# Pa.02 – Lab Assignment


## Nobel Gabriel (nobelgab), von Wartburg Rebekka (vonwareb)


## 09.03.2023

### 1. Decision problem for a bank

Answer a)

general

| | $Y = 0$ | $Y = 1$ |
|---|---|---|
| $D = 0$ | $0$ | $0$ |
| $D = 1$ | $-H$ | $z \cdot H$ |

concret

| | $Y = 0$ | $Y = 1$ |
|---|---|---|
| $D = 0$ | $0$ | $0$ |
| $D = 1$ | $-H$ | $0.1 \cdot H$ |

- What's the profit of the bank if an applicant is accepted and repays the loan?
  - Answer: 0.1 * H
- What's the profit of the bank if an applicant is accepted but defaults on the loan?
  - Answer: - H
- What's the profit of the bank if an applicant is NOT accepted but would have repaid the loan?
  - Answer: 0
- What's the profit of the bank if an applicant is NOT accepted and would have NOT repaid the loan?
  - Answer: 0

Answer b)

$$
\begin{aligned}
E[u] &= (1-p)\cdot u_{10} + p \cdot u_{11} \\
&\phantom{=}\ (1-p)\cdot(-H) + p \cdot z \cdot H = \\
&\phantom{=}\ -H + Hp + Hpz = 0 \\
&\phantom{=}\ Hp + Hpz = H \\
&\phantom{=}\ p(H + Hz) = H \\
p &= \frac{H}{H + Hz} \\
p &= \frac{H}{H(1+z)} \\
p &= \underline{\underline{\frac{1}{1+z}}}
\end{aligned}
$$

For a repayment probability p of 0.9090 D = 1 would be the better decision

The rule above doesn't depend on H

Answer c)

$$p = \frac{1}{1+0.1} = \frac{1}{1.1} = \underline{\underline{0.90\overline{90}}}$$

The rule:

- All applicants with values p >= 0.91 will become the loan
- All applicants with values p < 0.91 doesn't become the loan

This rule doesn't depend on the specifics of the population

## Answer d)

```
def calc_p(z):
    return 1 / (1+z)
H=1
z = 0.1
p = calc_p(z)
print('p:',p)

def calc_expected_val(H, p, z):
    return (1 - p) * -H + p * z * H

loan_applications_over_threshold = loan_applicants[loan_applicants['p'] >= p]
applicant_number = len(loan_applications_over_threshold)
print('Number of accepted applicants:', applicant_number)

loan_applications_over_threshold['E'] = calc_expected_val(loan_applications_over_threshold['H'],
                                                          loan_applications_over_threshold['p'],
                                                          loan_applications_over_threshold['z'])

total_profit = loan_applications_over_threshold['E'].sum()
print('Total Profit:', total_profit)
```

```
p: 0.9090909090909091
Number of accepted applicants: 884
Total Profit: 43.587058396754784
```

## Answer e)

```
def calc_mean_profit_per_person(total, nr, H):
    return (total / nr) / H * 100

print('Profit per Person in %:', calc_mean_profit_per_person(total_profit, applicant_number, H))
loan_applications_over_threshold
```

```
Profit per Person in %: 4.93066271456502
```

| | p | H | z | E |
|---|---|---|---|---|
| **1** | 0.950714 | 1 | 0.1 | 0.045786 |
| **11** | 0.969910 | 1 | 0.1 | 0.066901 |
| **33** | 0.948886 | 1 | 0.1 | 0.043774 |
| **34** | 0.965632 | 1 | 0.1 | 0.062195 |
| **43** | 0.909320 | 1 | 0.1 | 0.000252 |
| **...** | ... | ... | ... | ... |
| **9971** | 0.976080 | 1 | 0.1 | 0.073688 |
| **9972** | 0.956992 | 1 | 0.1 | 0.052691 |
| **9978** | 0.950612 | 1 | 0.1 | 0.045673 |
| **9983** | 0.909824 | 1 | 0.1 | 0.000806 |
| **9997** | 0.946708 | 1 | 0.1 | 0.041379 |

### Answer f)

```
# 10% more profit is p * 1.1 =~ 0.99
import numpy as np
for p_i in np.arange(0.0, 1.0, 0.01):
    laot_p_i = loan_applicants[loan_applicants['p'] >= p_i]
    profit_p_i = calc_mean_profit_per_person(calc_expected_val(laot_p_i['H'], laot_p_i['p'],
                                                               laot_p_i['z']).sum(),
                                                               len(laot_p_i), 1)
    print('Profit in percentage with p:', p_i, 'is', profit_p_i)
```

```
Profit in percentage with p: 0.8200000000000001 is 0.06152165403000742
Profit in percentage with p: 0.8300000000000001 is 0.7096671205481778
Profit in percentage with p: 0.84 is 1.2095940521542554
Profit in percentage with p: 0.85 is 1.7235611437097138
Profit in percentage with p: 0.86 is 2.253825933114002
Profit in percentage with p: 0.87 is 2.8139072412593933
Profit in percentage with p: 0.88 is 3.446086765055587
Profit in percentage with p: 0.89 is 3.9046656842781737
Profit in percentage with p: 0.9 is 4.498569915862962
Profit in percentage with p: 0.91 is 5.032220199136512
Profit in percentage with p: 0.92 is 5.4804285381425
Profit in percentage with p: 0.93 is 5.924775846123662
Profit in percentage with p: 0.9400000000000001 is 6.518547795391965
Profit in percentage with p: 0.9500000000000001 is 7.143709841623931
Profit in percentage with p: 0.96 is 7.693653494773212
Profit in percentage with p: 0.97 is 8.34123429274182
Profit in percentage with p: 0.98 is 8.873753339498423
Profit in percentage with p: 0.99 is 9.448521920592828
```

## 2. Decision problem for different subpopulations

### Answer a)

```
def get_share_gender_specific(p):
    loan_applications_over_threshold = loan_applicants[loan_applicants['p'] > p]
    return len(loan_applications_over_threshold[loan_applications_over_threshold['g'] == 1]) / len(loan_applicants[loan_applicant
res = get_share_gender_specific(0.7)
print(res[0] * 100, 'percent or',
      res[0] * len(loan_applicants[loan_applicants['g'] == 1]),
      'out of',
      len(loan_applicants[loan_applicants['g'] == 1]),
      'Men and',
      res[1] * 100,
      'percent or',
      res[1] * len(loan_applicants[loan_applicants['g'] == 0]),
      'out of',
      len(loan_applicants[loan_applicants['g'] == 0]),
      'Woman')
```

- A share of **665 women of 10000 women (6.65%)** receives a loan with a threshold of $p > 0.7$.
- A share of **1484 men of 10000 men (14.85%)** receives a loan with a threshold of $p > 0.7$.

### Answer b)

- A share of **192 women of 10000 women (1.92%)** receives a loan with a threshold of $p > 0.8$.
- A share of **598 men of 10000 men (5.98%)** receives a loan with a threshold of $p > 0.8$.

## Answer c)

It is crucial what criteria, regardless of gender, are used to decide whether someone gets a loan or not.

Possible deciding criteria:

- 1: income
- 2: workload
- 3: gender based bias

So why are women less likely to receive credit?

- Cause of 1: Women tend to have a lower income
- Cause of 2: Women are more often in part-time jobs (childcare)
- Cause of 3: Because of gender based bias in banking (male dominated sector or historical bias)

## Answer d)

- Three possibility:
    - p_0 = 0 (everyone accepted)
    - different p_0 for women and men
    - optimization of decision criterias (trained features)

## Answer e)

```
def calc_share_of_gender(threshold, g, Y):
    loan_applicants_over_threshold = loan_applicants[loan_applicants['p'] > threshold]
    y_1 = loan_applicants.loc[(loan_applicants['g'] == g) & (loan_applicants['Y'] == Y), 'g'].count()
    y_1_accepted = loan_applicants_over_threshold.loc[(loan_applicants_over_threshold['g'] == g)
                                                      & (loan_applicants_over_threshold['Y'] == Y), 'g'].count()
    return y_1_accepted, np.round(y_1_accepted / y_1 * 100, 2), y_1
```

```
res = calc_share_of_gender(.7, 0, 1)
print(res[0], 'or', res[1], 'percent Woman out of', res[2], 'who actually repay are accepted')
```

```
304 or 6.25 percent Woman out of 4863 who actually repay are accepted
```

```
res = calc_share_of_gender(.7, 1, 1)
print(res[0], 'or', res[1], 'percent Men out of', res[2], 'who actually repay are accepted')
```

```
747 or 14.91 percent Men out of 5009 who actually repay are accepted
```

## Answer f)

```
res = calc_share_of_gender(.8, 0, 1)
print(res[0], 'or', res[1], 'percent Woman out of', res[2], 'who actually repay are accepted')
res = calc_share_of_gender(.8, 1, 1)
print(res[0], 'or', res[1], 'percent Men out of', res[2], 'who actually repay are accepted')
```

```
89 or 1.83 percent Woman out of 4863 who actually repay are accepted
294 or 5.87 percent Men out of 5009 who actually repay are accepted
```

## 3. Video recommendations

### Answer a)

- read a book or whatch the movie $\Rightarrow$ D
- enjoys or hates the movie $\Rightarrow$ Y

- no watching $\hat{=}$ reading a book $\hat{=}$ $D = 0$
- watching the movie $\hat{=}$ $D = 1$
- enjoys the movie $\hat{=}$ $Y = 1$
- hates the movie $\hat{=}$ $Y = 0$

Utility :
- no watching the movie $\Rightarrow$ $+0$ $\Rightarrow$ $D = 0$
- enjoys the movie $\Rightarrow$ $+1$ $\Rightarrow$ $Y = 1$
- hates the movie $\Rightarrow$ $-1$ $\Rightarrow$ $Y = 0$
- reading a book $\Rightarrow$ $+0.3$ $\Rightarrow$ $D = 0$

| Utility matrix W | Y=0 | Y=1 |
|---|---|---|
| D=0 | 0.3 | 0.3 |
| D=1 | −1 | 1 |

### Answer b)

$$E(U|D=0) = (1-p)*0.3 + p*0.3$$
$$= 0.3 - 0.3p + 0.3p$$
$$= 0.3$$
$$E(U|D=1) = (1-p)*(-1) + p*1$$
$$= -1 + p + p$$
$$= 2p - 1$$

Which means

$$E(U|D=1) > E(U|D=0)$$
$$2p - 1 > 0.3$$
$$2p > 1.3$$
$$p > 0.65$$

Answer c)

Average utility

Reading a book until $p > 0.65$

$\Rightarrow 7 \cdot 0.3 = 2.1$

Watching movie for $p = 0.7$,

$p = 0.8$, $p = 0.9$, $p = 1.0$

$\Rightarrow (2 \cdot 0.7) - 1 + (2 \cdot 0.8) - 1 + (2 \cdot 0.9) - 1$

$+ (2 \cdot 1) - 1 = 2.8$

Average $= (2.1 + 2.8) / 11 = 0.445$

Answer d)

Average biased utility

Reading a book until $p > 0.65$

$\Rightarrow 6 \cdot 0.3 = 1.8$

Watching movie for $p = 0.7$,

$p = 0.6$, $p = 0.8$, $p = 0.9$, $p = 1.0$

$\Rightarrow (2 \cdot 0.6) - 1 + (2 \cdot 0.7) - 1 + (2 \cdot 0.8) - 1$

$(2 \cdot 0.9) - 1 + (2 \cdot 1) - 1 = 3.0$

Average $= (1.8 + 3.0) / 11 = 0.436$

Difference of utility = ~0.01 (0.445 – 0.436)

This occures because at **true p = 0.6** with **biased p = 0.7** a movie would be watched which has a worse utility than **0.3** ((0.6 * 2 ) -1)

## 4. Properties of predictors

Answer a)

Accuracy for different types are:

- **A** = 0.8
- **B** = 1.0
- **C** = 0.7

Mean is (0.8 + 1 + 0.7) / 3 = **0.8333**

Answer b)

Value r corresponds to the true probability: **r = p**

| Type | r = 0 | r = 0.5 | r = 1 |
|---|---|---|---|
| A | 100 | 0 | 0 |
| B | 0 | 100 | 0 |
| C | 0 | 0 | 100 |

## Answer c)

Assumption for p=0 → 0 individuals with Y=1 (100*0)

$$N(r=0)=40$$
$$N1(r=0)=0$$
$$N1(r=0)/N(r=0)=0$$

Assumption for p=0.5 → 50 individuals with Y=1 (100*0.5)

$$N(r=.5)=220$$
$$N1(r=.5)=110$$
$$N1(r=.5)/N(r=.5)=0.5$$

Assumption for p=1 → 100 individuals with Y=1 (100*1)

$$N(r=1)=40$$
$$N1(r=1)=40$$
$$N1(r=1)/N(r=1)=1$$

This means P3 is calibrated because all the N(r)'s are equal to r.

## Answer d)

Example:

Assume we have 100 Students. Everyone of these students has a different level of intelligence. Let's say in SEP fifty percent of students pass and fifty percent fail. This means if we assign a probability of 0.5 we could think that we have a perfect probability, which is not the case because this would imply that everyone has the same probability of passing the SEP.

Therefore, the score and the individual probability are very different, although they can be quickly mixed up!