Pravděpodobnost a statistika

Domácí úkoly

Martin Pustka

28.3.2021

Contents

1 1P

Děláno we wordu :(

2 2P

2.1

Za poslední tři roky firma uzavřela 9 projektů se státní dotací. K těmto projektům vyhotovila závěrečné zprávy, přičemž v závěrečné dokumentaci ke 3 projektům jsou závažné chyby. Auditor si z 9 projektů vybral 4 ke kontrole, náhodně. Jestliže X vyjadřuje počet projektů bez chyb v dokumentaci mezi projekty vybranými ke kontrole určete:

2.1.A

Pravděpodobnostní a distribuční funkci této náhodné veličiny X. Pravděpodobnostní funkci zapište do tabulky a distribuční funkci zadejte předpisem.

```
# Pouzijeme Hypergeometrickou vetu, vybirame pocet uspechu
# X ... pocet projektu bez chyb mezi 4 vybranymi
# X ~ H(N = 9, M = 6, n = 4)

x = 0:4 #pocet projektu, ktere nas zajimaji
N = 9 #celkem 9 projektu
M = 6 #celkovy pocet projektu, ktere nas zajimaji
n = 4 #pocet projektu ve vyberu

# graf pravdepodobnostni funkce pro P(X = 0 az 4)
p = dhyper(x, M, N - M, n) # hodnoty pravd. funkce pro x
#kontrola - melo by dat 1
#sum(p) == 1
#plot(x, p)

#jednotlive pravdepodobnosti p priradime k odpovidajicim
#hodnotam vytyhnutemu poctu spravnych projektu
pravdep = rbind(c("P(xi)",p))
colnames(pravdep) = c("xi",0:4)
rownames(pravdep) = ""
pravdep #tabulka pravdepodobnostni funkce
pravd.f(x, p) #vykresleni

#pravdepodobnosti prubezne scitame a priradime k nim omezeni
distrib = cbind(cumsum(p),c("x<=0","0<x<=1","0<x<=1","1<x<=2","x>3"))
colnames(distrib) = c("","")
rownames(distrib) = c("","","F(x) = ","","")
distrib #predpis distribucni funkce
dist.f(x,p) #vykresleni
```

2.1.B

Střední hodnotu, rozptyl, směrodatnou odchylku a modus náhodné veličiny X.

```
# stredni hodnota je soucet vsech hodnot vynasobenych jejich pravdepodobnosti
# rozptyl pomoci E(x^2) - E(x)^2
# smerodatna odchylka je odmocnina rozptylu
# modus je hodnota nejvetsi pravdepodobnosti
# vypocet modus by bylo treba nahradit y[matchAll(max(p),p)]
# misto y[match(max(p),p)], ale chtel jsem to mit spustitelne
# bez knihoven (matchAll je asi v knihovne tuple?)
souhrn(x,p)
```

2.1.C

Jaká je pravděpodobnost, že mezi projekty vybranými ke kontrole bude alespoň jeden projekt uzavřený s chybami v závěrečné dokumentaci?

```
# vybirame alespon jeden spatny projekt,
# takze 1 nebo 2 nebo 3 spravne (4 spatne neexistuji)
# P(x<=3)
sum(p[1:4])
```

2.1.D

Je stanoveno, že pokud auditor shledá v projektu závažná pochybení, firma dostane za každý projekt uzavřený s chybami pokutu 200 000 Kč. Určete očekávanou výši pokuty, kterou by měla firma zaplatit.

```
# celkem 4 projekty a odecteme sanci na vytazeni spravnych projektu
(4 - sum(x*p)) * 200000
```

2.1.E

S jakou pravděpodobností bude pokuta větší než 250 000 Kč?

```
# vetsi nez 250 tisic bude v pripade, ze vybereme alespon dva spatne projekty
# P(x<3)
sum(p[1:3])
```

2.2

Náhodná veličina Y je dána distribuční funkcí

2.2.A

Načrtněte graf distribuční i pravděpodobnostní funkce náhodné veličiny Y.

```
# definujeme distribucni funkci
F = c(0, 0.3, 0.5, 0.9, 1)
y = c(1,10,100,1000)

# z distribucni funkce zjistime pravdepodobnosti pro omezeni
p = diff(F)
pravd.f(y, p) #vykreslime pravdepodobnosti funkci
dist.f(y, p) #vykreslime distribucni funkci
```

2.2.B

Pomocí distribuční funkce vyjádřete a vypočtěte pravděpodobnost, že Y=100, a že hodnota Y je alespoň 10.

```
# P(Y=100)
p[match(100,y)] #dohledame hodnotu pro 100
# P(Y >= 10)
#hledame vetsi nez, takze musime odecist pravdepodobnost 10
1-F[match(10,y)]
# nebo by jsme mohli secist zmeny vetsich nez
p[match(10,y)] + p[match(100,y)] + p[match(1000,y)]
```

2.2.C

Určete střední hodnotu, rozptyl, směrodatnou odchylku a modus náhodné veličiny Y.

```
souhrn(y, p) # podobne jako 1B
```

2.2.D

Jestliže pro náhodnou veličinu R platí R=log(Y), určete střední hodnotu a rozptyl náhodné veličiny R.

```
R = log10(y) # prepocteme si omezeni podle logaritmu

# mohlo by dojit k nesetrizenym omezenim, takze je treba setridit
idx_sorted = order(R)
R = R[idx_sorted]
p_R = p[idx_sorted]

souhrn(R, p_R)
```

2.3

Rozdělení pravděpodobnosti spojité náhodné veličiny (SNV) X je dáno hustotou

2.3.A

Určete hodnotu konstanty c. Načrtněte graf hustoty f(x).

```
# integral od -2/3 do 0 pro f(x)dx = 1 kde f(x) = c(6x+8)
# c(3*x^2+8*x) => c(0*0)-c(3*(-2/3)^2+8*(-2/3)) =>
# => c(0) - c(-4) => c*4 dame rovno jedne => c = 0.25
f = function(x){return(6*x+8)} # f(x) = 6x+8
dolniMez = -2/3
horniMez = 0
1/integrate(f, dolniMez, horniMez)$value

# na papire si vyberu dva krajni body -2/3 a 0 a vypocitam jejich hodnoty
# jelikoz jde o primku tak tyto dva body bych spojil a uzavreny interval
# zbytek grafu hustoty bude na nule do nekonecen s otevrenymi intervaly

f.dens = function(x){
    res = 6/4*x+2 # 1/4(6*x+8)
    res[x < -2/3] = 0
    res[x > 0] = 0
    return(res)
}

x = seq(from = -1, to = 0.5, by = 0.01)
FX = f.dens(x)
plot(x, FX, cex = 0.2, main="Graf hustoty")
```

2.3.B

Určete distribuční funkci F(x).

```
# integral od -2/3 do 0 pro F(t) = f(x)dx
#        0                        t < -2/3
# F(t)   6/8*t^2 + 2*t + 1        t <-2/3;0>
#        1                        t > 0
# pocitam integral od -2/3 do t, kde t je mensi nez 0, pro f(x)dx
# 6/4*x+2
# 6/8*x^2+2*x
# 6/8*t^2+2*t - (6/8*(-2/3)^2+2*(-2/3)) => 6/8*t^2 + 2*t + 1 => dosadim do F(t)

F.dist = function(x){
    res = 6/8*x^2 + 2*x + 1
    res[x < -2/3] = 0
    res[x > 0] = 1
    return(res)
```

```
}

x = seq(from = -1, to = 0.5, by = 0.01)
FX = F.dist(x)
plot(x, FX, type = 'l', main="Distribucni funkce")
```

2.3.C

Vypočtěte pravděpodobnosti P(-1¡=X¡=-1/3) a P(X¿-1/3).

```
# vypocet se provadi integraci hustoty pravdepodobnosti v limitach danych zadanim,
# pripadne oseknutou limitami hustoty integrace provedena v 3B,
# pouze dosazujeme (6/8*(-1/3)^2 + 2*(-1/3) + 1) - (6/8*(-2/3)^2 + 2*(-2/3) + 1)
# P(-1 <= X <= -1/3)
integrate(f.dens, -2/3, -1/3)$value

# P(X > -1/3)
# dosazujeme (6/8*(0)^2 + 2*(0) + 1) - (6/8*(-1/3)^2 + 2*(-1/3) + 1)
integrate(f.dens, -1/3, 0)$value
```

2.3.D

Vypočtěte číselné charakteristiky SNV X. (E(X)=?, D(X)=?, sigma.X=?)

```
x_fx = function(x){
    fx = f.dens(x)
    return(x*fx)
}
xx_fx = function(x){
    fx = f.dens(x)
    return(x*x*fx)
}

# pouzijeme vzorce E(x) = integral x*f(x)dx na celem nenulovem intervalu
# E(x^2) = integral x^2*f(x)dx, pro D(x) = E(x^2) - E(x)^2
E_X = integrate(x_fx, -2/3, 0)$value
E_XX = integrate(xx_fx, -2/3, 0)$value

D_X = E_XX - E_X^2
std_X = sqrt(D_X)

cbind(c("E(X)","D(X)","sigma x"),c(E_X,D_X,std_X))
```

2.3.E

Pro SNV Y platí, že Y=4-2X. Určete střední hodnotu E(Y) a rozptyl D(Y).

```
# pro Y = 4-2x vyuzijeme vzorecku
# E(aX + b) = a*E(x)+b
# D(aX + b) = a^2*D(x)
```

```
# a hodnot, ktere jsme vypocitali v 3D
E_Y = 4 - 2*E_X # E(4-2x)
D_Y = (-2)^2*D_X # D(4-2x)
std_Y = sqrt(D_Y)

cbind(c("E(Y)","D(Y)","sigma y"),c(E_Y,D_Y,std_Y))
```

2.4

Spojitá náhodná veličina X je popsána distribuční funkcí

2.4.A

Určete hustotu pravděpodobnosti f(x).

```
# hustota se pocita derivaci distribucni funkce
# zderivujeme sin x => cos x
# f(x)   cos x <0,pi/2>
#        0 (-inf,0) nebo (pi/2,inf)
f = function(x){
     res = cos(x)        # x^2+2x+1
     res[x < 0] = 0 # 0 pro x<=0
     res[x > pi/2] = 0  # 1 pro x>1
     return(res)
}

x = seq(from = -0.5, to = 2, by = 0.01)
FX = f(x)
plot(x, FX, cex = 0.2, main="Graf hustoty")
```

2.4.B

Určete medián x0,5.

```
# median nalezi do <0,pi/2> a pocita se z distribucni funkce
F.dist = function(x){
     res = sin(x)        # x^2+2x+1
     res[x <= 0] = 0 # 0 pro x<=0
     res[x > pi/2] = 1  # 1 pro x>1
     return(res)
}
```

2.4.C

Najděte w takové, aby pravděpodobnost, že hodnota SNV X bude větší než w, byla 60 %.

```
# podobne jako median spocitame i pravdepodobnost 60%,
# akorat musime pocitat s 1-0.6
x = seq(from = -0.5, to = 2, by = 0.001)
FX = F.dist(x)
plot(x, FX, type='l', main="Pravdepodobnostni funkce a pravdepodobnost 60%")
lines(c(-0.5, 2),c(0.4, 0.4))

asin(0.4)
x[FX >= 0.4][1]
```

3 3P

3.1

V přírodní rezervaci se v průměru nachází 1 stádo divoké zvěře na 25 km^2. Celá rezervace je, co se týče vegetace, zdrojů potravy či vody, vyvážená, tudíž neočekáváme, že by v některých oblastech docházelo k nerovnoměrné kulminaci více stád.

3.1.A

a) Jaká je pravděpodobnost, že ve zkoumané oblasti rezervace o rozloze 100 km^2 bude nejvýše 6 stád?

```
# X... pocet stad na 1km^2
# X - pouzijeme Poissonovu vetu
lambda = 1/25
t = 100
lt = lambda*t # parametr Poissonova rozdeleni

# P(x <= 6)
# V oblasti 100km^2 bude nejvyse 6 stad
ppois(6, lt) # s pravdepodobnosti 88,93 %
```

3.1.B

b) Jaká je pravděpodobnost, že ve zkoumané oblasti rezervace o rozloze 100 km^2 bude alespoň 1 stádo ale nejvýše 7 stád?

```
# P(1<=X<=7) = P(X <= 7) - P(X <= 0)
# V oblasti 100km^2 bude alespon 1 ale nejvyse 7 stad
ppois(7, lt) - ppois(0, lt) # s pravdepodobnosti 93,06 %
```

3.1.C

c) Jaká je pravděpodobnost, že ve zkoumané oblasti rezervace o rozloze 100 km^2 bude alespoň 5 stád, jestliže už 2 stáda v ní byla nalezena?

```
# V oblasti 100km^2 bude alespon 5 stad, jestlize 2 stada byla nalezena
(1-ppois(4,lt)) / (1-ppois(1,lt)) # 0.4085801
```

3.1.D

d) Redakce přírodovědeckého časopisu chce provést focení stáda. Za tímto účelem vybere 10 oblastí a do každé oblasti umístí sledovací zařízení, přičemž každé zařízení zvládne monitorovat oblast o rozloze 5 km^2. Aby focení bylo úspěšné, musí v alespoň jedné monitorované oblasti zachytit alespoň jedno stádo (přepokládáme, že zařízení jsou spolehlivá, takže pokud se stádo v oblasti vyskytne, pak pořízené fotky jsou použitelné). Jaká je pravděpodobnost, že získají fotky pro blížící se vydání časopisu?

```
lambda = 1/25
t = 50
# P(X >= 1) = P(X > 0) = 1 - P(X <= 0)
# V oblasti 50km^2 bude alespon jedno stado
1 - ppois(0, lambda*t) # s pravdepodobnosti 86,47 %
```

3.2

Učitel opraví v průměru 6 domácích úkolů za hodinu.

3.2.A

a) Jaká je pravděpodobnost, že za osmihodinovou pracovní dobu stihne opravit více než 50 úkolů?

```
# X... pocet opravenych domacich ukolu za 1h
# X - pouzijeme Poissonovu vetu
lambda = 6 # cetnost vyskytu za hodinu
t = 8 # behem 8mi hodin
lt = lambda*t # parametr Poissonova rozdeleni

# P(x > 50) = 1 - P(X <= 49)
# Opravi vice jak 50 ukolu
1 - ppois(50, lt) # s 35,13% pravdepodobnosti
```

3.2.B

b) Jaká je pravděpodobnost, že za osmihodinovou pracovní dobu stihne opravit alespoň 45 ale nejvýše 55 úkolů?

```
# Opravi 45 az 55 ukolu
ppois(55, lt) - ppois(45-1, lt) # s 54,67% sanci
```

3.2.C

c) Jaká je pravděpodobnost, že za osmihodinovou pracovní dobu stihne opravit alespoň 55 úkolů, jestliže už 40 úkolů opravil?

```
# Opravi 55 ukolu jestlize 40 opravil
(1 - ppois(54, lt)) / (1 - ppois(39, lt)) # s pravdepodobnosti 19,4 %
```

3.2.D

d) Mezi úkoly je vždy 20 % kompletně špatně vypracovaných úkolů. Učitel opravuje úkoly tak dlouho, dokud postupně nenarazí na 3 takové špatné úkoly, pak si dá čokoládový bonbon na nervy. Určete pravděpodobnost, že učitel bude muset opravit více než 17 úkolů, než si dá bonbon.

```
# X ... pocet pokusu nez vybereme spatny ukol
# X ~ NB(k = 3,p = 0.2)
x = 17
k = 3
p = 0.2
# P(X > 15) = 1 - P(X <= 15)
# Narazi na tri spatne ukoly nejdrive za 17 pokusu
1-pnbinom(x - k, k, p) # s pravdepodobnosti 30,96 %
```

3.3

Doba, po kterou server funguje bez problémů, aniž by potřeboval restartovat, má exponenciální rozdělení. Průměrná doba do nutného restartu je 1 rok.

3.3.A

a) Načrtněte hustotu pravděpodobnosti uvedené náhodné veličiny a její distribuční funkci.

```
# X ... doba do restartu ve dnech
## X ~ Exp(lambda), kde E(X)=1/lambda
lambda = 1/365

x = seq(from = 0, to = 2500, by = 1)
f_x = dexp(x-1, lambda)
plot(x, f_x, cex = 0.1, main="Hustota pravdepodobnosti po dnech")
grid()

F_x = pexp(x-1, lambda)
plot(x, F_x, type='s', main="Distribucni funkce po dnech")
grid()
```

3.3.B

b) Jaká je pravděpodobnost, že server bude potřebovat restartovat nejdříve za 13 měsíců? Výsledek zaznačte do náčrtku hustoty pravděpodobnosti z bodu a).

```
# Restart bude potreba nejdrive za 13 mesicu
1 - pexp(365/12*13, lambda) # s pravdepodobnosti 33,85%
x = seq(from = 0, to = 2500, by = 1)
f_x = dexp(x-1, lambda)
plot(x, f_x, cex = 0.1, main="Hustota pravdepodobnosti a restart po 13 mesicich")
lines(c(0, 2500),c(dexp(365/12*13, lambda), dexp(365/12*13, lambda)))
lines(c(365/12*13, 365/12*13),c(0, max(f_x)))
grid()
```

3.3.C

c) Server jede už 12 měsíců bez restartu, jaká je pravděpodobnost, že bude potřeba jej restartovat v následujících 14 dnech?

```
# Pravdepodobnost restartu po roce pouzivani v nasledujicich 14 dnech
pexp(365 + 365/52*2, lambda) - pexp(365, lambda) # je 1,38 %.
```

3.3.D

d) Do jaké doby bude nutné provést restart serveru s 90 % pravděpodobností?

```
# P(X<t)=0,9 -> F(t)=0,9 -> t... 90% kvantil
# Server bude treba restartovat s 90 % pravdepodobnosti
qexp(0.9, lambda) # po 840,4 dnech.
```

3.4

Systolický krevní tlak dospělých má normální rozdělení se střední hodnotou 112 mmHg a směrodatnou odchylkou 10 mmHg.

3.4.A

a) Načrtněte hustotu pravděpodobnosti uvedené náhodné veličiny a její distribuční funkci.

```
# Systolicky krevni tlak, normalni rozdeleni se stredni hodnotou 112
# a smerodatnou odchylkou 10
mu = 112
sigma = 10

# vykreslime si Hustotu pravdepodobnosti
x = seq(from = 70, to = 150, by = 0.1)
f_x = dnorm(x, mean=mu, sd=sigma)
plot(x, f_x, cex = 0.01, main="Hustota pravdepodobnosti")
grid()

# vykreslime si Distribucni funkci
F_x = pnorm(x, mean=mu, sd=sigma)
plot(x, F_x, type = 'l', main="Distribucni funkce")
grid()
```

3.4.B

b) Kolik procent dospělých má systolický krevní tlak mimo ideální rozmezí, které je 90 až 120 mmHg? Výsledek zaznačte do náčrtku hustoty pravděpodobnosti z bodu a).

```
# Tlak mimo rozmezi 90 az 120g
1- (pnorm(120, mean=mu, sd=sigma) - pnorm(90, mean=mu, sd=sigma))
# s pravdepodobnosti 22,58 %

# Zaznaceni do grafu
# x = c(seq(from = 50, to = 90, by = 0.1),seq(from = 120, to = 180, by = 0.1))
# Vysrafovana oblast je pod krivkou v levo od hodnoty 90
# a v pravo od hodnoty x=120 a y<0
x = seq(from = 70, to = 150, by = 0.1)
f_x = dnorm(x, mean=mu, sd=sigma)
plot(x, f_x, cex = 0.01, main="Hustota pravdepodobnosti")
lines(c(90,90),c(0,max(f_x)))
lines(c(120,120),c(0,max(f_x)))
grid()
```

3.4.C

c) Kolik procent dospělých má systolický krevní tlak nižší než 105 mmHg? Výsledek zaznačte do náčrtku distribuční funkce z bodu a).

```
# Tlak nizsi nez 105mmHg
pnorm(105, mean=mu, sd=sigma) # s pravdepodobnosti 24,2 %
F_x = pnorm(x, mean=mu, sd=sigma)
plot(x, F_x, type = 'l', main="Distribucni funkce s tlakem")
lines(c(105,105),c(0,max(F_x)))
lines(c(70,150),c(pnorm(105, mean=mu, sd=sigma),pnorm(105, mean=mu, sd=sigma)))
grid()
```

3.4.D

d) Určete hodnotu 82. percentilu uvedené náhodné veličiny. Slovně ji interpretujte.

```
# 82. percentil
qnorm(0.82, mean=mu, sd=sigma) # je 121.15 mmHg
# Vysvetleni percentilu: 82% lidi ma hodnotu systolickeho
# krevniho tlaku do hranice 121.15 mmHg
```