GP modellek III.

Fegyverneki Sándor
Miskolci Egyetem
Alkalmazott Matematikai Intézeti Tanszék
matfs@uni-miskolc.hu

2021. február 22.

Full Screen

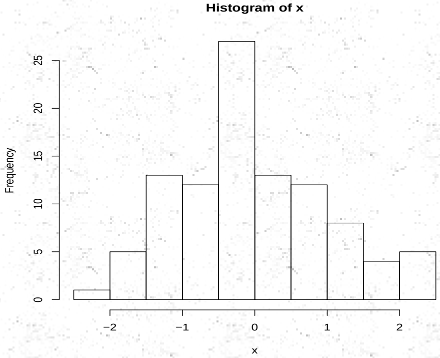
Search

Close

1 Bevezetés

```
> x=rnorm(100)
> hist(x)
```

Figure 1: A hisztogram



2 ELOSZLÁSOK

Γ–ELOSZLÁS

$$\Gamma(\alpha+1)=\int_0^\infty t^\alpha e^{-t}dt$$

$$\Gamma(n+1)=n!,\quad \Gamma(1)=1,\quad \Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$$

$$f(x)=\frac{1}{\Gamma(\alpha)}\lambda^\alpha x^{\alpha-1}e^{-\lambda x},\quad \text{ha}\quad x>0.$$

$$E(\xi)=\frac{\alpha}{\lambda},\quad D^2(\xi)=\frac{\alpha}{\lambda^2}$$

Speciális esetek:

1. Exponenciális eloszlás: $\alpha=1$.
2. χ^2_n-eloszlás: $\alpha=\frac{n}{2}$, $\lambda=\frac{1}{2}$

3 INTERVALLUMBECSLÉSEK

Véletlen intervallum – legalább az egyik végpontja valószínűségi változó.

Példa: $1 < \xi < 2 \quad \Longleftrightarrow \quad 2 < 2\xi$ és $\xi < 2$

$$P(1 < \xi < 2) = P(\xi < 2 < 2\xi)$$

Példa: $\xi \sim \chi^2_{16}$. Mennyi a valószínűsége, hogy $26.3 \in [\xi, 3.3\xi]$?

$$P(\xi < 26.3 < 3.3\xi) = P(7.97 < \xi < 26.3) \approx 0.90.$$

Az intervallum hossza: 2.3ξ.

Várható értéke: $E(2.3\xi) = 2.3 \cdot 16 = 36.8$.

FELADAT: az intervallum hossza ugyanolyan valószínűség mellett legyen rövid!!!

– kontrollkártyák (átlag, szórás, medián, mintaterjedelem)

Paraméter választás – feltételek, feladat. Mit akarunk ellenőrizni?

$\theta \in \Theta$ – paramétertér: $\Theta_0 \subset \Theta$ és $P(\theta \in \Theta_0) = 1-\alpha$

$1-\alpha$ megbízhatóságú konfidenciaintervallum (tartomány).

CSEBISEV-EGYENLŐTLENSÉG:

létezik $E(\xi) = m$, $D(\xi) = \sigma$, ekkor

$$P(|\xi - m| \leq k\sigma) \geq 1 - \frac{1}{k^2}.$$

$$E(\bar{\xi}) = m, \qquad D(\bar{\xi}) = \frac{\sigma}{\sqrt{n}}.$$

Minőségellenőrzés – 6σ, 8σ szabály.

Normális (GAUSS) eloszlás

$$\xi_1, \xi_2, \ldots, \xi_n \sim N(m, \sigma^2), \quad \text{függetlenek}$$

– centrális határeloszlás-tétel (mintaelemszám$> 30?$)

FELADAT: $P(x_a < m < x_b) = 1-\alpha$ – a σ_0 szórás ismert.

$$\frac{\bar{\xi} - m}{\sigma_0}\sqrt{n} \sim N(0,1).$$

$$P\left(\bar{\xi} - u_{\frac{\alpha}{2}}\frac{\sigma_0}{\sqrt{n}} \le m \le \bar{\xi} + u_{\frac{\alpha}{2}}\frac{\sigma_0}{\sqrt{n}}\right) = 1 - \alpha,$$

ahol a σ_0 ismert és $\Phi(u_{\frac{\alpha}{2}}) = 1 - \dfrac{\alpha}{2}$.

FELADAT: $P(x_a < m < x_b) = 1 - \alpha$ – a σ szórás nem ismert.

$$\frac{\bar{\xi} - m}{s_n^\star}\sqrt{n} \sim t_{n-1}.$$

$$P\left(\bar{\xi} - t_{\frac{\alpha}{2}}\frac{s_n^\star}{\sqrt{n}} \le m \le \bar{\xi} + t_{\frac{\alpha}{2}}\frac{s_n^\star}{\sqrt{n}}\right) = 1 - \alpha,$$

ahol az $F_n(t_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$.

FELADAT: $P(x_a < \sigma^2 < x_b) = 1 - \alpha.$

$$\frac{ns_n^2}{\sigma^2} \sim \chi^2_{n-1}.$$

$$P\left(\frac{ns_n^2}{b} \le \sigma^2 \le \frac{ns_n^2}{a}\right) = 1 - \alpha,$$

ahol a $\chi^2_{n-1}(a) = \frac{\alpha}{2}$ és $\chi^2_{n-1}(b) = 1 - \frac{\alpha}{2}.$

Full Screen

Search

Close

Exponenciális eloszlás

$$\xi_1, \xi_2, \ldots, \xi_n \sim EXPO(\lambda), \quad \text{függetlenek}$$

FELADAT: $P(x_a < \lambda < x_b) = 1 - \alpha$

$$n\bar{\xi}\lambda \sim \Gamma_{n,1}, \quad P\left(\frac{a}{n\bar{\xi}} \leq \lambda \leq \frac{b}{n\bar{\xi}}\right) = 1 - \alpha,$$

ahol a $\Gamma_{n,1}(a) = \dfrac{\alpha}{2}$ és $\Gamma_{n,1}(b) = 1 - \dfrac{\alpha}{2}$.

ÁLTALÁNOS FORMULA

$$\xi_1, \xi_2, \ldots, \xi_n \sim F_\theta, \quad \text{függetlenek}$$

$$-\sum_{i=1}^{n} \ln F_\theta(\xi_i) \sim \Gamma_{n,1}.$$

4 Illeszkedésvizsgálat

Legyen az $A_1, A_2, \dots, A_k$ teljes eseményrendszer. Végezzünk el n Bernoulli-kísérletet a megfigyelésükre és jelölje Y_i az A_i gyakoriságát, ekkor az

$$(Y_1, Y_2, \dots, Y_k)$$

vektor polinomiális eloszlású. Írjuk fel a következő hipotéziseket.

$$\begin{aligned} H_0 &: \quad p_i = p_i(\vartheta), \quad \text{ha} \quad \vartheta \in \Theta_0, \\ H_1 &: \quad p_i \quad \text{tetszőleges.} \end{aligned} \tag{1}$$

Ekkor

(2) $$2\ln L(H_0,H_1)=2\sum_{i=1}^{k}Y_i\ln\hat p_i-2\sum_{i=1}^{k}Y_i\ln p_i(\hat\vartheta)=$$

$$=2\sum_{i=1}^{k}Y_i\ln\left(\frac{\hat p_i}{p_i(\hat\vartheta)}\right),$$

ahol $\hat p_i=\dfrac{Y_i}{n}$ és $\hat\vartheta$ a ϑ maximum likelihood becslése a H_0 teljesülése esetén.

Vezessük be a következő jelöléseket:

(3) $$o_i=Y_i,\quad e_i=np_i(\hat\vartheta),\quad \delta_i=o_i-e_i.$$

Ekkor

$$
\begin{aligned}
2\ln L(H_0,H_1) =& 2\sum_{i=1}^{k} Y_i \ln\left(\frac{\hat{p}_i}{p_i(\hat{\vartheta})}\right)\\
=& 2\sum_{i=1}^{k} o_i \ln\left(\frac{o_i}{e_i}\right)\\
=& 2\sum_{i=1}^{k} (\delta_i+e_i)\ln\left(1+\frac{\delta_i}{e_i}\right)\\
=& 2\sum_{i=1}^{k} (\delta_i+e_i)\sum_{m=1}^{\infty}(-1)^{m+1}\frac{\delta_i^m}{me_i^m}\\
\approx& \sum_{i=1}^{k}\frac{\delta_i^2}{e_i}=\sum_{i=1}^{k}\frac{(o_i-e_i)^2}{e_i} \qquad (4)
\end{aligned}
$$

Full Screen
Search
Close

Ez utóbbit szokás Pearson-féle χ^2 statisztikának nevezni, mert ha H_0 esetén $\vartheta \in \mathbf{R}^m$ és becsüljük, akkor

(5) $$2 \ln L(H_0, H_1) \sim \chi^2_{k-m-1}$$

aszimptotikusan.

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$

Ne felejtsük el k az intervallumok száma és m a becsült paraméterek száma.

Példa: Egy dobókocka dobálása során a következő gyakoriságokat kaptuk:

1 – 7db	2 – 6db	3 – 10db
4 – 6db	5 – 8db	6 – 3 db

Ekkor

$$\chi^2 = \sum_{i=1}^{6} \frac{(o_i - 40 \cdot \frac{1}{6})^2}{40 \cdot \frac{1}{6}}$$

(6) $\overline{X} = 3.275, \qquad \chi^2 = 4.1 < 11.071 \approx \chi^2_{5,0.05}.$

```
> qchisq(0.95,5)
[1] 11.0705
```

GP modellek
Bevezetés
ELOSZLÁSOK
INTERVALLUMBEC...
Illeszkedésvizsgálat
A Poisson-eloszlás s...
Vége

Page 17 of 34

Full Screen
Search
Close

A Pearson-féle χ^2-próba alkalmazható tetszőleges eloszlás vizsgálatára, azaz illeszkedésvizsgálatra. Adott az F eloszlásfüggvény. Osszuk fel a $(-\infty, +\infty)$ intervallumot. Legyen

$$-\infty = x_0 < x_1 < \cdots < x_k = +\infty, \tag{7}$$

$$p_i = F(x_{i+1}) - F(x_i), \quad (i = 1, 2, \ldots, k). \tag{8}$$

Megjegyzés:
A felosztás módjára nincs általános szabály. A szokásos alkalmazások: illeszkedésvizsgálat, függetlenség- és homogenitásvizsgálat.

Megjegyzés:

A hisztogram az alapstatisztikák közé tartozik, de csak most jutottunk el odáig, hogy a Pearson-féle χ^2-próba kívánalmai szerint készítsük el.

Az $[a, b]$ intervallum tartalmazza az adatokat.

$$a = d_0 < d_1 < \cdots < d_k = b. \tag{9}$$

A felosztáskor figyeljük a darabszámot, kiugró értékeket és általában legyenek egyenlő hosszúak az intervallumok (kivéve a széleken). Adjuk meg a $[d_{i-1}, d_i)$ intervallumba eső adatok számát (o_i) minden i-re. Az o_i gyakorisággal arányos oszlopot rajzolunk a $[d_{i-1}, d_i)$ intervallumra.

Gyakoriságisztogram:

$$\sum_{i=1}^{k} \frac{o_i}{d_i - d_{i-1}}(d_i - d_{i-1}) = n. \tag{10}$$

Sűrűséghisztogram:

$$\sum_{i=1}^{k} \frac{\frac{o_i}{n}}{d_i - d_{i-1}}(d_i - d_{i-1}) = 1. \tag{11}$$

Full Screen

Search

Close

```
> set.seed(135);z=rnorm(1000)
> cat("Atlag=",mean(z),"\n")
Atlag= -0.02837767
> cat("sn2=",(length(z)-1)/
length(z)*var(z),"\n")
sn2= 1.0345
> cat("Konf.szorasnegyzet",
(length(z)-1)*var(z)/
qchisq(0.975,length(z)-1),", ",
(length(z)-1)*var(z)/qchisq(0.025,
length(z)-1),"\n")
Konf.szorasnegyzet
0.9504016 ,  1.132704
> a=c(-6,-1.5,-0.5,0.5,1.5,6)
> p=vector()
> for (i in 1:(length(a)-1))
p[i]=pnorm(a[i+1])-pnorm(a[i])
> np=length(z)*p
```

```
> z.cut=cut((z-mean(z))/sd(z),
breaks=a)
> nu=vector()
> for (i in 1:(length(a)-1))
nu[i]=table(z.cut)[[i]]
> chi2.z=sum((nu-np)^2/np)
> cat("p=",1-pchisq(chi2.z,length(a)-4),"\n")
p= 0.19279
```


Full Screen

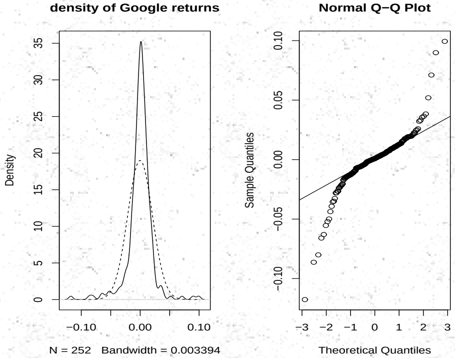
Search

Close

```
set.seed(135);ww=rnorm(10000)
> k=5
> q=seq(0,1,length.out = k+1)
> a=vector()
> for (i in 1:(k+1)) a[i]=qnorm(q[i])
> p=vector()
> for (i in 1:k) p[i]=1/k
> np=length(ww)*p
> ww.cut=cut(((ww-mean(ww))/sd(ww)),
breaks=a)
> nu=vector()
> for (i in 1:k) nu[i]=table(ww.cut)[[i]]
> chi2.ww=sum((nu-np)^2/np)
> cat("p=",1-pchisq(chi2.ww,k-3),"\n")
p= 0.2420769
```

```
price=read.csv("c:/A_GP_munka/GOOG.csv")
#GOOG has 253 values
x <- diff(log(price[,6]))
par(mfrow=c(1,2))
plot(density(x),
main="density of Google returns")
z <- seq(min(x),max(x),length=201)
y <- dnorm(z,mean=mean(x),sd=sd(x))
lines(z,y,lty=2)
qqnorm(x)
qqline(x)
```

Figure 2: Google



5 A Poisson-eloszlás származtatása

Tekintsük a következő tulajdonságokkal rendelkező időben lejátszódó folyamatot: legyenek a bekövetkezések egymástól függetlenek, hosszú intervalumon az intenzitás egyenletes (a bekövetkezések várható száma λh, h az intervallum hossza), míg egységnyi intervallumon a várható számuk $\lambda > 0$. Továbbá, nagyon kicsi δt intervallumon legfeljebb majdnem csak egyszer következik be.

Az $\mathrm{o}(g(x))$ (kis ordó) jelölést alkalmazva folyamatunk matematikai modellje a következő:

$\mathrm{o}(g(x)) = h(x)$, ha $x \to a$, azt jelenti, hogy

$$\lim_{x \to a} \frac{h(x)}{g(x)} = 0.$$

Legyen $p_k(t) = P($ pontosan k darab bekövetkezés t hosszúságú intervallumon). Ekkor

$$p_k(t+\delta t) = p_k(t)p_0(\delta t)+p_{k-1}(t)p_1(\delta t)+\mathsf{o}(\delta t), \quad \text{ha } k \geq 1,$$

$$p_0(t+\delta t) = p_0(t)p_0(\delta t)+\mathsf{o}(\delta t).$$

A kezdeti feltételek pedig

$$p_0(0) = 1, \qquad p_k(0) = 0, \quad \text{ha } k \geq 1.$$

Továbbá, felírva nagyon kicsi intervallumra a folyamat tulajdonságait, a bekövetkezések várható száma alapján, a következő összefüggéseket kapjuk

$$p_1(\delta t) = \lambda \delta t + \mathsf{o}(\delta t),$$

$$p_0(\delta t) = 1 - \lambda \delta t + \mathsf{o}(\delta t).$$

Ezeket behelyettesítve

$$p_k(t+\delta t) = p_k(t)(1-\lambda\delta t)+p_{k-1}(t)\lambda\delta t+\mathsf{o}(\delta t), \quad \text{ha } k \geq 1,$$

Full Screen
Search
Close

$$p_0(t+\delta t) = p_0(t)(1-\lambda\delta t) + \mathsf{o}(\delta t).$$

Rendezzük át az egyenleteket:

$$\frac{p_k(t+\delta t) - p_k(t)}{\delta t} + \lambda p_k(t) = \lambda p_{k-1}(t) + \mathsf{o}(1), \quad \text{ha } k \geq 1,$$

$$\frac{p_0(t+\delta t) - p_0(t)}{\delta t} + \lambda p_0(t) = \mathsf{o}(1).$$

Ha $\delta t \to 0$, akkor

$$p_k'(t) + \lambda p_k(t) = \lambda p_{k-1}(t), \quad \text{ha } k \geq 1,$$

$$p_0'(t) + \lambda p_0(t) = 0.$$

A második sorból rögtön adódik, hogy

$$\frac{d}{dt}\left[e^{\lambda t} p_0(t)\right] = 0,$$

$$e^{\lambda t} p_0(t) = p_0(0).$$

A $p_0(0) = 1$ kezdeti feltétel szerint pedig $e^{\lambda t} p_0(t) = 1$.

Hasonlóan a $k \geq 1$ esetekre

$$\frac{d}{dt}\left[e^{\lambda t} p_k(t)\right] = \lambda e^{\lambda t} p_{k-1}(t),$$

$$\left[e^{\lambda t} p_k(t)\right]_0^x = \int_0^x \lambda e^{\lambda t} p_{k-1}(t) dt.$$

A $p_k(0) = 0$ $(k \geq 1)$ kezdeti feltétel alapján

$$e^{\lambda x} p_k(x) = \int_0^x \lambda e^{\lambda t} p_{k-1}(t) dt.$$

Vezessük be a $Q_k(x) = e^{\lambda x} p_k(x)$ jelölést, ekkor

$$Q_k(x) = \lambda \int_0^x Q_{k-1}(t)dt, \quad \text{ha } k \geq 1,$$

$$Q_0(x) = 1.$$

Ebből rögtön adódik, hogy

$$Q_1(x) = \lambda \int_0^x dt = \lambda x, \quad \text{azaz} \quad p_1(x) = \lambda x e^{-\lambda x};$$

$$Q_2(x) = \lambda \int_0^x \lambda t dt = \frac{\lambda^2 x^2}{2}, \quad \text{azaz} \quad p_2(x) = \frac{\lambda^2 x^2}{2} e^{-\lambda x};$$

$$Q_3(x) = \lambda \int_0^x \frac{\lambda^2 t^2}{2} dt = \frac{\lambda^3 x^3}{6}, \quad \text{azaz} \quad p_3(x) = \frac{\lambda^3 x^3}{6} e^{-\lambda x}; \ldots$$

Általában pedig

$$Q_k(x) = \frac{\lambda^k x^k}{k!}, \quad \text{azaz} \quad p_k(x) = \frac{\lambda^k x^k}{k!} e^{-\lambda x}.$$

Végül egy rögzített t hosszúságú intervallumon, mivel a bekövetkezések várható száma $\mu = \lambda t$, megkapjuk a Poisson-eloszlást.

$$P(\xi = k) = p_k(t) = \frac{\lambda^k t^k}{k!} e^{-\lambda t} = \frac{\mu^k}{k!} e^{-\mu}.$$

Köszönöm

a figyelmet!

References

[1] Deák I.: *Véletlenszámgenerátorok és alkalmazásaik*, Akadémiai Kiadó, Budapest, 1986.

[FR11] Fegyverneki Sándor, Raisz Péter: *Sztochasztikus modellezés*,elektronikus jegyzet, 2011, TÁMOP 4.1.2-08/1/A-2009-0001 project, https://www.uni-miskolc.hu/~ matfs/

[FS11] Fegyverneki Sándor: *Valószínűség-számítás és matematikai statisztika*, elektronikus jegyzet, Kempelen Farkas elktronikus könyvtár, 2011, TÁMOP 4.1.2-08/1/A-2009-0001 project, https://www.uni-miskolc.hu/~ matfs/

[FE78] W. Feller: *Bevezetés a valószínűségszámításba és alkalmazásaiba*, Műszaki Könyvkiadó, Budapest, 1978.

Full Screen

Search

Close

[2] I.M. Szobol: *A Monte-Carlo módszerek alapjai*, Műszaki Könyvkiadó, Budapest, 1981.