<u>אותות אקראיים</u>

<u>עבודה 1</u>

מרצה: ד"ר יצחק לפידות

תאריך: 2.1.2020

מגיש:

יזיד בישארה 315914689

חלק 1

נבצע גנרציה של סדרה D באורך 1024 לשני גאוסיאנים G0, G1 באופן אקראי.

נעשה גנרציה לשבעה סדרות k=0,1...,6

כל סדרה מחלקים ל-m סדרות אורך כל סדרה הוא 2^k

לפי התפלגות ברנולי קובעים מאיזה גאוסיאן.

פרמטרים של הגאוסיאן:

לפי תעודת זהות שלי 315914689

תעודת הזהות השנייה היא משותף בקורס 321194276(בחרתי את המספר השני כדי שיהיה לנו שונות שונה מהגאוסיאן הראשון)

$$\mu_0 = \frac{9}{10} = 0.9 \qquad \mu_1 = \frac{6}{10} = 0.6$$

$$\sigma_0^2 = \frac{3}{2} = 1.5 \qquad \sigma_1^2 = \frac{2}{2} = 1$$

$$G_0 \sim N(\ 0.9, 1.5\ ) \qquad G_1 \sim N(0.6, 1.5)$$

randn(m,n) : עושה גנרציה למטריצה בגודל mxn שהערכים שיש בה מתפלגים נורמלית

$$\forall i \cdot x_i \sim N(0,1)$$

כדי ליצור הפרמטרים הרצויים מכפילים כל איבר בשונות הרצויה ומוסיפים את התוחלת

$$y_i = x_i * \sigma + \mu$$

חלק 2

משערך ML

באמצעות משערך ML משעריכים את הסדרות

$$P\left(\mathcal{D}\middle|\mu,\sigma^2\right)=\prod_{n=1}^{N}\frac{1}{\sqrt{2\pi\sigma}}e^{-\frac{1}{2\sigma^2}(x_n-\mu)^2}=\left(\frac{1}{\sqrt{2\pi\sigma}}\right)^N e^{-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2}$$

$$L\left(\mathcal{D}\middle|\mu,\sigma^2\right)=-N\ln\left(\sqrt{2\pi\sigma}\right)-\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2$$

$$Define:\quad \overline{x}\equiv\frac{1}{N}\sum_{n=1}^{N}x_n\qquad S\equiv\sum_{n=1}^{N}(x_n-\mu)^2$$

$$L\left(\mathcal{D}\middle|\mu,\sigma^2\right)=-N\ln\left(\sqrt{2\pi\sigma}\right)-\frac{1}{2\sigma^2}\left[N\left(\mu-\overline{x}\right)^2+S\right]$$

$$\mu_{ML}=\overline{x}=\frac{1}{N}\sum_{n=1}^{N}x_n$$

$$\sigma_{ML}^2=\frac{\sum_{n=1}^{N}\left(x_n-\mu_{ML}\right)^2}{N}$$

מימוש בקוד:

```
sum_x=0;
mu_ml=0;
for i= 1:N%%mean ML
    sum_x= sum_x +Xn(i);
end
mu_ml=sum_x/N;
S=0;
for i = 1:N %%S
    S = S + (Xn(i)-mu_ml)^2;
end
L0 = -N*log((var0*2*pi)^0.5)- (N*(mu_ml-mu0)^2+S)/(2*var0);
L1 = -N*log((var1*2*pi)^0.5)- (N*(mu_ml-mu1)^2+S)/(2*var1);
```

אחר כך משווים בין שני ה Lים רואים לאיזה מהם יש סיכוי יותר גדול.

בשערוך הזה נמצא את התוחלת והשונות ואחר כך נמצא את הנראות, פה יש לנו הסטה בשונות

$$\boxed{\frac{N-1}{N}\sigma^2}$$

(bias)

בגלל שהנחנו שאנחנו לא יודעים את התוחלת והשונות אז כששערכנו את המספר השתמשנו בתוחלת המשעורכת זה מה שגרם ל סטיה,

אבל יש לנו Asymptotically unbiased estimator זה אומר כשכל ה N (מספר הדגימות) גדול יותר הסטייה תשאוף לאפס.

חישוב KL divergence:

KL divergence מדד כמה שתי ההתפלגיות שומות אחדת לשניה, במצב שלנו השוני בין ההתפלגות המשוערכת להתפלגות האמיתית. ככל שההבדיל בין ההתפלגיות קטן יותר הערך שמוציאה קטן יותר, בהתחשבות בהסתברות כל קטע בההתפלגות והערך.

כלומר כשערך ה KLגדול יותר השוני גדול יותר

אז נשתמש בחישוב הזה כמדד כמה הערך שיצא שלנו מדויק לערך האמיתי

החישוב הינו:

$$\mathrm{KL}_x(f_1, f_2) = \int f_1(x) \log \frac{f_1(x)}{f_2(x)} dx$$

כאשר f1וf2 הינן 2 התפלגויות שנרצה לבדוק את הדמיון ביניהם (התפלגות מקור ומשוערכת).

עבור המקרה הפרטי שלנו, נרצה לבדוק את השוני עבור 2 גאוסיאנים:

$$\int [\log(p(x)) - log(q(x))]\, p(x)dx$$

$$= \int \left[-\tfrac{1}{2}\log(2\pi) - \log(\sigma_1) - \tfrac{1}{2}\left(\tfrac{x-\mu_1}{\sigma_1}\right)^2 + \tfrac{1}{2}\log(2\pi) + \log(\sigma_2) + \tfrac{1}{2}\left(\tfrac{x-\mu_2}{\sigma_2}\right)^2\right]$$

$$\times \tfrac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\tfrac{1}{2}\left(\tfrac{x-\mu_1}{\sigma_1}\right)^2\right] dx$$

$$= \int \left\{\log\left(\tfrac{\sigma_2}{\sigma_1}\right) + \tfrac{1}{2}\left[\left(\tfrac{x-\mu_2}{\sigma_2}\right)^2 - \left(\tfrac{x-\mu_1}{\sigma_1}\right)^2\right]\right\} \times \tfrac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\tfrac{1}{2}\left(\tfrac{x-\mu_1}{\sigma_1}\right)^2\right] dx$$

$$= E_1 \left\{\log\left(\tfrac{\sigma_2}{\sigma_1}\right) + \tfrac{1}{2}\left[\left(\tfrac{x-\mu_2}{\sigma_2}\right)^2 - \left(\tfrac{x-\mu_1}{\sigma_1}\right)^2\right]\right\}$$

$$= \log\left(\tfrac{\sigma_2}{\sigma_1}\right) + \tfrac{1}{2\sigma_2^2} E_1\left\{(X-\mu_2)^2\right\} - \tfrac{1}{2\sigma_1^2} E_1\left\{(X-\mu_1)^2\right\}$$

$$= \log\left(\tfrac{\sigma_2}{\sigma_1}\right) + \tfrac{1}{2\sigma_2^2} E_1\left\{(X-\mu_2)^2\right\} - \tfrac{1}{2}$$

(Now note that
$(X-\mu_2)^2 = (X-\mu_1+\mu_1-\mu_2)^2 = (X-\mu_1)^2 + 2(X-\mu_1)(\mu_1-\mu_2) + (\mu_1-\mu_2)^2$)

$$= \log\left(\tfrac{\sigma_2}{\sigma_1}\right) + \tfrac{1}{2\sigma_2^2}\left[E_1\left\{(X-\mu_1)^2\right\} + 2(\mu_1-\mu_2)E_1\left\{X-\mu_1\right\} + (\mu_1-\mu_2)^2\right] - \tfrac{1}{2}$$

$$= \log\left(\tfrac{\sigma_2}{\sigma_1}\right) + \tfrac{\sigma_1^2 + (\mu_1-\mu_2)^2}{2\sigma_2^2} - \tfrac{1}{2}$$

אז מה אנחנו רוצים:

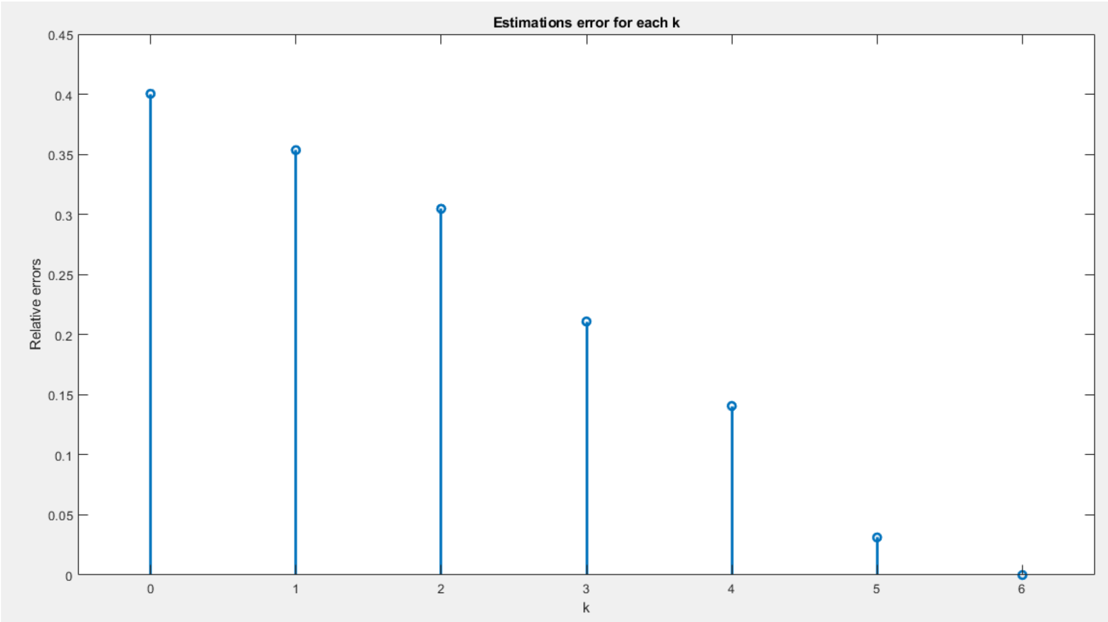
$$KL(p,q) = \log \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2}$$

כאשר $\sigma1=\sigma2$, $\mu1=\mu2$ נקבל 0 כלומר ההתפלגויות זהות.

*ההכוחה נמצא בארתר הזה:
https://stats.stackexchange.com/questions/7440/kl-divergence-between-two-univariate-gaussians.

תוצאות הניסוי

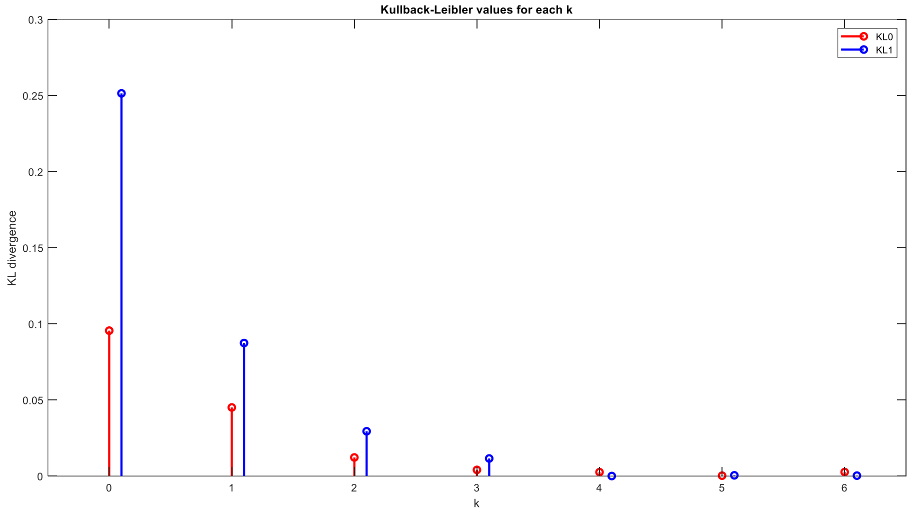
שגיאת ה ML ביחס לגודל כל ניסוי



בכל ניסוי מ-0 ל-6 נספרו כמות השגאות בכל שערוך של הסדרה Dm,

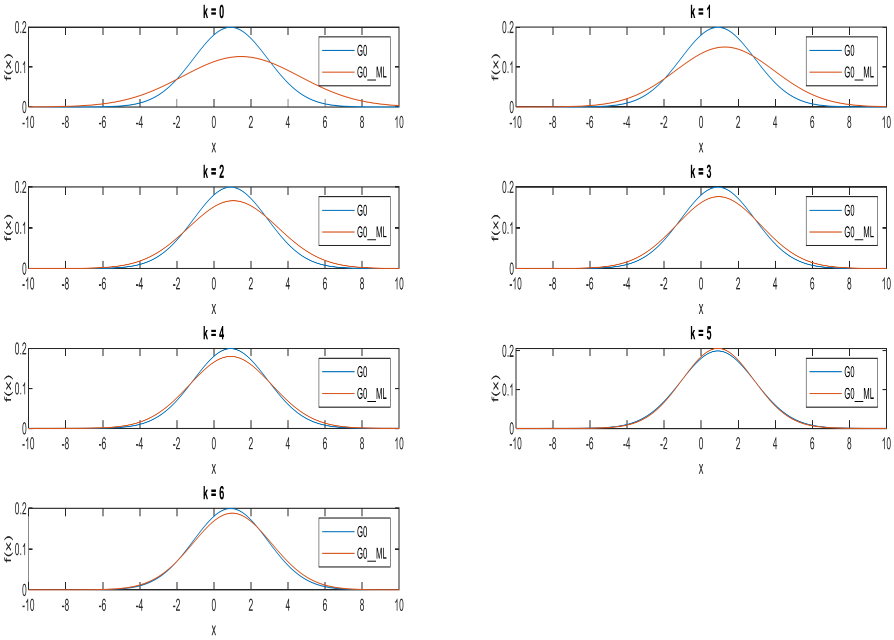
אחרי נירמול כלומר חילוק כמות השגיאות בקמות כל הנסויים יהיה לנו את הגרף למעלה, כפי שרואים אחוז השגיאה הולך וקטן עם הגדלת הסדרה, זה מהסיבה שמשערך ML מתקרב לערך האמיתי עם הגדלת הערכים, גם ההטיה הולכת וקטנת כמו שראינו למעלה.

גרף KL:

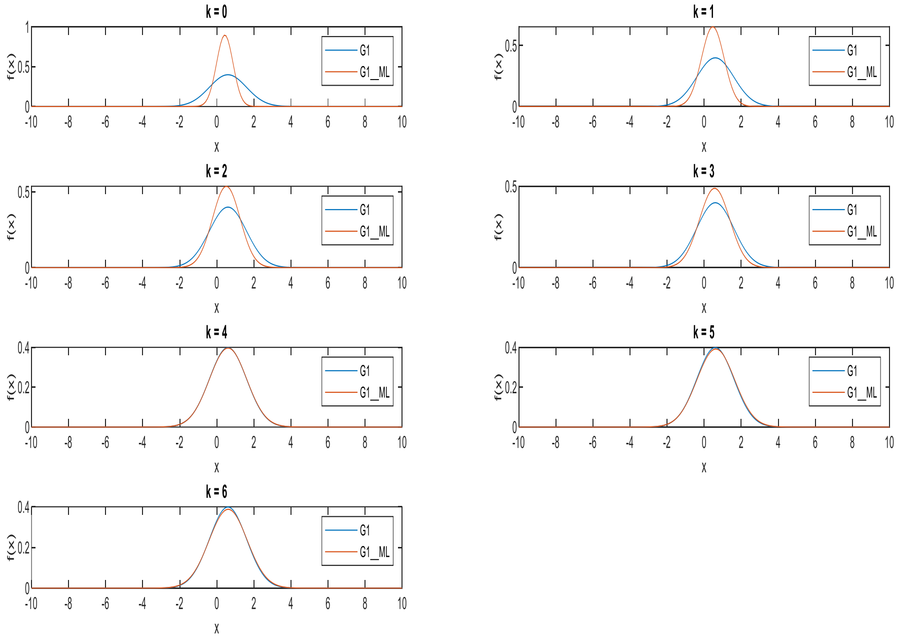


ככל שאורך הסדרה גדל ערך הKL קטן, מזה נובע שהדמיון בין ההסתבריות הולך וגדל ביחס לאורך השגיאה, וזה מכיוון לסדרה גדולה יש יותר מידע והשערוך נהיה יותר מדויק.

G0:



:1G



סיכום ומסקנות

1. שינוי קטן במשוואה נותן תוצאות לא נכונות למשל כמו הוצאת השורש מחוץ לLOG.
 זה יכול להיות בגלל שהחישוב יוצא מדויק יותר כשקודם עושים שורש ואחר כך מכפילים

$$L\left(\mathcal{D}\middle|\mu,\sigma^2\right)=-N\ln\left(\sqrt{2\pi}\sigma\right)-\frac{1}{2\sigma^2}\left[N\left(\mu-\bar{x}\right)^2+S\right]$$

2. השינוי לא באיברי הסדרה (שזה 1024) אלא כמות תתי הסדרה(DM).

3. ככל שמגדילים כמות המידע התוצאה תהיה יותר מדוייקת וזה יושב עם ההגעון של משערך הML.

4. KL divergence נותן ערך מספרי שבאמצעתו ניתן לדעת עד כמה השערוך היה מדויק. כלומר יכול להיות שזה את הנקודות ששייכים לגוסיאן אבל לא בנה גאוסייאן בצורה נכונה, לכן הוא תלוי בכמות המידע ואיך משתמש בהם.

קוד:

```
%% Home Work 1
%% yazid bisharat 315914689
%% MAIN %%
clear all;
close all;
clc;
% parameters
mu0 = 0.9;  % G0 mean & variance
var0 = 2;
mu1 = 0.6;    % G1 mean & variance
var1 = 1;
D = 1024 ;              % overall sequence length
k = 0:6 ;               % number of experiemnets(7)
Dm = 2.^k ;             % length of Dm with respect to k
m = D./Dm ;             % number of sequences Dm

% Init arrays
Gen = zeros(length(k),D);           % create matrix of 7x1024, each
row = experiment
failure = zeros(1,length(k));       % fails counter for each k
KL0 = zeros(1,length(k));           % KL for G0 ,for each k
KL1 = zeros(1,length(k));           % KL for G1 ,for each k

for experiment = 1:length(k)
    g0_index = 1;
    g1_index = 1;
    G0_ML=[];
    G1_ML=[];
    for seq = 1:m(experiment) % run on small sequences
        qm = binornd(1,1/2) ; % Bernoulli with p=0.5,every iteration
get 0 or 1
% Generation
        if (qm == 0)  %Generate subgroup from G0
            Gen(experiment, ((seq-
1)*Dm(experiment)+1):(seq*Dm(experiment))) =
sqrt(var0).*randn(1,Dm(experiment)) + mu0 ;
        else          %Generate subgroup from G1
            Gen(experiment, ((seq-
1)*Dm(experiment)+1):(seq*Dm(experiment))) =
sqrt(var1).*randn(1,Dm(experiment)) + mu1 ;
        end

% ML estimation to determine which Gaussian is Dm
        Xn = Gen(experiment, ((seq-
1)*Dm(experiment)+1):(seq*Dm(experiment)));
        N = length(Xn);
        sum_x=0;
        mu_ml=0;
        for i= 1:N%%mean ML
            sum_x= sum_x +Xn(i);
        end
        mu_ml=sum_x/N;
        S=0;
        for i = 1:N%%S (VARIANCE without norG1_ML
            S = S + (Xn(i)-mu_ml)^2;
        end
        L0 = -N*log((var0*2*pi)^0.5)- (N*(mu_ml-mu0)^2+S)/(2*var0);
        L1 = -N*log((var1*2*pi)^0.5)- (N*(mu_ml-mu1)^2+S)/(2*var1);
        if (L0>L1)
```

```
            qm_est = 0;
            G0_ML(g0_index:g0_index+Dm(experiment)-1) = Xn ;
            g0_index = g0_index + Dm(experiment) ;
        else
            qm_est = 1 ;
            G1_ML(g1_index:g1_index+Dm(experiment)-1) = Xn ;
            g1_index = g1_index + Dm(experiment) ;
        end

        if (qm ~= qm_est)
            failure(experiment) = failure(experiment)+1 ;   % failure
counter
        end

    end

% got all subgroups of the current experiment, the features are:
    mu0_est = mean(G0_ML);  var0_est = var(G0_ML);
    mu1_est = mean(G1_ML);  var1_est = var(G1_ML);

% KL calculation for each experiment:
    KL0(experiment) = log(sqrt(var0_est)/sqrt(var0)) + (var0+(mu0-
mu0_est)^2)/(2*var0_est) - 0.5;
    KL1(experiment) = log(sqrt(var1_est)/sqrt(var1)) + (var1+(mu1-
mu1_est)^2)/(2*var1_est) - 0.5;

%%%%%%%%%%% Plot results:

% G0 & G1
    figure(1)
    x = -10:.1:10;
    subplot(6,2,experiment);
    norm = normpdf(x,mu0,var0);
    plot(x,norm); hold on ;
    norm = normpdf(x,mu0_est,var0_est);
    plot(x,norm);
    str = sprintf('k = %d',experiment-1);
    title(str);
    xlabel('x'); ylabel('f(x)');
    legend('G0','G0__ML');

    figure(2)
    x = -10:.1:10;
    subplot(6,2,experiment);
    norm = normpdf(x,mu1,var1);
    plot(x,norm); hold on ;
    norm = normpdf(x,mu1_est,var1_est);
    plot(x,norm);
    str = sprintf('k = %d',experiment-1);
    title(str);
    xlabel('x'); ylabel('f(x)');
    legend('G1','G1__ML');
end


% Errors in estimation qm:
error_rate = (failure./m );
figure(3)
stem(k,error_rate,'LineWidth',2);
```

```
xlim([-0.5,6.5]);
title('Estimations error for each k');
xlabel('k'); ylabel('Relative errors');

% KL divergence:
figure(4)
stem(k,KL0,'LineWidth',2,'Color','r');
xlim([-0.5,6.5]);hold on;
stem(k+0.1,KL1,'LineWidth',2,'Color','b');
title('Kullback-Leibler values for each k');
xlabel('k'); ylabel('KL divergence'); legend('KL0','KL1');
```