Matematična Fizika 1, domača naloga: Preizkus generatorja slučajnih števil porazdeljenih po Poissonovi porazdelitvi

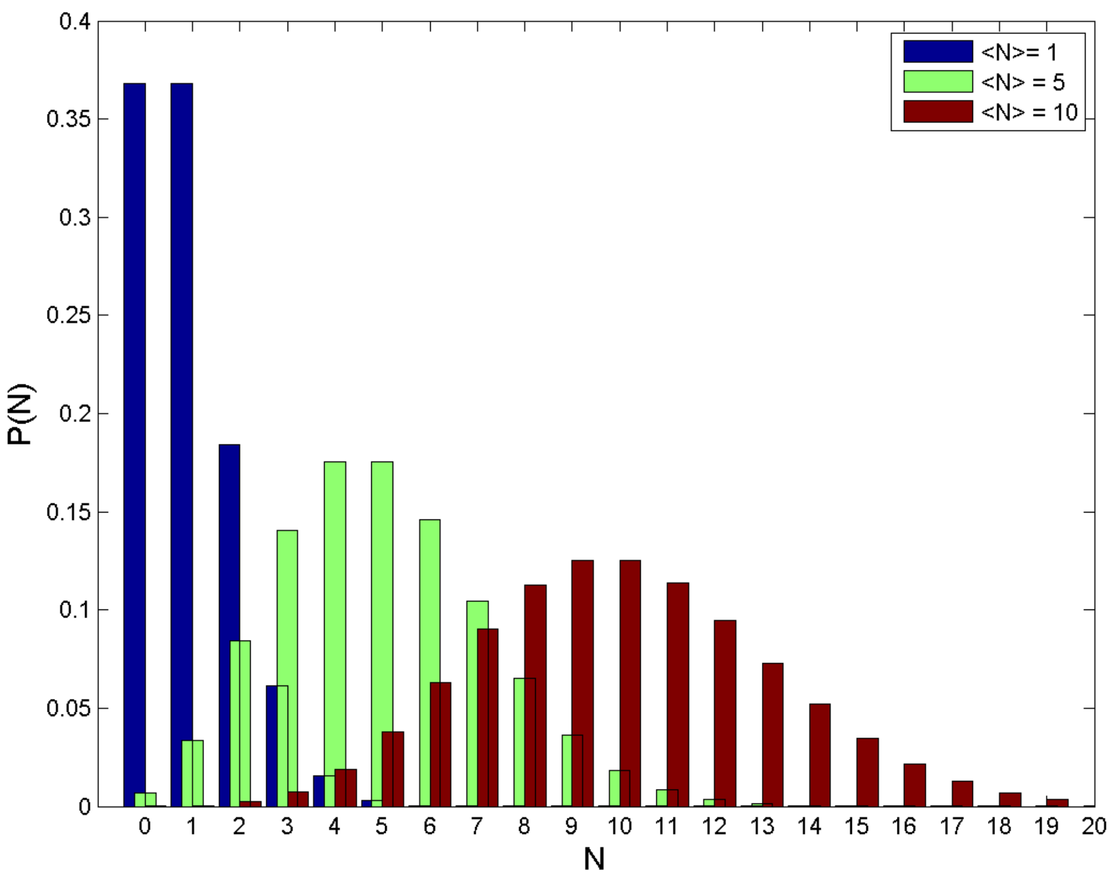
Rok Mihevc

Fakulteta za matematiko in fiziko
Univerza v Ljubljani
rok@mihevc.org

5. oktober 2012

1 Uvod

Za preskus generatorjev slučajnih števil uporabljajo tudi takoimenovani »run-up« test: v danem vzorcu števil preštejemo, koliko je v njem striktno naraščajočih (padajočih) nizov z dolžino M. Izračunamo pričakovane vrednosti števila nizov z dolzino 1, 2, 3, ..., v naključnem vzorcu 1 milijon števil, porazdeljenih po Poissonovi porazdelitvi. (Slika 1)



Slika 1: Poissonova porazdelitev

2 Analitični izračun

2.1 Poissonova porazdelitev

Poissonova porazdelitev je diskretna verjetnostna porazdelitev, ki nam poda *verjetnost danega števila dogodkov N, v danem prostorskem ali časovnem intervalu, pri znani pogostosti teh dogodkov* $\bar{N}$.
Izpeljemo jo iz binomske verjetnostne porazdelitve kot njen limitni primer.

$$P(N) = e^{-\bar{N}} \cdot \frac{\bar{N}^N}{N!}$$

Kumulativno verjetnost pa zapišemo kot:

$$F(N) = e^{-\bar{N}} \cdot \sum_{i=0}^{N} \frac{\bar{N}^i}{i!}$$

Za potrebe te naloge se bomo zaradi preprostosti omejili na primer $\bar{N} = 1$.

2.2 Verjetnost monotono naraščajočega zaporedja

Odločimo se da bom preučili nize števil, ki strogo naraščajo $(N_2 > N_1)$. Ko iz poissonove naključne porazdelitve zajamamemo vrednost N_1, je verjetnost, da smo izbrali N_1 in da bo naslednje poissonovo naključno zajeto število N_2 večje enaka:

$$P(N_2 > N_1) = P(N_1) \sum_{N_2=N_1+1}^{\text{inf}} P(N_2) = P(N_1) \cdot (1 - F(N_1))$$

Zaradi praktičnosti uvedemo notacijo:
$P_+(M) = P(N_1 < N_2 < \cdots < N_{M-1} < N_M)$ - verjetnost za strogo naraščujoč niz dolžine M
$P_-(M) = P(N_1 > N_2 > \cdots > N_{M-1} > N_M)$ - verjetnost za strogo padajoč niz dolžine M
$P_=(M) = P(N_1 = N_2 = \cdots = N_{M-1} = N_M)$ - verjetnost za niz enakih vrednosti N dolžine M
P_+ - verjetnost za verjetnost za strogo naraščujoč niz
$P(M)$ - verjetnost za niz dolžine M
$N_+(M) = N(N_1 < N_2 < \cdots < N_{M-1} < N_M)$ - število strogo naraščujočih nizov dolžine M
$N_-, N_=, P_-, P_=$ - analogno

Prejšen izraz posplošimo na vse možne N_i in vse možne nize števil $N_1 < N_2 < \cdots < N_{M-1} < N_M$ pri dolžini niza M dobimo rahlo nepregledno formulo:

$$P_+(M) = \sum_{N_1=0}^{\text{inf}} \left(P(N_1) \sum_{N_2=N_1+1}^{\text{inf}} \left(P(N_2) \cdots \sum_{N_{M-1}=N_{M-2}+1}^{\text{inf}} (P(N_{M-1})(1 - F(N_{M-1}))\right)\right)$$

Za reševanje te enačbe se zdi primerno orodje rekurzivna funkcija:

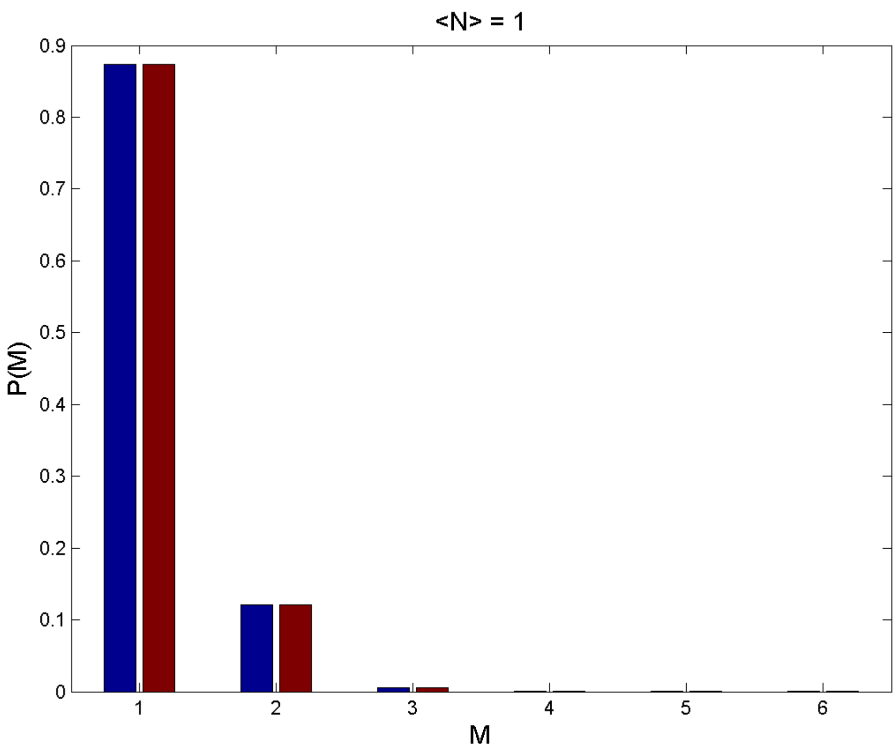
```
%%%%%%%%%%%%%%%%%%%%
% zap
% Rekurzivno izracuna verjetnost da generator
% poissonove nakljucne porazdelitve pri lambda
% vrne strogo narascajoc niz stevil dolzine M.
%
% N = zacetna tocka podniza
% M = dolzina narascujocega niza

function P = zap(Nmin,inf,M,lambda)

  if M == 1
    P = 1 - poisscdf(Nmin-1,lambda);

  else
    P = 0;
    for i=Nmin:inf,
        P = P + poisspdf(i,lambda) * zap(i+1,inf,M-1,lambda);
    end
  end

end
```

Možno pa je tudi reševanje z matriko kombinacij - zapišemo matriko vseh možnih zaporedij, kjer so vrstice možna zaporedja. Nato izračunamo verjetnosti za te elemente in jih zapišemo v matriko enakih dimenzij. Elemente te matrike zmnožimo po vrsticah, da dobimo verjetnost vsakega od možnih zaporedij in jih zapišemo v vektor. Elemente tega vektoja seštejemo in dobimo verjetnost, da bo dolžina strogo naraščujočega zaporedja enaka številu stolpcev naše originalne matrike. Primer kode:

```
zaporedja = combntns(0:4,3)                # dobimo matriko kombinacij, vrednosti od 0:4 na 3 mesta
verjetnosti = poisspdf(poti,1)                 # verjetnosti za posamezne elemente
verjetnosti_zaporedij = prod(verjetnosti',1)'  # verjetnost posameznega zaporedja
verjetnost = sum(verjetnosti_zaporedij)        # skupna verjetnost zaporedij
```

Uporabil sem oba pristopa in (pri $\bar{N} = 1$) izračunal verjetnosti za različne dolžine strogo naraščujočih nizov. Dobljene rezultate sem normiral, tako da $\sum_{m=1}^{\inf} P_+(m) = 1$.

Rezultat obeh pristopov je viden na grafu predvidene verjetnostne porazdelitve $P_+(M)$ (Slika 2):



Slika 2: Verjetnost dolžine niza $P_+(M)$ v odvisnosti od njegove dolžine M, $\bar{N} = 1$

3 Numerični račun

3.1 "Run-up"test

Za preizkus generatorjev naključnih števil uporabimo "run-up"test: v danem vzorcu številu preštejemo striktno naraščujoče nize z dolžino M in jih primerno normaliziramo, da dobimo delež teh nizov P'.

$$P'_+ = \frac{N_+}{N_+ + N_- + N_=}$$

$$P'(M) = \frac{N(M)}{\sum_{m=1}^{\inf} N(m)}$$

$$P'_+(M) = P'(+) \cdot P'(M) = \frac{N_+ N(M)}{(N_+ + N_- + N_=) \cdot \sum_{m=1}^{\inf} N(m)}$$

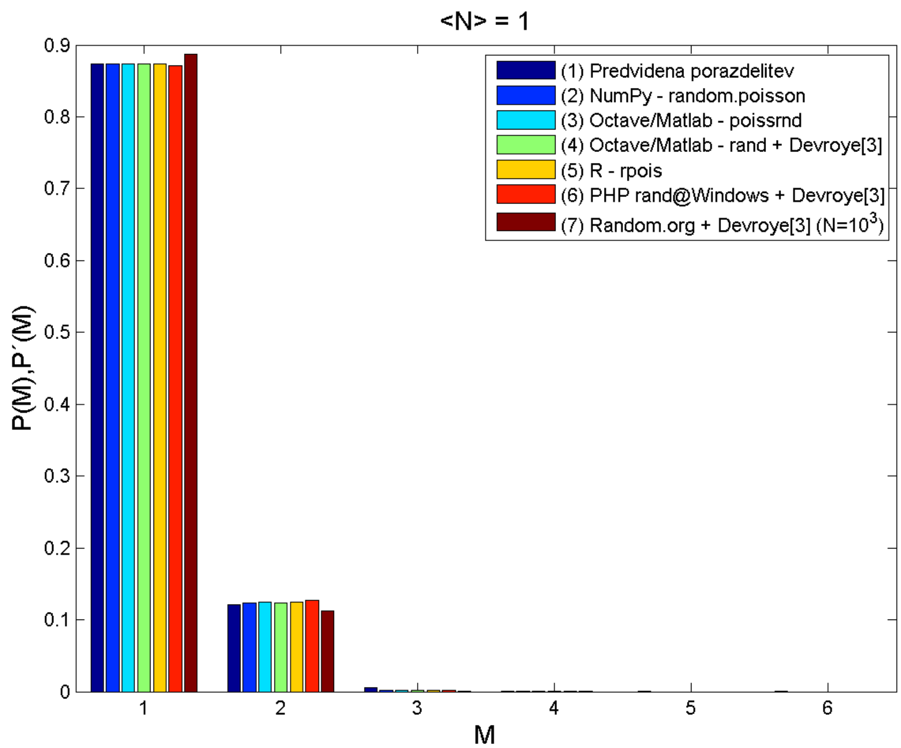
Delež strogo naraščujočih nizov še normaliziramo, da je primerljiv z izračunano verjetnostjo: $\sum_{m=1}^{\inf} P'_+(m) = 1$.

3.2 Generatorji

Za analizo uporabimo sete števil dolžine 10^6 iz različnih virov, porazdeljenih po naključni poissonovi porazdelitvi z $\bar{N} = 1$. Večina generatorjev te porazdelitve uporablja algoritem opisan v [3, str. 504], ki iz uniformne porazdelitve generira poissonovo.
Uporabimo generatorje iz različnih okolij: numpy.random.poisson, Octave/Matlab - poissrnd, Octave/Matlab - rand + Devroye, R - rpois, PHP rand@Windows + Devroye.

S spletne strani random.org je mogoče dobiti tudi prave (tako trdi avtor) naključne podatke, vendar v omejenih količinah - uporabili smo niz 10^3 uniformno porazdeljenih števil iz intervala [0,1).

Rezultate prikažemo na sliki 3 in približane na sliki 4:

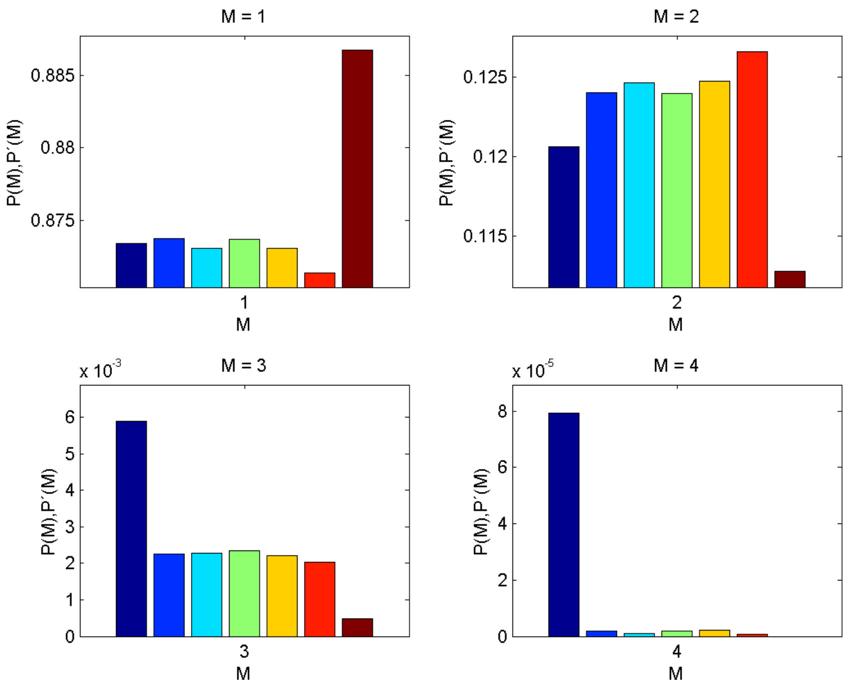


Slika 3: Normaliziran delež naraščujočih nizov ($P'_+(M)$) in verjetnost za dolžino naraščujočih nizov ($P_+(M)$) v odvisnosti od njihove dolžine M. $\bar{N} = 1$

4 Zaključek

Izkaže se da je računanje verjetnosti za daljše strogo naraščujoče nize v poissonovi porazdelitvi računsko zahtevna naloga, vendar imajo pri previdno izbranem $\bar{N}$ kratek rep, kar nalogo precej olajša.

Po grafih sklepamo, da večina priljubljenih računskih paketov generira zelo dobre nize poissonovih naključnih števil, saj so razlike med deleži v generiranih nizih, deleži v nizu naključnih števil in pričakovanimi verjetnostmi dolžin velikostnega reda 10^3.



Slika 4: Normaliziran delež naraščujočih nizov ($P'_+(M)$) in verjetnost za dolžino naraščujočih nizov ($P_+(M)$) v odvisnosti od njihove dolžine M. $\bar{N} = 1$

Literatura

[1] I. Kuščer in A. Kodre: Matematika v fiziki in tehniki, DMFA, Ljubljana, 1994

[2] Dokumentacija programa Matlab - http://www.mathworks.com/help/techdoc/

[3] Luc Devroye: Non-uniform random variate generation, Springer-Verlag New York, 1986 (http://www.eirene.de/Devroye.pdf)