732A66 Decision Theory

Assignment 4

Johannes Hedström

STIMA
Department of Computer and Information Science
Linköpings universitet

2024-12-17

1 Question 1

Assume some budget calculations depend on whether a certain cost will be at least SEK 120 000 or lower than this amount. A reasonable model for this cost is a normal distribution with standard deviation SEK 12 000 (independent of the mean) and a mean that can be modeled as normally distributed with mean 115 000 (SEK) and standard deviation 9 000 (SEK). No trend is anticipated for this cost and for the 6 previous periods the average cost was SEK 121 000. Note that the hypotheses are about the actual cost, not the expected cost.

1.1 a

Show that the prior odds for the hypothesis that the cost will exceed SEK 120 000 (against the alternative that it will not) is about 0.59. [hint: write the observed variables a sum of two independent random variable $\tilde{x} = \tilde{\mu} + \tilde{\epsilon}$]

$$\tilde{\mu} \sim N(115000, 9000)$$

Standard deviation of $\tilde{x}$ should then be

$$\sqrt{9000^2 + 12000^2} = 15000$$

Now you have the mean and the standard deviation so you can calculate the standardized value for 120000.

$$z = \frac{120000 - 115000}{15000}$$

From the normal cumulative distribution(standard) we get that the probability for x to be less than 120 000 to be 0.63056:

so the prior odds to exceed 120000 is then

$$\frac{1 - 0.63056}{0.63056} \approx 0.59$$

```r
sd <- sqrt(9000^2+12000^2 )

p <- pnorm(120000,115000,sd) # normal distribution mean of x< 120000

p <- round(p,5)

odds <-  round(((1-p)/p),3)
paste('The probability for the cost to be less than 120 0000 is', p)
```

```
## [1] "The probability for the cost to be less than 120 0000 is 0.63056"
```

```r
paste('The prior odds for the cost to exceed 120 000 is',odds) # odds
```

```
## [1] "The prior odds for the cost to exceed 120 000 is 0.586"
```

rounds to 0.59.

1.2 b

Show that the Bayes factor (considering the average cost for the previous 6 periods) for the hypothesis that the cost will exceed SEK 120 000 (against the alternative that it will not) is about 1.63 [Not about 1.60 or about 1.70].

$$B = \frac{Odds(H_0|data, I)}{Odds(H_A|data, I)} = \frac{P(H_0|data, I)/P(H_A|data, I)}{P(H_0|I)/P(H_A|I)}$$

$$H_A : \text{The cost is lower or equal to } 120000$$

$$H_0 : \text{The cost is exceeds } 120000$$

Prior odds for $H_0|I$ is calculated in a.

As the prior and the data are normal, the posterior is normal.

The posterior is proportional to :

$$N(\theta|\mu_n, \tau_n^2)$$

where

$$\frac{1}{\tau_n^2} = \frac{n}{\sigma^2} + \frac{1}{\tau_0^2}$$

n = 6 and σ_n^2 = 15000^2

$$\mu_n = w\bar{x}(1-w)\mu_o$$

where $\bar{x} = 121000$

and w :

$$w = \frac{\frac{n}{\sigma_n^2}}{\frac{n}{\sigma_n^2} + \frac{1}{\tau_0^2}}$$

```
datamean <- 121000

sigma <- 15000

tau2 <- 1/(6/sigma^2 + 1/15000^2)
w <- (6/sigma^2)/((6/sigma^2)+1/15000^2)
mu_n <- w*121000 + ((1 - w)*115000)

z_data <- (mu_n-120000) / sqrt(tau2)
```

```
HA <-  pnorm(z_data)  # P(cost <= 120000 | data)
H0 <- 1 - HA


bayes_factor <- (H0/HA)/odds

bayes_factor
```

```
## [1] 1.639227
```

1.3 c

If the loss of accepting the hypothesis that the cost will be lower than SEK 120 000 while the opposite will be true is SEK 4 000, and the loss of accepting the hypothesis that the cost will be at least SEK 120 000 while the opposite will be true is SEK 6 000, which decision should be made for the budget (according to the rule of minimizing the expected loss)?

The probabilities are taken from the variables HA and H0 created in question b.

$$EL[lower] = 0 \cdot 0.5113018 + 0.4886982 \cdot 4000 = 1954.793$$

$$EL[higher] = 6000 \cdot 0.5113018 + 0.4886982 \cdot 0 = 3067.811$$

Action 1, has the lowest loss and is the optimal action according to the rule of minimizing the expected loss.

2 Question 2

Consider a big box filled with an enormous amount of poker chips. You know that either 70% of the chips are red and the remainder blue, or 70% are blue and the remainder red. You must guess whether the big box has 70% red / 30% blue or 70% blue / 30% red. If you guess correctly, you win US dollar 5 . If you guess incorrectly, you lose US $ 3 . Your prior probability that the big box contains 70% red / 30% blue is 0.40, and you are risk neutral in your decision making (i.e. your utility is linear in money).

2.1 a

If you could purchase sample information in the form of one draw of a chip from the big box, how much should you be willing to pay for it? Assume now that the cost of sampling is US$0.25 (i.e. 25 US cents) per draw.

```
d_matrix <- data.frame('actions'=c('Guess 70 % red', 'Guess  70 % blue'), 'State 1(70%red)'=c(5,-3), 'St

colnames(d_matrix) = c('actions','State 1(70%red)','State 2(70%blue)')
knitr::kable(d_matrix)
```

actions	State 1(70%red)	State 2(70%blue)
Guess 70 % red	5	-3
Guess 70 % blue	-3	5

Expected utility:

$$\text{Action red} = 5 \cdot 0.4 - 3 \cdot 0.6 = 0.2$$

$$\text{Action blue} = -3 \cdot 0.4 + 5 \cdot 0.6 = 1.8$$

Optimal prior action is to guess for 70% blue.

Posterior distribution:

$$P(\theta|BUY(sample)) = \frac{P(BUY|\theta) \cdot P(\theta)}{P(BUY)}$$

Where θ is the states.

Probability of drawing a red chip then (blue is then the opposite):

$$P(Red) = 0.7 \cdot 0.4 + 0.3 \cdot 0.6 = 0.46$$

Posterior probabilities:

$$P(State1|Red) = \frac{0.7 \cdot 0.4}{0.46} = 0.6087$$

$$P(State2|Red) = 1 - P(State1|Red) = 0.3913$$

$$P(State1|Blue) = \frac{0.3 \cdot 0.4}{1 - 0.46} = 0.2222$$

$$P(State2|Blue) = 1 - P(State1|Blue) = 0.7778$$

Expected utility after sample

Guessing 70% red (after observing red):

$$E[R(Red) \mid Red] = 5 \cdot 0.6087 - 3 \cdot 0.3913 = 1.8696$$

Guessing 70% blue (after observing red):

$$E[R(Blue) \mid Red] = -3 \cdot 0.6087 + 5 \cdot 0.3913 = 0.1304$$

Guessing 70% red (after observing blue):

$$E[R(Red) \mid Blue] = 5 \cdot 0.2222 - 3 \cdot 0.7778 = -1.889$$

Guessing 70% blue (after observing blue):

$$E[R(Blue) \mid Blue] = -3 \cdot 0.2222 + 5 \cdot 0.7778 = 3.2223$$

When utility is linear in money, we could also define VSI as

$$VSI(y_n) = E(R(a^{''}|y_n)|y_n) - E(R(a^{'})|y_n)$$

Posterior expected utility from posterior optimal action - Posterior expected utility with optimal action from prior.

So the VSI in this case is:

$$VSI(Red) = 1.8696 - 0.1304 = 1.7392$$

No change in the action compared to prior

$$VSI(Blue) = 3.2223 - 3.2223 = 0$$

Expected value of sample information(EVSI)

$$EVSI = \sum_y VSI(y_n) \cdot f(y_n) dy_n$$

$$EVSI = 0.46 \cdot 1.7392 + 0.54 \cdot 0 = 0.800032$$

The amount you should be willing to pay for information is 0.800 \$.

2.2 b

What is the ENGS for a sample of 10 chips using a single-stage sampling plan.

Expected net gain of sampling as a function of the sample size n:

$$ENGS(n) = EVSI(n) - CS(n)$$

Cost of 10 draws is 2.5\$

The only samples that will make us change from the prior actions are the ones that have more red chips than blue, so we only have to consider the probabilities from drawing 6, 7, 8, 9 and 10 red chips out och 10 chips.