer would return it with 1% probability, independent of the others. Calculate the probability that out of 1000 customers who purchased the product, at least 100 would later return it.

Describe:

Desired: $P(X \geq 100)$ where X is the number of people who return the product out of 1000 customers.

Given: $P(A|buy)$, $P(B|buy)$, $P(Return|A \text{ and } buy)$, and $P(Return|B \text{ and } buy)$.

Identify:

X is binomial distributed with $n = 1000$ and $p = P(\text{return}|\text{buy})$.

$$P(\text{return}|\text{buy}) = P(\text{return and A}|\text{buy}) + P(\text{return and B}|\text{buy}).$$

$$P(\text{return and A}|\text{buy}) = P(\text{return}|A \text{ and buy})P(A|\text{buy}).$$

Plan:

Use the above formula to compute $p = P(\text{return}|\text{buy})$, then use the CDF of the binomial distribution. Noting that if F is the CDF of X ,

$$P(X \geq 100) = 1 - P(X \leq 99) = 1 - F(99).$$

Execute:

Let p be the probability that a customer who purchases the product returns it, we have

$$p = 0.2P(A|\text{buy}) + 0.01P(B|\text{buy}) = (0.2)(0.499) + (0.01)(0.501) \approx 0.105.$$

Let X be the number of customers who return the product out of the 1000 given customers. $X \sim \text{Binomial}(n = 1000, p)$. Let F be its CDF. The desired probability is

$$P(X \geq 100) = 1 - F(99) \approx 0.708.$$

```
[5]: from scipy.stats import binom
      return_prob=p_a_buy*0.2+(1-p_a_buy)*0.01
      dist=binom(1000,return_prob)
      1-dist.cdf(99)
```

0.7082061756245909

(d) Suppose that there is 100,000 customers in total. Moreover, a returned product yields zero revenue. What is the expected value and standard deviation of revenue from each segment, after accounting for returns?

Describe:

Desired: expected value and standard deviation of revenue from each segment.

Given: each (unreturned) purchase yields revenue of 160, the size of each segment, the purchase probability of each segment, and the return probability of each segment.

Identify:

Let revenue_A and purchases_A be the revenue from segment A and the number of (unreturned) purchases from this segment respectively. We have

$$E[\text{revenue}_A] = 160E[\text{purchases}_A].$$

$$SD(\text{revenue}_A) = 160SD(\text{purchases}_A).$$

The random variable purchases_A is binomial distributed with $n = 30000$ and p being the probability of buying and not returning. The probability of buying and not returning is equal to the probability of buying (0.369) multiplied by the probability of not returning (0.8).

Plan:

Calculate the probability of purchasing and not returning for each segment. Then use the binomial distribution formula to compute the mean and standard deviation of the total number of (unreturned) purchases from each segment. Finally, multiplying each by the price to find the mean and standard deviation of profit.

Execution: Let p_A and p_B be the probability of buying and not returning from a randomly chosen individual of segment A and B respectively, we have

$$p_A = (1 - 0.2)P(\text{buy}|A) = (1 - 0.2)(0.369) \approx 0.296,$$

$$p_B = (1 - 0.01)P(\text{buy}|B) = (1 - 0.01)(0.159) \approx 0.157.$$

Let X_A and X_B be the number of unreturned purchases from each segment. These are distributed $X_A \sim \text{Binomial}(n = 30000, p_A)$ and $X_B \sim \text{Binomial}(n = 70000, p_B)$.

Let R_A and R_B be the revenue from each segment, we have $R_A = 160X_A$ and $R_B = 160X_B$. Therefore,

$$E[R_A] = 160E[X_A] = 160(30000)p_A \approx 1.42 \times 10^6,$$

$$SD[R_A] = 160SD[X_A] = 160\sqrt{(30000)p_A(1 - p_A)} \approx 1.26 \times 10^4.$$

Similarly, $E[R_B] \approx 1.76 \times 10^6$ and $SD(R_B) \approx 1.54 \times 10^4$.

```
[6]: import numpy as np
    p_A=b_A*(1-0.2)
    p_B=b_B*(1-0.01)
    n_A=30000
    n_B=70000
    mu_A=n_A*p_A*160
    sigma_A=np.sqrt(n_A*p_A*(1-p_A))*160
    mu_B=n_B*p_B*160
    sigma_B=np.sqrt(n_B*p_B*(1-p_B))*160
    print('Segment A: expected revenue,',mu_A,'standard deviation',sigma_A)
    print('Segment B: expected revenue,',mu_B,'standard deviation',sigma_B)
```

Segment A: expected revenue, 1418654.7462979725 standard deviation 12645.1064224288

Segment B: expected revenue, 1759169.4555919957 standard deviation 15403.163278617487