Set 10

1. We have to find estimator of parameter p in geometric distribution using MLE method.

$X_1, X_2, X_3, \dots, X_n$ are random samples from geometric distribution, so our likelihood function $L(p) = (1-p)^{(X_1-1)} p (1-p)^{(X_2-1)} p \dots$

$\dots (1-p)^{(X_n-1)} p = (1-p)^{\sum_{i=1}^{n} X_i - n} \cdot p^n$

$\ln(L(p)) = \ln\left((1-p)^{\sum_{i=1}^{n} X_i - n} \cdot p^n\right) = \ln (1-p)^{\sum_{i=1}^{n} X_i - n} + \ln p^n =$

$= \left(\sum_{i=1}^{n} X_i - n\right) \ln(1-p) + n \ln(p)$

Let's find maximum of this function using derivative:

$$\frac{d[\ln L(p)]}{dp} = -\frac{\sum_{i=1}^{n} X_i - n}{1-p} + \frac{n}{p} = 0$$

$$\frac{n}{p} = \frac{\sum_{i=1}^{n} X_i - n}{1-p}$$

$$n - \underline{np} = p \sum_{i=1}^{n} X_i - \underline{pn}$$

$$n = p \sum_{i=1}^{n} X_i \iff p = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{n}{n\bar{X}} = \frac{1}{\bar{X}}$$

Now let's check the second derivative to be sure it's maximum

$$\frac{d^2[\ln L(p)]}{dp^2} = -\frac{\sum_{i=1}^{n} X_i - n}{(1-p)^2} - \frac{n}{p^2} =$$

$= -\left(\frac{n(\bar{X}-1)}{(1-p)^2} + \frac{n}{p^2}\right)$ - which is negative for all p.

So our maximum is $p = \underline{\frac{1}{\bar{X}}}$.

Probability class

Task 2.

List 10

21 maja 2020

The data is the observations $x_1, ..., x_n$ from the following below distributions. Find MLE estimator of listed parameters:
Pareto distribution, $f(x; a, k) = \frac{ka^k}{x^{k+1}}$, $x \in (a, \infty)$, k known, parameter a.

MLE estimator – point in the parameter space that maximizes the likelihood function.

Let's find the likelihood function L relative to parameter a

$$f(x_1, ..., x_n; a, k) = \prod_{i=1}^{n} \frac{ka^k}{x_i^{k+1}} = L(a, k)$$

Now we want this function to reach its maximum at some point a which is the value of our estimator. As increasing function gives the same answer we will look to maximize the function $lnL(a, k)$;

$$ln\, L(a, k) = ln\left(\prod_{i=1}^{n} \frac{ka^k}{x_i^{k+1}}\right) = ln\left(\frac{k^n a^{nk}}{\prod_{i=1}^{k} x_i^{k+1}}\right) = nlnk + nklna \text{ - } (k{+}1)\left(\sum_{i=1}^{k} lnx_i\right)$$

As we can see only the product $nklna$ depends on the value of parameter a so to maximize the value of the whole function we want to maximize the value of $nklna$. As we know that $x \in (a, \infty)$, we can say that the maximum value of a is the minimum value of x_i. Therefore MLE estimator of L is $min(x_i)$.

Problem Set 10

Paweł Głowacki 308023

21.05.2020

1 Task 3

Find MLE of k in Pareto distribution, $f(x;a,k) = \frac{ka^k}{x^{k+1}}$, a known.
Likelihood function:

$$L(k) = \prod_{i=1}^{n} \frac{ka^k}{x_i^{k+1}} = (ka^k)^n \prod_{i=1}^{n} \frac{1}{x_i^{k+1}} = (ka^k)^n (\prod_{i=1}^{n} \frac{1}{x_i})^{k+1}$$

We want maximum value of likelihood function (we can calculate ln(L(k)), ln is monotonically increasing)

Logarithm of $L(k)$:

$$ln(L(k) = n \cdot ln(ka^k) + (k+1) ln(\prod_{i=1}^{n} \frac{1}{x_i}) = n \cdot ln(k) + nk \cdot ln(a) - (k+1) \sum_{i=1}^{n} ln(x_i)$$

Now we calculate derivative ln(L(k)):

$$\frac{\partial ln(L(k))}{\partial k} = \frac{n}{k} + nln(a) - \sum_{i=1}^{n} ln(x_i) = 0$$

$$k = \frac{n}{\sum_{i=1}^{n} ln(x_i) - nln(a)} = \frac{n}{\sum_{i=1}^{n} ln(\frac{x_i}{a})}$$

Aleksander Szymański, 309785, List 10

(5) $f(x,\lambda) = \lambda \cdot e^{-\lambda x}, \; x \in (0,\infty)$

$$L(\lambda) = f(x_1, x_2, \dots, x_n, \lambda) = \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2} \cdot \ldots \cdot \lambda \cdot e^{-\lambda x_n} = \lambda^n \cdot e^{-\lambda \sum_{i=1}^{n} x_i}$$

$$L(\lambda) = \lambda^n \cdot e^{-\lambda \sum_{i=1}^{n} x_i} \quad / \ln$$

$$\ln(L(\lambda)) = \ln\left(\lambda^n \cdot e^{-\lambda \sum_{i=1}^{n} x_i}\right)$$

$$\ln(L(\lambda)) = \ln(\lambda^n) + \ln\left(e^{-\lambda \sum_{i=1}^{n} x_i}\right)$$

$$\ln(L(\lambda)) = n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

$$\frac{d}{d\lambda} \ln(L(\lambda)) = \frac{d}{d\lambda} n \ln(\lambda) - \frac{d}{d\lambda} \lambda \sum_{i=1}^{n} x_i = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = \frac{n}{\lambda} - \bar{x} \cdot n$$

$$\frac{d}{d\lambda} \ln(L(\lambda)) = 0 \Rightarrow \frac{n}{\lambda} - \bar{x} n = 0 \Rightarrow \underbrace{\lambda = \frac{1}{\bar{x}}}$$

$$\frac{d^2}{d\lambda^2} \ln(L(\lambda)) = -\frac{n}{\lambda^2} \leq 0 \Rightarrow \lambda = \frac{1}{\bar{x}} \text{ is maximum of } L \Rightarrow$$

$\Rightarrow \lambda = \frac{1}{\bar{x}}$ is MLE estimator.

Dencity function:

$$f(x;k,\lambda) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} \exp\left(-\left(\frac{x}{\lambda}\right)^{k}\right)$$

Likelihood function:

$$L(\lambda) = f(x_1, x_2, \dots, x_n; k, \lambda) = \frac{k^n}{\lambda^{nk}} \exp\left(-\sum\left(\frac{x_i}{\lambda}\right)^k\right) \prod x_i^{k-1}$$

Using logarithm on both sides:

$$\log L(\lambda) = n \log k - nk \log \lambda - \sum\left(\frac{x_i}{\lambda}\right)^k + (k-1)\sum x_i$$

Differentiating we have:

$$\frac{d \log L(\lambda)}{d\lambda} = -\frac{nk}{\lambda} + k\sum \frac{x_i^k}{\lambda^{k+1}}$$

$$-\frac{nk}{\lambda} + k\sum \frac{x_i^k}{\lambda^{k+1}} = 0$$

$$-\frac{n}{\lambda} + \frac{1}{\lambda^{k+1}}\sum x_i^k = 0$$

$$\sum x_i^k = n \cdot \lambda^k$$

$$\lambda = \left(\frac{\sum x_i^k}{n}\right)^{\frac{1}{k}}$$

Questions 6 and 7

Andrzej Tkaczyk, 309181

May 21, 2020

Exercise 6

$X \sim \chi^2(n), \quad Y \sim \chi^2(k)$

$f_X(x) = \frac{x^{\frac{n}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}$

$f_Y(y) = \frac{y^{\frac{k}{2}-1}e^{-\frac{y}{2}}}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})}$

$f_{(X,Y)}(x,y) = f_X(x) * f_Y(y) = \frac{x^{\frac{n}{2}-1}e^{-\frac{x}{2}}y^{\frac{k}{2}-1}e^{-\frac{y}{2}}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})2^{\frac{k}{2}}\Gamma(\frac{k}{2})} = \frac{x^{\frac{n}{2}-1}y^{\frac{k}{2}-1}e^{-\frac{x+y}{2}}}{2^{\frac{n+k}{2}}\Gamma(\frac{n}{2})\Gamma(\frac{k}{2})}$

Exercise 7

$F = \frac{X}{Y}\frac{k}{n}, \quad T = Y$

$X = \frac{n}{k}FT, \quad Y = T$

$$|J| = \begin{vmatrix} \frac{n}{k}T & \frac{n}{k}F \\ 0 & 1 \end{vmatrix} = \frac{n}{k}T$$

$$g_{(F,T)}(f,t) = f_{(X,Y)}(x(f,t), y(f,t)) * |J| = \frac{(\frac{n}{k}ft)^{\frac{n}{2}-1}t^{\frac{k}{2}-1}e^{-\frac{t(\frac{n}{k}f+1)}{2}}}{2^{\frac{n+k}{2}}\Gamma(\frac{n}{2})\Gamma(\frac{k}{2})} * \frac{n}{k}t = \frac{(\frac{n}{k})^{\frac{n}{2}}f^{\frac{n}{2}-1}}{2^{\frac{n+k}{2}}\Gamma(\frac{n}{2})\Gamma(\frac{k}{2})} * t^{\frac{n+k}{2}-1}e^{-\frac{t(\frac{n}{k}f+1)}{2}}$$

$$g_F(f) = \int_0^\infty g_{(F,T)}(f,t)\ dt = \int_0^\infty \frac{(\frac{n}{k})^{\frac{n}{2}}f^{\frac{n}{2}-1}}{2^{\frac{n+k}{2}}\Gamma(\frac{n}{2})\Gamma(\frac{k}{2})} * t^{\frac{n+k}{2}-1}e^{-\frac{t(\frac{n}{k}f+1)}{2}}\ dt = \frac{(\frac{n}{k})^{\frac{n}{2}}f^{\frac{n}{2}-1}}{2^{\frac{n+k}{2}}\Gamma(\frac{n}{2})\Gamma(\frac{k}{2})} * \int_0^\infty t^{\frac{n+k}{2}-1}e^{-\frac{t(\frac{n}{k}f+1)}{2}}\ dt$$

$$\int_0^\infty t^{\frac{n+k}{2}-1}e^{-\frac{t(\frac{n}{k}f+1)}{2}}\ dt = \begin{vmatrix} s = \frac{t(\frac{n}{k}f+1)}{2} \\ t = \frac{2s}{\frac{n}{k}f+1} \\ dt = \frac{2}{\frac{n}{k}f+1}\ ds \end{vmatrix} = \int_0^\infty (\frac{2s}{\frac{n}{k}f+1})^{\frac{n+k}{2}-1}e^{-s}\frac{2}{\frac{n}{k}f+1}\ ds = \int_0^\infty (\frac{2}{\frac{n}{k}f+1})^{\frac{n+k}{2}}s^{\frac{n+k}{2}-1}e^{-s}\ ds =$$

$$= (\frac{2}{\frac{n}{k}f+1})^{\frac{n+k}{2}} * \Gamma(\frac{n+k}{2})$$

$$g_F(f) = \frac{(\frac{n}{k})^{\frac{n}{2}}f^{\frac{n}{2}-1}}{2^{\frac{n+k}{2}}\Gamma(\frac{n}{2})\Gamma(\frac{k}{2})} * \frac{2^{\frac{n+k}{2}}}{(\frac{n}{k}f+1)^{\frac{n+k}{2}}} * \Gamma(\frac{n+k}{2}) = \frac{\Gamma(\frac{n+k}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{k}{2})} * (\frac{n}{k})^{\frac{n}{2}} * \frac{f^{\frac{n}{2}-1}}{(\frac{n}{k}f+1)^{\frac{n+k}{2}}}$$

List 10 Task 8

21 maja 2020

Points $(x_1, y_1), \cdots, (x_n, y_n)$ are given. We are looking for a regression coefficients in the form of $y = a + bx + cx^2$. Justify that parameters a, b, c are solution of the system of linear equations

$$\begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{bmatrix}$$

Let

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} \qquad \beta = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \qquad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

We know that $\beta = (X^T X)^{-1} X^T Y$ from which we derive $(X^T X)\beta = X^T Y$. Which means that

$$\left(\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \end{bmatrix} \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} \right) \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\begin{bmatrix} \sum 1 & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{bmatrix}$$

TASK 9 SZYMON KILCZAK

POINTS GIVEN: $(x_1, y_1, z_1), \dots, (x_n, y_n, z_n)$

REGRESSION COEFFICIENTS IN FORM OF $z = a + bx + cy$

WE HAVE TO PROVE THAT THE PARAMETERS a, b, c ARE SOLUTION OF:

$$\underbrace{\begin{bmatrix} n & \sum x_i & \sum y_i \\ \sum x_i & \sum x_i^2 & \sum x_i y_i \\ \sum y_i & \sum x_i y_i & \sum y_i^2 \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} a \\ b \\ c \end{bmatrix}}_{B} = \underbrace{\begin{bmatrix} \sum z_i \\ \sum z_i x_i \\ \sum z_i y_i \end{bmatrix}}_{C}$$

WE CAN DO CALCULATIONS OF LEFT SIDE OF EQUATION:

$$\underbrace{\begin{bmatrix} n & \sum x_i & \sum y_i \\ \sum x_i & \sum x_i^2 & \sum x_i y_i \\ \sum y_i & \sum x_i y_i & \sum y_i^2 \end{bmatrix}}_{A} \cdot \underbrace{\begin{bmatrix} a \\ b \\ c \end{bmatrix}}_{B} = \underbrace{\begin{bmatrix} an + b\sum x_i + c\sum y_i \\ a\sum x_i + b\sum x_i^2 + c\sum x_i y_i \\ a\sum y_i + b\sum x_i y_i + c\sum y_i^2 \end{bmatrix}}_{D}$$

$$an + b\sum x_i + c\sum y_i = \sum z_i$$

$$a\sum x_i + b\sum x_i^2 + c\sum x_i y_i = \sum a x_i + \sum b x_i^2 + \sum c x_i y_i = \sum z_i x_i$$

$$a\sum y_i + b\sum x_i y_i + c\sum y_i^2 = \sum a y_i + \sum b x_i y_i + \sum c y_i^2 = \sum z_i y_i$$

SO WE CAN SEE THAT MATRIX D IS EQUAL TO MATRIX C

List 10 Exercise 10

Szymon Kosakowski 309980

21 maja 2020

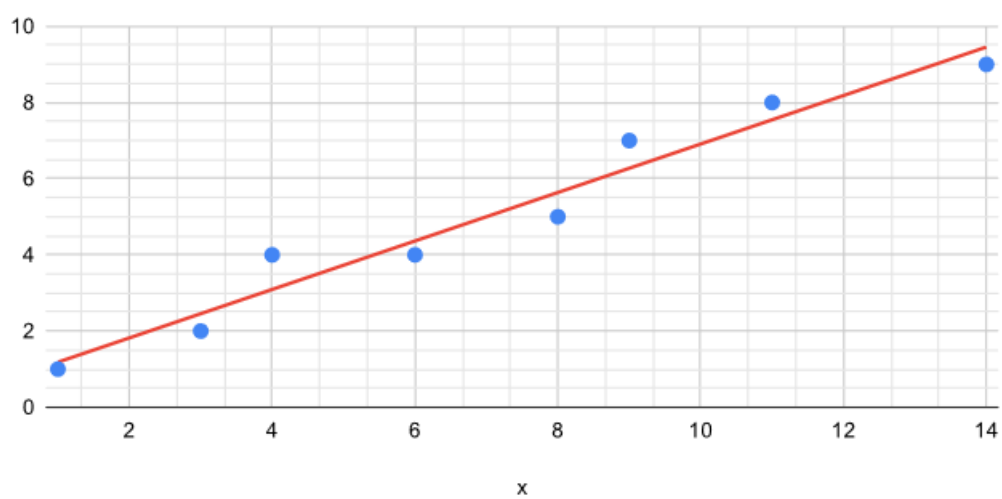
$$\beta = \frac{n * \sum x_k y_k - \sum x_k * \sum y_k}{n * \sum x_k^2 - (\sum x_k)^2}$$

$$\alpha = \bar{y} - \beta * \bar{x}$$

$$\hat{y_k} = \alpha + \beta * x_k$$

	A	B	C	D	E
1					
2	n	xk	yk	xk*yk	xk^2
3	8	1	1	1	1
4		3	2	6	9
5		4	4	16	16
6		6	4	24	36
7		8	5	40	64
8		9	7	63	81
9		11	8	88	121
10		14	9	126	196
11	suma:	56	40	364	524
12					
13					
14		avg:			
15		y	5		
16		x	7		

b =	0,6363636364
a =	0,5454545455



Task 11

$$P \cdot V^k = C$$

$$P = C \cdot V^{-k}$$

$$\ln(P) = -k \cdot \ln(V) + \ln(C)$$

$$y = a \cdot x + b$$

$$a = \frac{COV(\ln(V); \ln(P))}{V(\ln(P))}$$

$$a = \frac{\sum_{i}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i}^{n} (x_i - \bar{x})^2} = -1{,}404204$$

$$k = -a = 1{,}404204$$

$$C = e^{b} = 15971{,}804$$

P для V=100

$$P = \frac{C}{V^k} = 24{,}828246$$

	A	B	C	D	E	F	G	H	I	J
1	V	P		ln(v)	ln(p)		ln(v)-avg(ln(v))	ln(p)-avg(ln(p))	G*H	G^2
2	54.3	61.2		3.994524227	4.11414719		-0.4937271773	-0.374104214771336	0.184705418	0.243766526
3	61.8	49.5		4.123903364	3.90197267		-0.3643480398	-0.586278734714971	0.213609508	0.132749494
4	72.4	37.6		4.282206299	3.62700405		-0.2060451049	-0.861247353893767	0.177455801	0.042454585
5	88.7	28.4		4.485259889	3.346389145		-0.002991515	-1.14186225912246	0.003415898	8.94916E-06
6	118.6	19.2		4.775756487	2.954910279		0.28750508227	-1.53334112525588	-0.44084337	0.082659172
7	194	10.1		5.267858159	2.312535424		0.77960675477	-2.1757159804424	-1.69620287	0.607786692
8										
9				e	2.718281828					
10						Sum			-1.55785962	1.109425419
11										
12										
13				v	avg(ln(v))	4.488251404			-1.55785962	
14					avg(ln(p))	3.376159793		B1		-1.404204
15									1.109425419	
16										
17										
18								B0	9.67858039	
19					(-a)					
20					k	1.404204005				
21										
22					(e^B0)					
23					C	15971.80709				
24										
25						C	15971.8070878			
26					P(V=100)			24.8282463659601		
27						V^k	643.291791631			