Problem 1

a) To find the true probability we use the formula $p = \frac{1}{TO}$;

Norge - Kyprus

$$\frac{1}{1.30} = 0.7692\ ;\ \frac{1}{4.10} = 0.2439\ ;\ \frac{1}{8.10} = 0.12345\ ;$$

Tyskland - Frankrike

$$\frac{1}{2.35} = 0.4255\ ;\ \frac{1}{3.0} = 0.33\ ;\ \frac{1}{2.65} = 0.3773\ ;$$

Portugal - Croatia

$$\frac{1}{2.20} = 0.4545\ ;\ \frac{1}{2.90} = 0.3448\ ;\ \frac{1}{2.85} = 0.3508\ ;$$

Nederland - Peru

$$\frac{1}{1.60} = 0.625\ ;\ \frac{1}{3.40} = 0.2941\ ;\ \frac{1}{4.45} = 0.2247\ ;$$

	1	X	2	Sum
Norge-Kyprus	0.7692	0.2439	0.1234	1.1365
Tyskland-Frankrike	0.4255	0.33	0.3773	1.1328
Portugal-Croatia	0.4545	0.3448	0.3508	1.1501
Nederland-Peru	0.625	0.2941	0.2247	1.1438

Looking at the result we can say that the sum of all probabilities are larger than 1 and vary from 1.1328 to 1.1501.

b)

We can set up a simulation algorithm following the next steps:

1) First, we generate game outcomes based on the given probabilities.

2) Then for every game with the right outcome we calculate the winnings following the formula $winnings = odds * stake$.

3) We sum all the $winnings$.

4) Calculate the profit with formula $profit = winnings - money\ spent$

d) Calculation of the expected values of P_s and P_c

$$E(P_s) = \sum_{i=1}^{n} S * TO_i * E(X_i) - nS;$$

$$E(P_c) = nS * E(X_i) * \left(\prod_{i=1}^{n} * TO_i\right) - nS$$

Problem 2

a) With the integral $\int_0^{24} \lambda(u)$ we want to calculate the expected number of visitors per day. By generating this number a large amount of times we can make probability estimations of x amount of visitors arriving during a day.

From the lectures we know that hit or miss method is obtained by simulating $X_1, X_2, \dots, X_n\ iid\ U[a,b]$ and $Y_1, Y_2, \dots, Y_n\ iid\ U[0,c]$ and calculation of:

$$\theta_{HM} = \frac{c(b-a)}{n} \sum_{i=1}^{n} I(Y_i \leq \lambda(X_i))$$

To find c we'll have to make a plot of the $\lambda(t)$. On the horizontal axis we will have the edges going from 0 to 24 and on the vertical axis edges going from 0 to c where $c \geq \lambda(t)$ for any of the $t \in [0,24]$. From the plot made in R we can estimate that a safe choice for c will be 210.

Thus the hit or mis method can be obtained by simulating $X_1, X_2, \dots, X_n\ iid\ U[0,24]$ and $Y_1, Y_2, \dots, Y_n\ iid\ U[0,210]$ and calculation of:

$$\hat{\theta}_{HM} = \frac{210 * 24}{n} \sum_{i=1}^{n} I(Y_i \leq \lambda(X_i))$$

Where 210*24 will be the area of the plot and $\frac{\sum_{i=1}^{n} I(Y_i \leq \lambda(X_i))}{n} = \hat{p}$ is the proportion of the uniformly distributed points inside the plot area that are within the $\lambda(x)$ function.

From the lectures we have that:

$$\widehat{SD}\left(\hat{\theta}HM\right) = c(b-a)\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

Which implies that to have $(1-\alpha)100\%$ probability that $\hat{\theta}_{CMC}$ is at most a margin e from θ we need

$$z_{a/2} c(b-a)\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} < e\ ;$$

$$n < \frac{z_{a/2}^2\ c^2(b-a)^2 \hat{p}(1-\hat{p})}{e^2}$$

We can estimate $\hat{p}$ by simulating the first simulation but we can also use $\hat{p}(1-\hat{p}) \leq 0.25$ since $0 \leq \hat{p} \leq 1$.

So, finally we have c = 210, a = 0, b = 24, e = 10 and $z_{a/2}$= 1.96. With that we have

$$n < \frac{1.96^2 * 210^2 * 24^2 * 0.25}{10^2} = 243956.9$$

To be more precise the first estimate of $\hat{p}$ can be used.

c) Thinning method is more suitable because it applies when $\lambda(t)$ is bounded on the interval of interest. Let $[a, b]$ be the interval of interest if $\lambda(t) \leq \lambda max$ for all $t \in [a, b]$.

A reasonable choice for λmax would be 210 based on all $\lambda(t)$ $where\ t \in$ $[0,24]$.

This proportion of times that will be deleted: $u_i \geq \lambda(s_i)/\lambda max$

i) To estimate the probability that the maximum number of active visitors exceeds a certain number a, we have to simulate the number of active visitors at the website during the day a large amount of times. And then find the proportion of days when the number of active visitors exceeded a certain number a.

ii) For medium and quantiles we use built in r functions. To calculate the number of active visitors at certain time point we have to run through the arrival and departure times and record the time and the number of customers in the system at certain time point t where there is a new arrival or departure.

e) From the lectures we know that to simulate an nonhomogeneous Poisson process we first have to simulate homogeneous Poisson process with intensity 1. Then we have to transform all the simulated arrival times by the inverse of the integrated intensity function.

So with the given intensity $\lambda_2 = 10 + 10t$, first let $\Lambda(t) = \int_0^t \lambda(u)du$.

$$\Lambda(t) = \int_0^t 10 + 10t dt = 5t^2 + 10t$$

Then we have to calculate the inverse of that function.

$$5t^2 + 10t = \omega$$

$$t^2 + 2t = \frac{\omega}{5}$$

$$(t+1)^2 - 1^2 = \frac{\omega}{5}$$

$$(t+1)^2 = \frac{\omega}{5} + 1$$

$$t + 1 = \sqrt{\frac{\omega}{5} + 1}$$

$$t = \sqrt{\frac{\omega}{5} + 1} - 1$$

Hence, $\Lambda^{-1}(\omega) = \sqrt{\frac{\omega}{5} + 1} - 1$

Now we have to calculate the amount of arrival times needed. So the expected number of events in the time interval 0 to 24 will be $\Lambda(24) = 5 * 24^2 + 10 * 24 = 3120$. We can generate N = 4000 arrival times to be certain that the largest simulated arrival time is larger than 24.

The algorithm then will be:

1. Simulate N = 4000 arrival times of an HPP with intensity 1, giving the simulated arrival times $\omega 1, \omega 2, \dots, \omega N$.

2. Calculate the arrival times $s1 = \sqrt{\frac{\omega 1}{5} + 1} - 1$, $s2 = \sqrt{\frac{\omega 2}{5} + 1} - 1$,

$sN = \sqrt{\frac{\omega N}{5} + 1} - 1$.

3. Delete the arrival times larger than 24.

4. Return $s_1, s_2, \dots, s_n$ where s_n is the largest arrival time smaller than 24.