MI-SPI 2013 – Domácí úkol č.1

Vedoucí týmu: Ondřej Paška (paskaond, 109)
Členové týmu: Tomáš Sušánka (susantom, 109), Jan Tvrdík (tvrdija4, 109)
Datum: 21.4.2015

1. Generování náhodného výběru a grafické ověřování jeho rozdělení:

1.I. Vygenerujte n = K*20 náhodných hodnot z rozdělení Exp(L) pomocí inverze distribuční funkce

Příkazy R dle instrukcí:

```
# Ondřej Paška
K = 6
L = 5
n = K*20

## 1

# The Uniform Distribution
u = runif(n, 0, 1)
u

# 1.I
# https://en.wikipedia.org/wiki/Exponential_distribution#Generating_exponential_variates
# log is natural log
x=-log(1-u)/L
x
```

- Výstup:

```
# u:
  [1] 0.243033744 0.010115521 0.333420235 0.101698445 0.434505911 0.636949205
  [7] 0.374808080 0.203311003 0.651506090 0.363400390 0.454744932 0.655830760
 [13] 0.190080107 0.168193895 0.102758160 0.053737915 0.253799674 0.204295923
 [19] 0.954511327 0.980875648 0.564616147 0.545511848 0.089053544 0.665786904
 [25] 0.945082399 0.413256222 0.118907021 0.300749463 0.304630415 0.524237602
 [31] 0.295175667 0.053908952 0.705883287 0.363084009 0.801324788 0.412771302
 [37] 0.544723849 0.336823134 0.105908954 0.324685741 0.680162727 0.361404799
 [43] 0.393721589 0.627131239 0.915163298 0.184334674 0.135721084 0.118556767
 [49] 0.218073163 0.556510590 0.578077265 0.450966891 0.479373802 0.065915993
 [55] 0.465728337 0.912873635 0.890855337 0.531048209 0.753786653 0.218783895
 [61] 0.254752929 0.373201734 0.579611704 0.350548999 0.958184577 0.748896915
 [67] 0.296224886 0.620510025 0.311832978 0.606646791 0.785579793 0.203857383
 [73] 0.764848048 0.751940470 0.369126887 0.624587501 0.888075179 0.198239201
 [79] 0.588137144 0.158527769 0.893685594 0.757955102 0.805506271 0.747274054
 [85] 0.716130599 0.066296686 0.958020038 0.712029431 0.990927031 0.792316538
 [91] 0.328184032 0.285628745 0.377804339 0.492527903 0.658124053 0.295524138
 [97] 0.007172716 0.906984000 0.497910799 0.093567505 0.804455474 0.656665123
[103] 0.577826630 0.450874576 0.399898923 0.151746929 0.417055239 0.087237704
[109] 0.423562906 0.306856987 0.459123811 0.578713486 0.432576970 0.863098788
[115] 0.214378593 0.416185759 0.505062947 0.162460457 0.917699744 0.622683484

# x:
  [1] 0.055687321 0.002033406 0.081119094 0.021449892 0.114011087 0.202642504
  [7] 0.093939321 0.045458179 0.210826905 0.090322875 0.121300316 0.213324353
 [13] 0.042163987 0.036831183 0.021685969 0.011047141 0.058552236 0.045705585
 [19] 0.618058385 0.791358553 0.166305443 0.157716687 0.018654232 0.219195296
 [25] 0.580384277 0.106633410 0.025318424 0.071549236 0.072662360 0.148567342
 [31] 0.069961336 0.011083294 0.244755722 0.090223503 0.323216778 0.106468186
 [37] 0.157370224 0.082142712 0.022389534 0.078515426 0.227988587 0.089696903
 [43] 0.100083195 0.197305754 0.493405404 0.040750230 0.029171949 0.025238935
 [49] 0.049198820 0.162616271 0.172586615 0.119919306 0.130544593 0.013637780
 [55] 0.125370168 0.488079149 0.443016219 0.151451062 0.280311371 0.049380693
 [61] 0.058807895 0.093426107 0.173315296 0.086325577 0.634898008 0.276378345
```

```
 [67] 0.070259283 0.193785420 0.074744741 0.186609464 0.307963521 0.045595389
 [73] 0.289504674 0.278817304 0.092130105 0.195945972 0.437985576 0.044188994
 [79] 0.177412972 0.034520453 0.448270896 0.283726408 0.327471072 0.275089919
 [85] 0.251848201 0.013719308 0.634112573 0.248979399 0.940491141 0.314348035
 [91] 0.079554167 0.067270497 0.094900134 0.135662710 0.214661467 0.070060242
 [97] 0.001439713 0.474996751 0.137795497 0.019647744 0.326393434 0.213809798
[103] 0.172467844 0.119885681 0.102131435 0.032915251 0.107932569 0.018255957
[109] 0.110177813 0.073303787 0.122912976 0.172888424 0.113330034 0.397699138
[115] 0.048256055 0.107634486 0.140664938 0.035457360 0.499476213 0.194934176
```

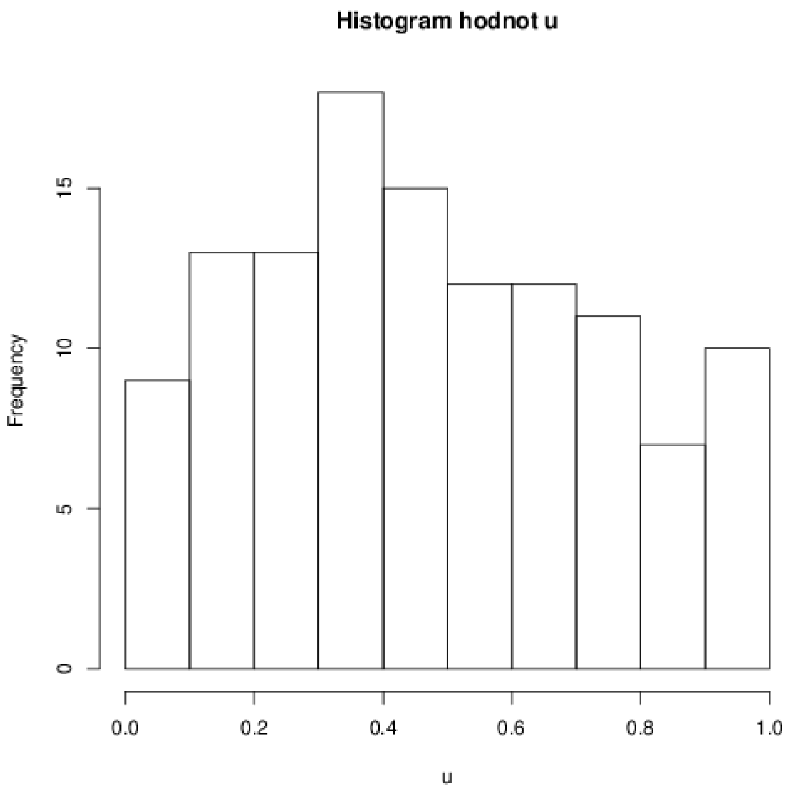
- Do proměnné *u* jsme pomocí příkazu výše uložili 120 náhodných rovnoměrně rozdělených hodnot v intervalu (0, 1). Pomocí vzorce $x = \frac{-\ln(1-u)}{L}$ jsme je následně transformovali do *x*. Tyto hodnoty **odpovídají rozdělení *Exp(L)***.

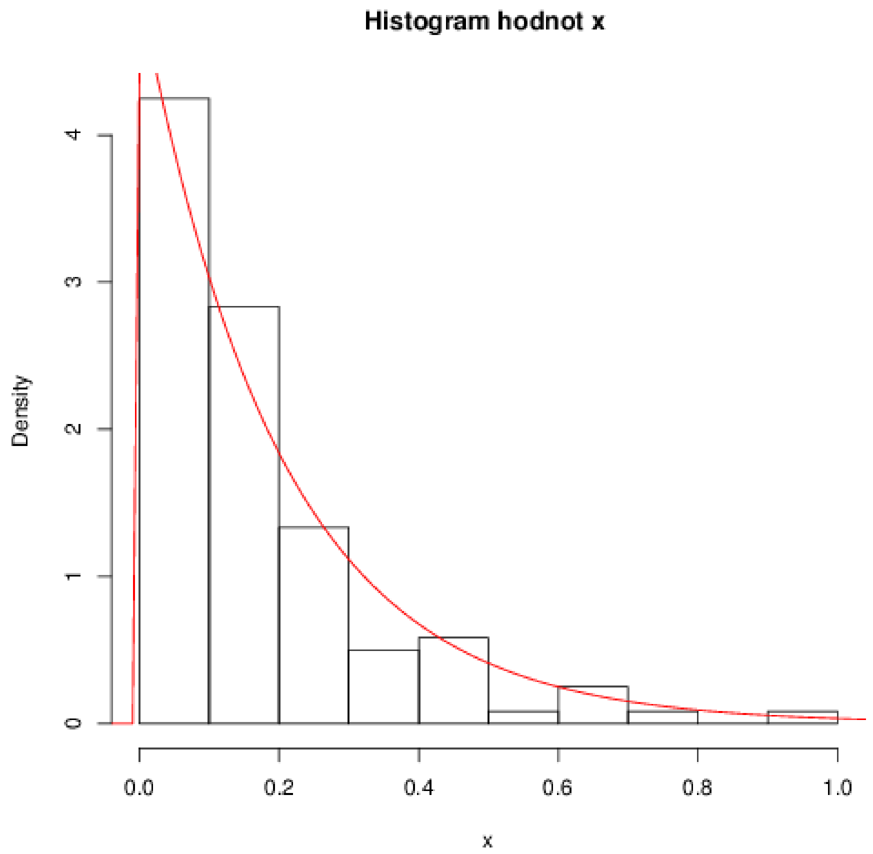
1.I. Vytvořte histogramy dat v u a x. Histogram x zobrazte spolu s grafem hustoty rozdělení Exp(L)

Příkazy R dle instrukcí:

```
hist(u, main="Histogram hodnot u")

hist(x, probability=TRUE, main="Histogram hodnot x ")
xWidth=max(x)-min(x)
xGrid=seq(min(x)-0.2*xWidth,max(x)+0.2*xWidth,length=n)
lines (xGrid,dexp(xGrid, rate=L), col='red')
```

- Výstup





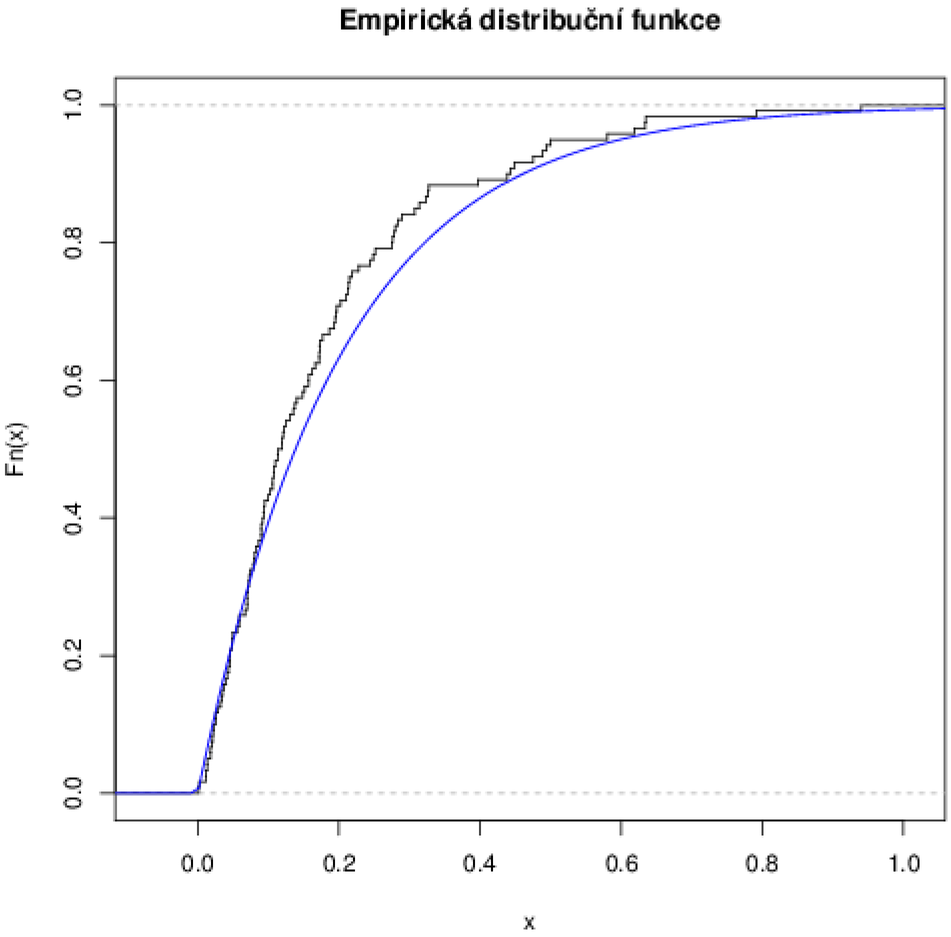
- Vytvořili jsme histogramy dat v *u* a *x*. Histogram *x* jsme doplnili grafem hustoty rozdělení *Exp(L)*. Na druhém grafu tedy porovnáváme hodnoty vygenerované pomocí vzorce, který jsme definovali výše, s grafem hustoty exponenciálního rozdělení. Toto rozdělení je vyzobrazeno červenou barvou.
- Na grafu lze vidět **menší odchylky**, které jsou dané relativně **malým vzorkem dat**, které generujeme do proměnné *u*. Odchylky však nejsou veliké a pro demonstraci je počet náhodných hodnot dostačující.

1.II. Vygenerujte graf empirické distribuční funkce pro data z x spolu s grafem distribuční funkce rozdělení Exp(L)

Příkazy R dle instrukcí:

```
plot(ecdf(x), verticals=TRUE, do.points = FALSE, main="Empirická distribuční funkce")
lines (xGrid,pexp(xGrid, rate=L), col='blue')
```

- Výstup



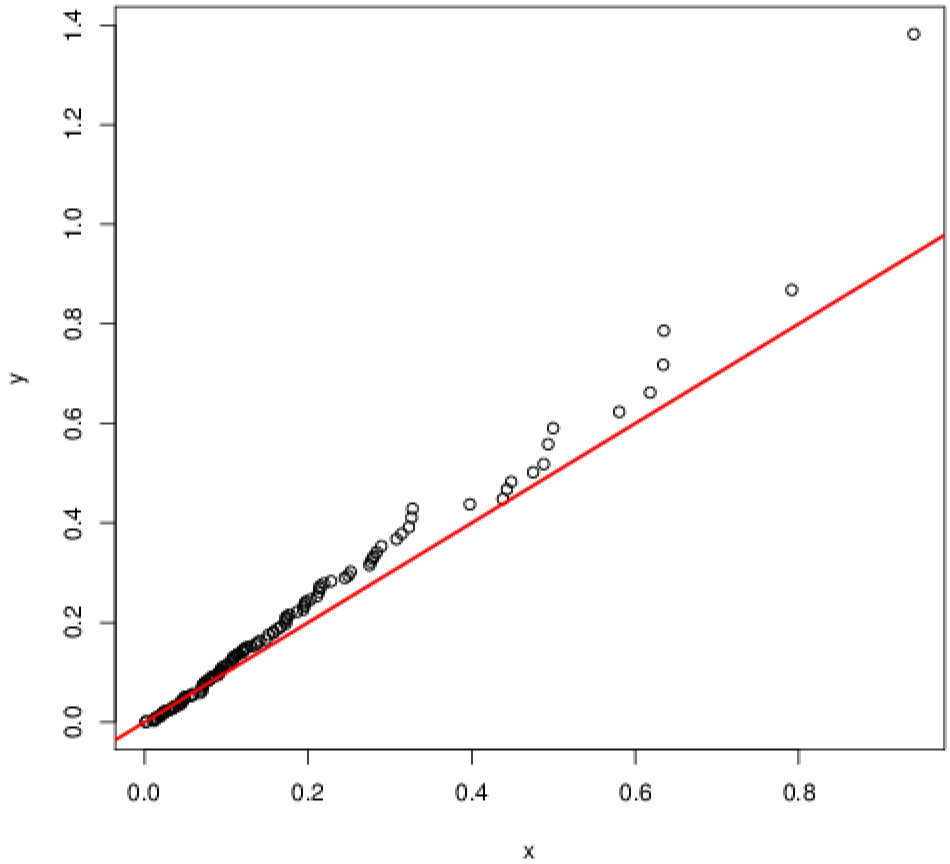
- Graf vyzobrazuje empirickou distribuční funkci pro data z proměnné x (černě) a graf distribuční funkce exp. rozdělení (modrou barvou). Porovnáme-li obě křivky je vidět, že vygenerovaná data **kopírují graf distribuční funkce** rozdělení *Exp(L)*. Náhodně vygenerovaná data mají pouze v intervalu 0.1 – 0.35 trochu vyšší zastoupení.

1.III. Vygenerujte „pravděpodobnostní papír" pro porovnání rozdělení dat x s rozdělením Exp(L)

Příkazy R dle instrukcí:

```
y=rexp(1000, rate=L)
qqplot(x, y);
abline(0, 1, col='red', lwd=2)
```

- Výstup



- Na grafu je vidět, že naše náhodná data *x* opravdu **tvoří exponenciální rozdělení**, neboť leží na jedné ose. Také lze pozorovat, že **vyšší hodnoty jsou zastoupeny výrazně méně** než ty nižší, což je dalším důsledkem, že se jedná o exp. rozdělení.

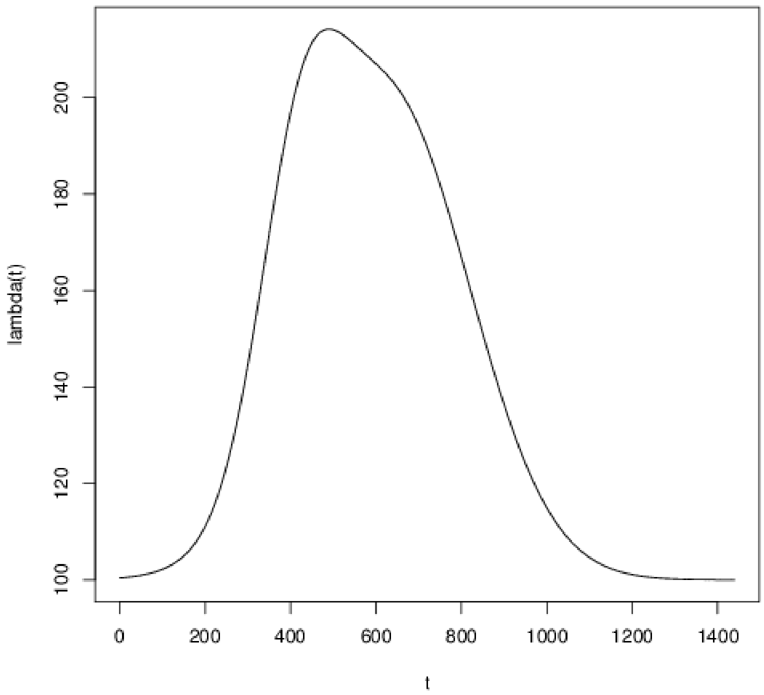
2. Generování nehomogenního Poissonova procesu a grafické ověřování jeho rozdělení

2.I. Uvažujte příchod požadavků na webový server dle Poissonova procesu s proměnlivou intenzitou lambda(t) příchodů za minutu

Příkazy R dle instrukcí:

```
lambda = function(t) { 100 + 50/exp((t - 420)^2/(3600*L)) + 100/exp((L*(t - 480 -
30*L)^2)/360000)}
size=24*60

# generates minutes
t=seq(0,size-1)
plot(t, lambda(t), type='l')
```

- Výstup – Intenzita dle Poissonova rozdělení



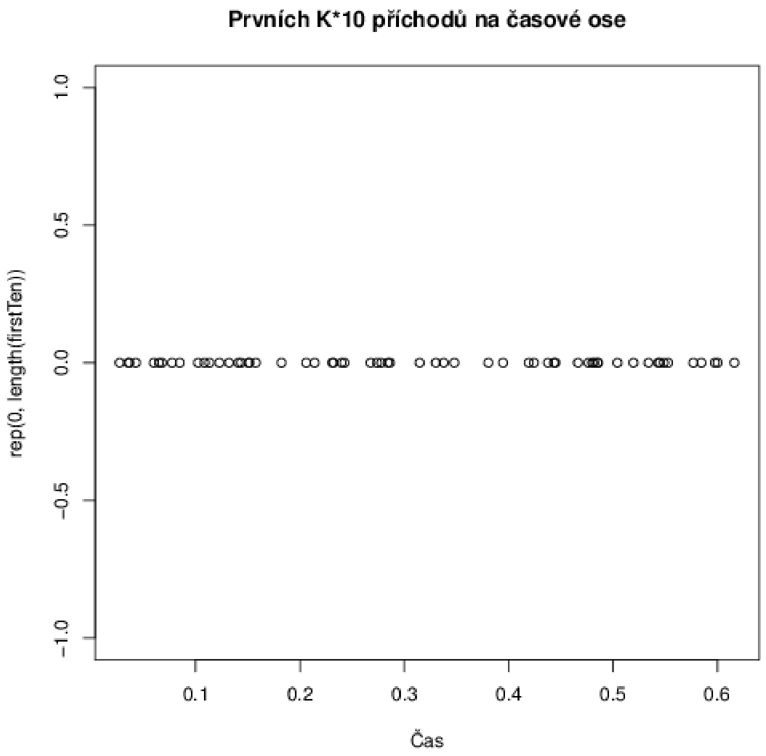
Graf ukazuje rozdělní příchozích požadavků na webový server dle daného rozdělení během dne. Nejvíce požadavků **přichází kolem 7 hodiny**, poté počet klesá až ke 100 požadavkům za hodinu.

2.II. Vygenerujte časy příchodů požadavků dle tohoto Poissonova procesu pro první den, čili pro první peridu 24*60 minut. Zobrazte prvních K*10 příchodů na časové ose.

Příkazy R dle instrukcí:

```
result=list()
t=0
i=0
while(t < size)
{
  lambdaT=lambda(t)

  t = t + rexp(1, lambdaMax)

  # random
  s = runif(1, min=0, max=1)
  if (s <= lambdaT/lambdaMax)
  {
         i = i + 1
         result[i] = t
  }
}

xWidth=max(unlist(result))
xGrid=seq(min(x)-0.1*xWidth,max(x)+0.1*xWidth,length=K*10)

firstTen = result[0:(K*10)]
plot(firstTen,rep(0, length(firstTen)), xlab="Čas", main="Prvních K*10 příchodů na
časové ose")
```

- Výstup



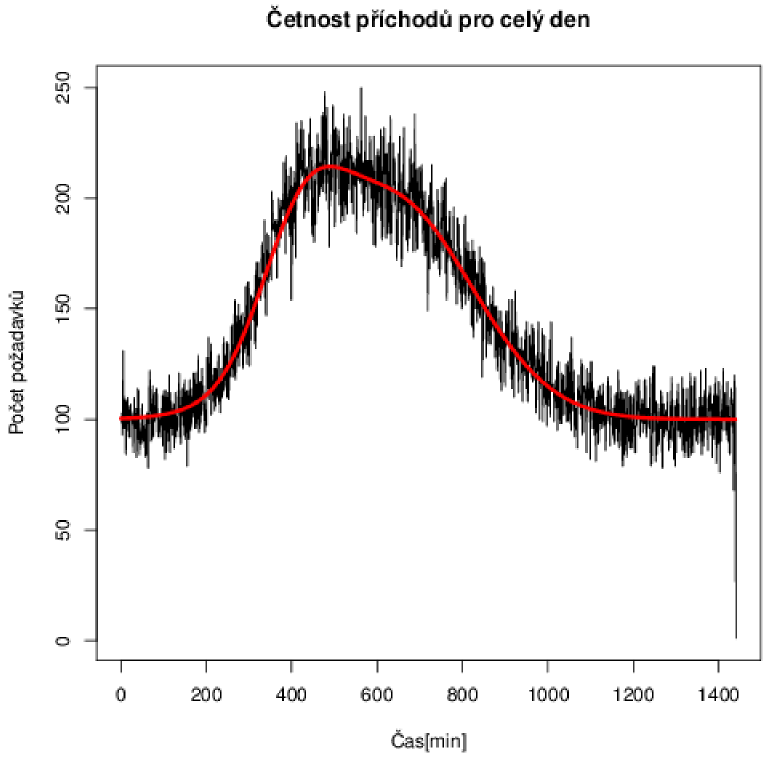
Tento graf zobrazuje prvních 60 příchozích požadavků. **Prodleva mezi dvěma požadavky je určena exponenciálním rozdělením**. Samotné rozložení požadavků je tudíž rovnoměrné. Dle grafu intenzity, by v první minutě mělo přijít přibližně 100 požadavků na server. Prvních 60 tedy v čase 0.6 minuty, graf tomuto stavu odpovídá.

2.III. Zobrazte četnosti příchodů pro celý den, t.j. pro všechy vygenerované příchody během peridy 24*60 minut. Četnosti zobrazte spolu s grafem intenzity lambda(t), podobně jako v následující ilustraci, kde červená křivka odpovídá předepsané intenzitě lambda(t).

Příkazy R dle instrukcí:

```
t=seq(0,size-1)
#Vykresleni krivky za pomoci dat z histogramu
h=hist(unlist(result), breaks=size, plot=FALSE)
plot(h$mids, h$counts, type="l", xlab = "Čas[min]", ylab="Počet požadavků",
main="Četnost příchodů pro celý den")
lines(t, lambda(t), col="red", lwd=3)
```

- Výstup



2.IV. Diskutujte kvalitu vygenerovaných dat. Diskutujte efektivitu algoritmu z pohledu ztráty generovaných náhodných hodnot.

Graf zobrazuje příchody za 24 hodin, je vidět, že **dobře kopírují teoretické hodnoty**. Odchylka od křivky dosahuje přibližně 25 příchodů / h.

3. Simulace internetového obchodu s tokem nákupů dle nehomogenního Poissonova procesu

3.I. Uvažujte internetový obchod s tokem objednávek dle nehomogenního Poissonova procesu z předchozího bodu. Ze všech zákazníků K/(K+L)*100% použije kurýrní službu, ostatní zvolí pro doručení objednávky státní poštu. Rozhodnutí o druzích doručení jsou náhodná a nezávislá mezi zákazníky i nezávislá od času zadání objednávky.

3.II. Použijte časů příchodů Poissonova procesu vygenerovaných v předchozím bodě. Rozdělte tento proces na dva procesy:

Příkazy R dle instrukcí:

```
courierProb = K/(K+L)

postList = list()

courierList = list()

i=0


for (i in 0:size)

{

        lambdaT = lambda(i)
```

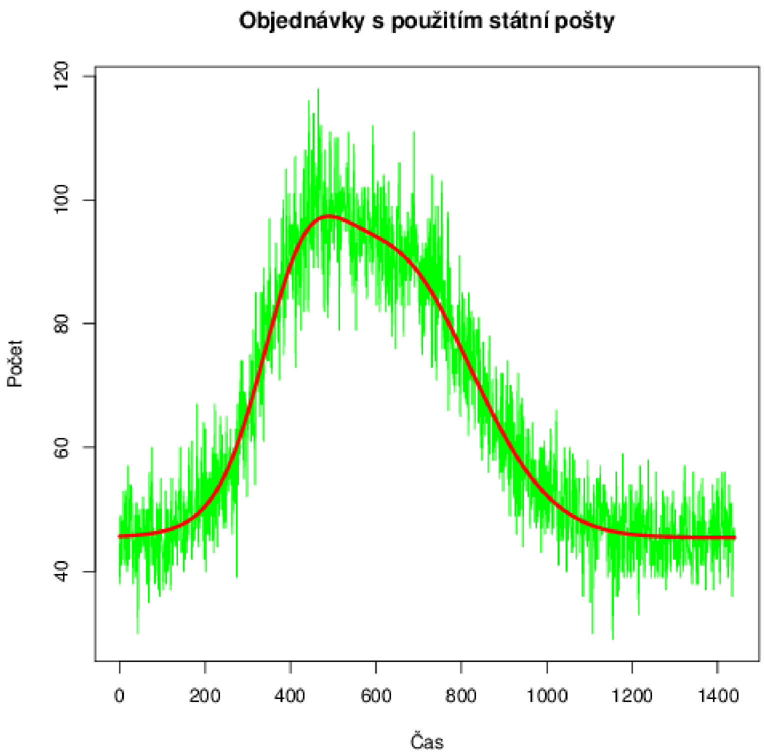
```
        courierList[i] = 0
        postList[i] = 0

        for (y in 0:lambdaT)
        {
                r = runif(1, min=0, max=1) # random(0, 1)
                # either courier or post
                if (r <= courierProb)
                {
                        courierList[i] = as.numeric(courierList[i]) + 1
                }
                else
                {
                        postList[i] = as.numeric(postList[i]) + 1
                }
        }
}
```

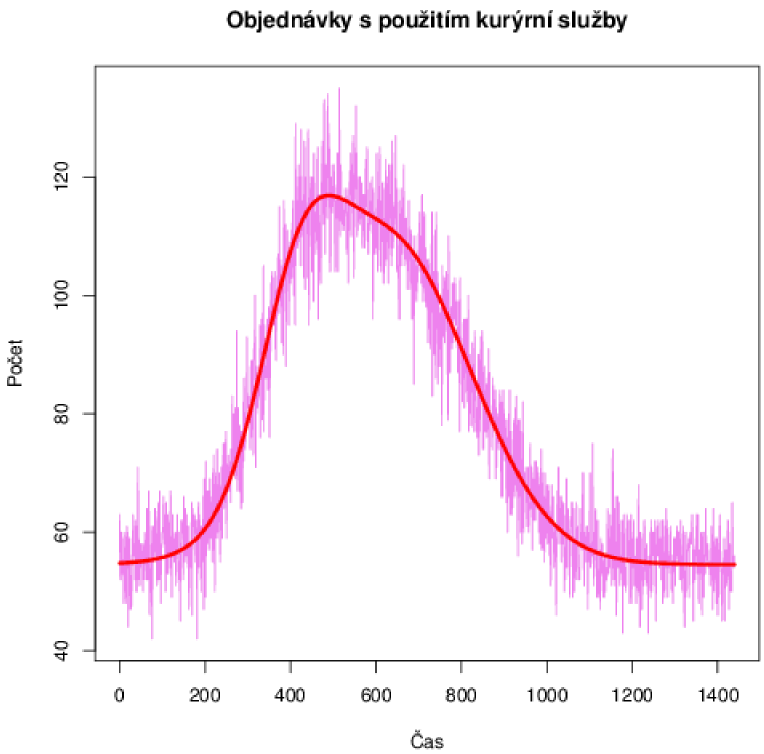
- Projdeme postupně všechny objednávky a každou z nich přiřadíme s pravděpodobností *courierProb* (v našem případě asi 55 %) kurýrovi resp. poště v opačném případě.

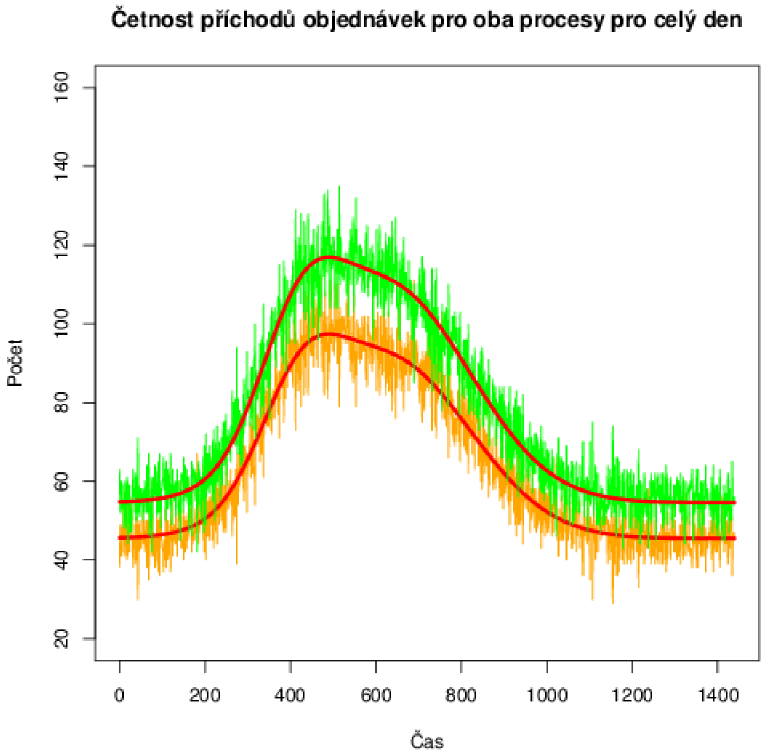
3.III. Zobrazte četnosti příchodů objednávek pro oba procesy pro celý den, t.j. během peridy 24*60 minut. Podobně jako v předchozím bodě, porovnejte četnosti s grafy intenzit lambda_kuryr(t) a lambda_pošta(t).

Příkazy R dle instrukcí:

```
# Post usage; postProb is opposite of courierProb
plot(t, postList, typ="l", col="green", xlab="Čas", ylab="Počet", main="Objednávky s použitím státní pošty")
lines(t, lambda(t)*(1-courierProb), col="red", lwd="3")

# Courier usage
plot(t, courierList, typ="l", col="violet", xlab="Čas", ylab="Počet", main="Objednávky s použitím kurýrní služby")
lines(t, lambda(t)*(courierProb), col="red", lwd="3")

# all together
plot(t, postList, typ="l", col="orange", xlab = "Čas", ylab="Počet", main="Četnost příchodů objednávek pro oba procesy pro celý den", ylim=c(20,160))
lines(t, lambda(t)*(1-courierProb), col="red", lwd="3")
lines(t, courierList, typ = "l", col="green")
lines(t, lambda(t)*(courierProb), col="red", lwd="3")
```


Objednávky s použitím státní pošty
Počet
Čas
0
200
400
600
800
1000
1200
1400
40
60
80
100
120


Objednávky s použitím kurýrní služby
Počet
Čas
0
200
400
600
800
1000
1200
1400
40
60
80
100
120



- Z grafů je vidět, že obě křivky mají **přibližně stejný tvar**, protože obě vychází ze stejného Poissonova rozdělení. Křivka pro kurýry je ale nabývá o trochu vyšších hodnot, protože zákazníci volí dopravu kurýrem s mírně vyšší pravděpodobností (55 %) než dopravu státní poštou.
- **Sečtením** obou křivem bychom dostali stejnou křivku jaku tu, která je uvedena výše (2.III).