Assignment 3

Wuhao Wang(wuhwa469)

2022-12-28

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 modelled 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.

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.

The two hypothesis are:

$$H_0 : \tilde{x} < 120000$$
$$H_1 : \tilde{x} \geq 120000$$

And the prior odds for H_1 is:

$$PO(H_1) = \frac{P(H_1)}{P(H_0)}$$

According to the *hints*, $\tilde{x}$ can be considered as sum of two variables:

$$\tilde{\mu} \sim N(115000, 9000^2)$$
$$\tilde{\epsilon} \sim (0, 12000^2)$$
$$\tilde{x} = \tilde{\mu} + \tilde{\epsilon}$$

The sum of two normal distribution is still a normal distribution, so the distribution of $\tilde{x}$ is:

$$\tilde{x} \sim N(115000, 12000^2 + 9000^2)$$

```
PR_H0 <- pnorm(120000, 115000, 15000)
PR_H1 <- 1-PR_H0
PO_H1 <- PR_H1/PR_H0
PO_H1
```

```
## [1] 0.5858953
```

The result shows it is close to 0.59.

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.

The Bayes factor is calculated from the posterior odds, so first we need to obtain the posterior distribution of two hypothesis. According to Bayes theory, we have:

the given distribution of $\tilde{x}$ is:

$$\tilde{x}|\theta, \sigma^2 \sim N(\theta, \sigma^2)$$

the prior of θ is:

$$\theta \sim N(\mu_0, \tau_0^2)$$

the posterior distribution is:

$$p(\theta|\tilde{x}) \propto N(\mu_n, \tau_n^2)$$

where

$$\begin{aligned}\frac{1}{\tau_n^2} &= \frac{n}{\sigma^2} + \frac{1}{\tau_0^2} \\ \mu_n &= w * x + (1-w)\mu_0 \\ w &= \frac{n/\sigma^2}{n/\sigma^2 + 1/\tau_0^2}\end{aligned}$$

From the given information, we have :

$$\begin{aligned}\sigma &= 12000^2 \\ \mu_0 &= 115000 \\ \tau_0^2 &= 9000^2 \\ x &= 121000 \\ n &= 6\end{aligned}$$

plug in all the information to formula, we can get the post and then calculate the Bayes Factor

```
w <- (6/(12000^2))/((6/(12000^2))+(1/(9000^2)))
mu_n <- w*121000 + (1-w)*115000
tau_n <- sqrt(1/((6/(12000^2))+(1/(9000^2))))
H0_post <- pnorm(120000, mu_n, sqrt(tau_n^2+12000^2))
H1_post <- 1-H0_post
PS_H1 <- H1_post/H0_post
PS_H1/PO_H1 # bayes factor
```

```
## [1] 1.629247
```

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)?

$$EL(Accept\ H1) = 6000 * Pr(H_0|x)$$
$$EL(Do\ not\ reject\ H_0) = 4000 * Pr(H_1|x)$$

From **problem b**, we know the

$$Pr(H_0|x) = 0.5116$$
$$Pr(H_1|x) = 0.4884$$

So,

$$EL(Accept\ H1) = 3069.6$$
$$EL(Do\ not\ reject\ H_0) = 1953.6$$

the H_0 (the certain cost will be less than120000) should not be rejected.

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$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).

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?

From the prior knowledge, we have

$$E\prime R(70r, 30b) = 5 * 0.4 - 3 * 0.6 = 0.2$$
$$E\prime R(70b, 30r) = 5 * 0.6 - 3 * 0.4 = 1.8$$

So, **Guess 70b,30r** is optimal action ($a\prime$)in prior context.

now, if we take a draw from the box and the chip is red, then posterior is:

$$P(70r, 30b|c = r) = \frac{0.7 * 0.4}{0.7 * 0.4 + 0.3 * 0.6} = 0.61$$
$$P(70b, 30r|c = r) = \frac{0.3 * 0.6}{0.7 * 0.4 + 0.3 * 0.6} = 0.39$$

also consider when the chip we draw is blue:

$$P(70r, 30b|c = b) = \frac{0.3 * 0.4}{0.3 * 0.4 + 0.7 * 0.6} = 0.22$$
$$P(70b, 30r|c = b) = \frac{0.7 * 0.6}{0.3 * 0.4 + 0.7 * 0.6} = 0.78$$

Then, we have

$$
\begin{aligned}
E''R(70r, 30b|c = r) &= 5 * 0.61 - 3 * 0.39 = 1.88 \\
E''R(70b, 30r|c = r) &= 5 * 0.39 - 3 * 0.61 = 0.12 \\
E''R(70r, 30b|c = b) &= 5 * 0.22 - 3 * 0.78 = -1.24 \\
E''R(70b, 30r|c = b) &= 5 * 0.78 - 3 * 0.22 = 3.24
\end{aligned}
$$

We make a notation:

$$
\begin{aligned}
a''|c = r \quad &-> Guess\ 70r, 30b \\
a''|c = b \quad &-> Guess\ 70b, 30r
\end{aligned}
$$

we have also

$$
a' \quad -> Guess\ 70b, 30r
$$

$$
\begin{aligned}
VSI(r) &= E''R(a''|c = r) - E''R(a'|c = r) = 1.88 - 0.12 = 1.76 \\
VSI(b) &= E''R(a''|c = b) - E''R(a'|c = b) = 3.24 - 3.24 = 0
\end{aligned}
$$

EVSI=1.76*(0.7*0.4+0.3*0.6)+0*(0.4*0.3+0.7*0.6)=0.8096

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

the table is too long to show, see attached file.

Question 3

Curiosity: Quite recently (2019) the international prototype kilogram (IPK) in Paris was replaced by a definition in terms of the Planck constant.Assume we have a scale with no systematic error and that we have weighed the IPK 3 times and obtained the values 1.0076, 1.0015 and 0.9971 (kg)

a) problem is too difficult to type here.

The Inverse Gamma is conjugate prior of normal distribution. The new parameters are

$$
\begin{aligned}
\alpha' &= \alpha + \frac{n}{2} \\
\beta' &= \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2
\end{aligned}
$$

Since we are using 0-1 loss here, the point estimator should be posterior mode, which is $\frac{\beta}{\alpha+1}$

```
sample <- c(1.0076, 1.0015, 0.9971)
n <- 3
alpha <- 2
beta <- 10^-5
alpha_pr <- alpha + (n/2)
beta_pr <- beta + 0.5*sum((sample-1)^2)
beta_pr/(alpha_pr+1)
```

```
## [1] 9.824444e-06
```

b) basically redo the last one with new prior.

now the new parameters are

$$\alpha\prime = \frac{n}{2}$$
$$\beta\prime = \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2$$

And the posterior mode is:

```
alpha_pr <- (n/2)
beta_pr <- 0.5*sum((sample-1)^2)
beta_pr/(alpha_pr+1)
```

```
## [1] 1.3684e-05
```