# تقدير لامعلمي لدالة كثافة احتمالية متعددة المتغيرات

**مناف يوسف حمود*، تهاني مهدي عباس** و قتيبة نبيل نايف*****

* قسم الإحصاء، كلية الإدارة والإقتصاد، جأمعة بغداد.

** كـلـيـــة العـلـــوم، جـامعـــة بغــداد.

*** قسم الإحصاء، كلية الإدارة والإقتصاد، جامعة بغداد.

**المستخلص**

في هذا البحث تم استعراض بعض الطرائق الخاصة بتقدير دالة كثافة احتمالية متعددة المتغيرات.اذ تم التركيز على توزيع طبيعي ثنائي المتغيرات وكذلك توزيع ملوث ثنائي المتغيرات.وقد تم مقارنة ثلاث مقدرات (تمثل المقدر الاول بمقدر معلمي وهو مقدر الامكان الاعظم مع مقدرين لامعلميين هما مقدر لامعلمي ذو مصفوفة كاملة للمعالم التمهيدية مع مقدر لامعلمي ذو مصفوفة قطرية للمعالم التمهيدية)، وذلك باستخدام اسلوب المحاكاة.

## 1- المقدمة

يعد تقدير دالة الكثافة متعددة المتغيرات باستخدام التقدير اللُبي من الاساليب الحديثة والمفيدة في الوصف البياني والمرئي( Visual) الامثل، ويستخدم هذا الاسلوب في فحص تركيب البيانات وبناء الانموذج الملائم بالاعتماد على البيانات المعطاة فضلا عن تزويدنا باداة تحليلية مرنة غالبا ما تستخدم في تمثيل البيانات بيانيا.

ان الهدف من تقدير دالة الكثافة اللامعلمية متعددة المتغيرات يتمثل بتقريب دالة الكثافة الاحتمالية $f(\underline{x}) = f(x_1, x_2, ..., x_d)$ للمتغيرات العشوائية $\underline{X} = (X_1, X_2, ...., X_d)^T$. يعرف مقدر اللّب لدالة الكثافة متعددة المتغيرات ذات البعد d وكالأتي:

$$\hat{f}(\underline{x}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_1.h_2....h_d}K\left(\frac{X_{i1}-x_1}{h_1}, ..., \frac{X_{id}-x_d}{h_d}\right) \quad ......................(1)$$

اذ يشير $K(.)$ الى دالة اللُب متعددة المتغيرات، مع الاشارة الى انه تم افتراض ان المعلمة التمهيدية في المعادلة المذكورة انفا تمثل متجه من المعالم التمهيدية $\underline{h} = (h_1, h_2, ..., h_d)^T$.

هنالك عدة صيغ تاخذها دالة اللّب منها دالة اللُب الناتجة من حاصل ضرب دوال لّب احادية المتغير:

$$K(\underline{z}) = \kappa(z_1).\kappa(z_2)....\kappa(z_d)$$

اذ يشير $k(.)$ الى دالة لّب احادية المتغير مثل دالة Gaussian:

$$\kappa(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} I(-\infty, \infty)$$

وهنالك صيغة اخرى تتمثل باستخدام دالة لّب متعددة المتغيرات:

$$K(\underline{z}) = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{1}{2}\underline{z}^T\underline{z}}$$

اما الصيغة الثانية لتقدير دالة الكثافة متعددة المتغيرات وتتمثل باستعمال مصفوفة معلمة تمهيدية متماثلة غير صفرية $H$ ومن ثم تصبح الصيغة العامة لمقدردالة الكثافةالأحتمالية متعددة المتغيرات كالاتي:

$$\hat{f}_h(\underline{x}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{|H|}K\left(H^{-1}(\underline{X}-\underline{x})\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}K_H(\underline{X}-\underline{x}) \quad ..........................(2)$$

## 2- بعض الصيغ لمصفوفة المعلمة التمهيدية

هنالك بعض الصيغ التي تتضمنها مصفوفة المعلمة التمهيدية تتمثل الصيغة الاولى باستعمال معلمة تمهيدية متساوية في جميع الابعاد وتناظر تلك الى قيمة المصفوفة:

$$H = hI_d$$

اذ يشير $I_d$ الى مصفوفة وحدة ذات بُعد $d\times d$.

في حين تتمثل الصيغة الثانية باستعمال قيم مختلفة للمعلمة التمهيدية كما في حالة المعادلة (1) المذكورة انفا،مما ينتج ان:

$$H = diag(h_1, h_2, ..., h_d)$$

اي تمثل مصفوفة قطرية بالعناصر $h_1, h_2, ..., h_d$.

اما الصيغة الثالثة والتي تعد الحالة العامة فتتمثل بكون المصفوفة غير قطرية ولها قيم خارج القطر ويتم استعمال مصفوفة معلمة تمهيدية متناسبة الى $\hat{\Sigma}^{-1/2}$ اي ان $H \alpha \hat{\Sigma}^{-1/2}$، اذ يشير $\hat{\Sigma}$ الى مصفوفة التباين المشترك للبيانات.

لذلك فان استعمال هذه المعلمة التمهيدية سوف يكون مطابقا الى تحويل البيانات الى مصفوفة وحدة للتباين المشترك، وكنتيجة لهذا فانه بالامكان استعمال مصفوفات معلمة تمهيدية لتصحيح الارتباط بين المركبات لـــ $\underline{X}$.

**3- الخصائص الاحصائية للمقدر متعدد المتغيرات**

نتيجة للافتراض القياسي على دالة اللّب غير السالبة:

$$\int K(\underline{z})d\underline{z} = 1 \quad ..........................................(3)$$

فأن التقدير $\hat{f}_H(\underline{x})$ يمثل دالة كثافة يحقق $\int \hat{f}_H(\underline{x})d\underline{x} = 1$ وان التقدير الاتي يكون:

$$\hat{f}_H(\underline{x}) = \frac{1}{n}\sum_{i=1}^{n} K_H(\underline{x} - \underline{X_i}) = f(\underline{x}) + o_P(1) \quad ..(4)$$

متسقا عند اي نقطة استمرارية للدالة f(.)، اي انه اذا كانت $n \to \infty$ و $|H| \to \infty$ فان $n|H| \to \infty$ مما ينتج ان الاشتقاق لمتوسط مربع الخطأ MSE ومتوسط مربعات الخطا التكاملي IMSE يكون مشابها الى حالة البعد الواحد.

ومن الجدير بالاشارة الى ان الرمز الذي يشير الى التكامل انما يشير تكامل متعدد $\int\int...\int_d$ وان $d\underline{x} = dx_1 dx_2 ... dx_d$.

ان معيار متوسط مربعات الخطأ التكاملي المحاذي AIMSE يتكون من جزئين هما التحيز التكاملي المحاذي Asymptotic integrated bias (AIB) والتباين المحاذي التكاملي Asymptotic integrated variance (AIV) اذ يكون التحيز للمقدار هو $E\hat{f}_H(\underline{x}) - f(\underline{x})$ وان تكامل مربع التحيز يكون:

$$ISB(H) = \int \{E\hat{f}_H(\underline{x}) - f(\underline{x})\}^2 d\underline{x} \quad ..............(5)$$

مع الاشارة الى ان تكامل مربع التحيز المحاذي يمثل حدا من الرتبة الاولى للمقدار ISB(H)، اي ان:

$$\frac{ISB(H) - AIB(H)}{AIB(H)} = o(1)$$

ويلاحظ انه تم افتراض ان $n \to \infty$ وان $|H| \to \infty$ وكذلك $n|H| \to \infty$. اما مقدار التباين التكاملي فيكون:

$$IV(H) = \int E\left\{\hat{f}_H(\underline{x}) - E\hat{f}_H(x)\right\}^2 d\underline{x} \quad ...........(6)$$

بدمج كلا من AIB(H) مع AIV(H) نحصل على متوسط مربعات الخطأ التكاملي المحاذي:

AIMSE (H) = AIV (H) + AIB (H)

ولاشتقاق المركبات لـــ AIMSE(H) وكما في حالة احادي المتغير نستعمل مفكوك تايلور من الرتبة الثانية وسوف يشار الى gradient بالرمز $\nabla$ ولمصفوفة الـــ Hessian والتي تمثل مصفوفة المشتقات الجزئية من الرتبة الثانية للدالة f ويرمز لها بـــ $H_f$ ومن ثم فان مفكوك تايلور لـــ f(.) حول $\underline{x}$ يكون:

$$f(\underline{x} + \underline{z}) = f(\underline{x}) + \underline{z}^T \nabla_f(\underline{x}) + \frac{1}{2}\underline{z}^T H_f(\underline{x})\underline{x} + o(\underline{z}^T \underline{z})$$

ان هذه المعادلة تقود الى المقدار:

$$E\hat{f}_H(\underline{x}) = \int K_H(\underline{z} - \underline{x}) f(\underline{z}) d\underline{z} = \int K(\underline{s}) f(\underline{x} + H\underline{s}) d\underline{s}$$

$$\approx \int K(\underline{s})\{f(\underline{x}) + \underline{s}^T H^T \nabla_f(\underline{x}) + \frac{1}{2}\underline{s}^T H^T H_f(\underline{x}) H\underline{s}\} d\underline{s} \quad ............(7)$$

وباستخدام الافتراض في المعادلة (3) فضلا عن الافتراضات الاتية:

$$\int \underline{z} K(\underline{z}) d\underline{z} = \underline{0}_d$$

$$\int \underline{z}\,\underline{z}^T K(\underline{z}) d\underline{z} = M_2(K) I_d$$

فان المعادلة (7) تصبح:

$$E\hat{f}_H(\underline{x}) - f(\underline{x}) \approx \frac{1}{2} M_2(K)\, tr(H^T H_f(\underline{x}) H)$$

لذلك فأن:

$$AIB(H)=\frac{1}{4}M_2^2(K)\int[tr(H^T H_f(\underline{x})H)]^2 d\underline{x}$$ .............(8)

اما الحد الرئيس لجزء التباين يمثل العزم الثاني للتقدير اي ان:

$$Var\{\hat{f}_H(\underline{x})\}=\frac{1}{n}\int\left\{K_H(\underline{z}-\underline{x})\right\}^2 d\underline{z}-\frac{1}{n}\left\{E\hat{f}_H(\underline{x})\right\}^2$$
$$\approx\int\frac{1}{n|H|}K^2(\underline{s})f(\underline{x}+H\underline{s})d\underline{s}$$
$$\approx\int\frac{1}{n|H|}K^2(\underline{s})f(\underline{x}+\underline{s}^T H^T\nabla_f(\underline{x}))d\underline{s}$$
$$\approx\frac{1}{n|H|}\|K\|_2^2 f(\underline{x})$$ ..........................(9)

اذ يشير $\|K\|_2$ الى $L_2$−norm ذات d من الابعاد للدالة K(.)، لذلك فأن التباين المحاذي التكاملي يكون:

$$AIV(H)=\frac{1}{n|H|}\|K\|_2^2$$ ...............................(10)

ومن ثم فان:

$$AIMSE(H)=\frac{1}{4}M_2^2(K)\int\left[tr(H^T H_f(\underline{x})H)\right]^2 d\underline{x}+\frac{1}{n|H|}\|K\|_2^2$$ .........................(11)

## 4- اختيار المعلمة التمهيدية لمقدر الكثافة متعدد المتغيرات

نتحول الان الى مسالة اخرى وتتمثل باختيار المعلمة التمهيدية المثلى، اذ تعمل هذه المعلمة على الموازنة بين التحيز والتباين في مقدار AIMSE. بافتراض ان h تمثل ثابتا بحيث ان $H=hH_0$ وان $|H_0|=1$ فأن المقدار AIMSE(H) يمكن كتابته كالاتي:

$$AIMSE(H)=\frac{1}{4}h^4M_2^2(K)\int\left[tr(H_0{}^T H_f(\underline{x})H_0)\right]^2 d\underline{x}+\frac{1}{nh^d}\|K\|_2^2$$ .............................(12)

اذ ان مسالة اختيار المعلمة التمهيدية H ذات اهمية عظمى في حالة متعدد المتغيرات ويمكن استخدام الاساليب المتبعة في حالة البعد الواحد في تقدير المعلمة التمهيدية في حالة متعدد المتغيرات (اذ نبحث عن المعلمة التمهيدية الشاملة H او الموضعية $H(\underline{x})$) ومن هذه الطرائق (الملئ Plug-in والعبور الشرعي Cross-Validation وقاعدة الابهامRule of thumb).

### 4.1 قاعدة الابهام *Rule of thumb*:

ان اختيار المعلمة التمهيدية المستندة على قاعدة الابهام يظهر من استخدام الصيغة لتوزيع المصدر (Reference distribution). من الواضح ان دالة الكثافة الاحتمالية لدالة التوزيع الطبيعي متعدد المتغيرات $N_d(\underline{\mu},\Sigma)$ تمثل مرشحا جيدا لتوزيع المصدر في متعدد المتغيرات.
بافتراض ان دالة اللّب $K(.)$ تمثل دالة Gaussian، اي انها تمثل PDF لدالة التوزيع الطبيعي $N_d(\underline{0}_d,I_d)$ مع ملاحظة ان $M_2(K)=1$ وان $\|K\|_2^2=2^{-d}\pi^{-d/2}$،لذلك ومن المعادلة (11) وحقيقة كون:

$$\int\left[tr(H^T H_f(\underline{x})H)\right]^2 d\underline{x}=\frac{1}{2^{d+2}\pi^{d/2}|\Sigma|^{1/2}}\left[2tr(H^T\Sigma^{-1}H)^2+\{tr(H^T\Sigma^{-1}H)\}^2\right]$$

الحالة الابسط تتمثل بكون كلا من $\Sigma$ و $H$ مصفوفتان قطريتان،اي ان:

$$H=diag(h_1,h_2,...,h_d)$$
$$\Sigma=diag(\sigma_1^2,\sigma_2^2,...,\sigma_d^2)$$

فنحصل على:

$$\tilde{h}_j=\left(\frac{4}{d+2}\right)^{1/(d+4)}n^{-1/(d+4)}\sigma_j$$ ................(13)

يلاحظ ان هذه الصيغة تتطابق مع صيغة قاعدة الابهام للباحث Silverman (1986) في حالة البعد الواحد اي عندما $d=1$.

وباستبدال $\sigma_j$ باالتقدير الملائم له نتوصل الى قاعدة Scott (1992) وهي:

$$\tilde{h}_j=n^{-1/(d+4)}\hat{\sigma}_j$$ .......................................(14)

من المعادلة (13) المذكورة انفا يلاحظ امكانية اختيار مصفوفة معلمة تمهيدية H متناسبة الى $\Sigma^{1/2}$، ففي هذه الحالة نحصل وكتعميم لقاعدة Scott (1992) على:

$$\hat{H}=n^{-1/(d+4)}\hat{\Sigma}^{1/2}$$ ....................................(15)

اذ تعد هذه القاعدة مساوية الى تطبيق تحويل مهنلوبس Mahalanbois للبيانات (اي تحويل مصفوفة التباين المشترك المقدرة الى مصفوفة وحدة) ومن ثم حساب تقدير اللّب مع معالم تهيدية متساوية $h = n^{-1/(d+4)}$ ومن ثم واخيراً نعيد تحويل دالة الكثافة المقدرة الى المقياس الاصلي.

**5- الجانب التجريبي**

في هذا المبحث تم استعمال اسلوب المحاكاة لغرض مقارنة المقدرات لدالة الكثافة الاحتمالية متعددة المتغيرات وهي مقدر الامكان الاعظم ومقدر كثافة لامعلمي ذو مصفوفة معالم تمهيدية كاملة وكذلك مقدر ذو مصفوفة قطرية للمعالم التمهيدية.

وقد تمت المقارنة بالاعتماد على بضعة حالات منها:

- استخدام توزيع طبيعي ثنائي المتغيرات لكن بتباينات مختلفة للمتغيرات المستخدمة وهي على التوالي:

$$X \sim N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)$$

$$X \sim N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1 & \rho\sqrt{2}\\ \rho\sqrt{2} & 2\end{pmatrix}\right)$$

$$X \sim N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}2 & \rho\sqrt{10}\\ \rho\sqrt{10} & 5\end{pmatrix}\right)$$

وقد تم استعمال قيم مختلفة للارتباطات هي $\rho = \mp 0.3, \mp 0.5, \mp 0.8$ . كذلك تم استخدام حجوم مختلفة للعينات ($n = 50,\ 100,\ 200$).

- استخدام توزيع طبيعي ثنائي المتغيرات ملوث بنسب $\alpha = 30\%,\ 40\%$ :

والتوزيع الملوث المستخدم في هذا البحث هو:

$$f(\underline{x}) = (1-\alpha)N(\underline{0},\Sigma) + \alpha N(\underline{0},\lambda^2\Sigma)$$

اذ ان $\lambda = 4$ .

والجداول الاتية توضح نتائج المحاكاة وفق المعطيات المذكورة انفا، كذلك تم عرض بعض الاشكال لعدد من الحالات المستخدمة.

علما ان معيار المستخدم لغرض المقارنة يتمثل بمعيار متوسط مربعات الخطأ MSE.

## جدول (1)

**يشير الى قيم $MSE(\hat{f}(\underline{x}))$ وفق المقدرات المستخدمة وقيم التباينات وحجوم العينات المستخدمة مع الاشارة الى ان نسبة ثنائي المنوال(التلوث) هي%0 ( لعدد الناتج مضروب في 10000).**

| n | $\rho$ | 0.3 | | | 0.5 | | | 0.8 | | | -0.3 | | | -0.5 | | | -0.8 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\sigma_1^2$ | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 |
| | $\sigma_2^2$ | 1 | 2 | 5 | 1 | 2 | 5 | 1 | 2 | 5 | 1 | 2 | 5 | 1 | 2 | 5 | 1 | 2 | 5 |
| 50 | $\hat{f}_{ML}$ | 4.212 | 2.02 | 0.43 | 15 | 7.16 | 1.68 | 370 | 174 | 34 | 4.23 | 2.15 | 0.41 | 16 | 7.79 | 1.57 | 333 | 174 | 35 |
| | $\hat{f}_{Hd}$ | 7.65 | 3.3 | 0.55 | 11 | 4.64 | 0.71 | 43 | 18 | 2.41 | 7.58 | 3.19 | 0.535 | 11 | 4.71 | 0.698 | 42 | 18 | 2.37 |
| | $\hat{f}_{Hf}$ | 7.8 | 3.37 | 0.56 | 12 | 4.86 | 0.75 | 45 | 20 | 2.76 | 7.73 | 3.26 | 0.545 | 11 | 4.95 | 0.738 | 45 | 19 | 2.72 |
| 100 | $\hat{f}_{ML}$ | 2.19 | 1.03 | 0.22 | 11 | 5.26 | 1.1 | 291 | 142 | 30 | 2.11 | 1.15 | 0.21 | 10 | 5.45 | 1.01 | 289 | 146 | 29 |
| | $\hat{f}_{Hd}$ | 6.57 | 2.74 | 0.393 | 9.67 | 3.95 | 0.54 | 40 | 16 | 2.11 | 6.63 | 2.70 | 0.383 | 9.79 | 3.93 | 0.532 | 40 | 17 | 2.14 |
| | $\hat{f}_{Hf}$ | 6.73 | 2.82 | 0.41 | 10 | 4.22 | 0.59 | 43 | 18 | 2.58 | 6.79 | 2.77 | 0.396 | 10 | 4.21 | 0.583 | 43 | 19 | 2.61 |
| 200 | $\hat{f}_{ML}$ | 1.18 | 0.629 | 0.61 | 8.09 | 4.18 | 0.82 | 296 | 133 | 27 | 1.29 | 0.59 | 0.808 | 8.19 | 4.13 | 0.841 | 269 | 136 | 27 |
| | $\hat{f}_{Hd}$ | 5.63 | 2.24 | 2.27 | 8.44 | 3.41 | 0.43 | 36 | 15 | 1.86 | 5.62 | 2.29 | 0.426 | 8.402 | 3.42 | 0.423 | 36 | 15 | 1.87 |
| | $\hat{f}_{Hf}$ | 5.80 | 2.33 | 2.36 | 9.07 | 3.73 | 0.48 | 40 | 17 | 2.39 | 5.79 | 2.37 | 0.483 | 9.03 | 3.74 | 0.498 | 40 | 17 | 2.40 |

## جدول (2)

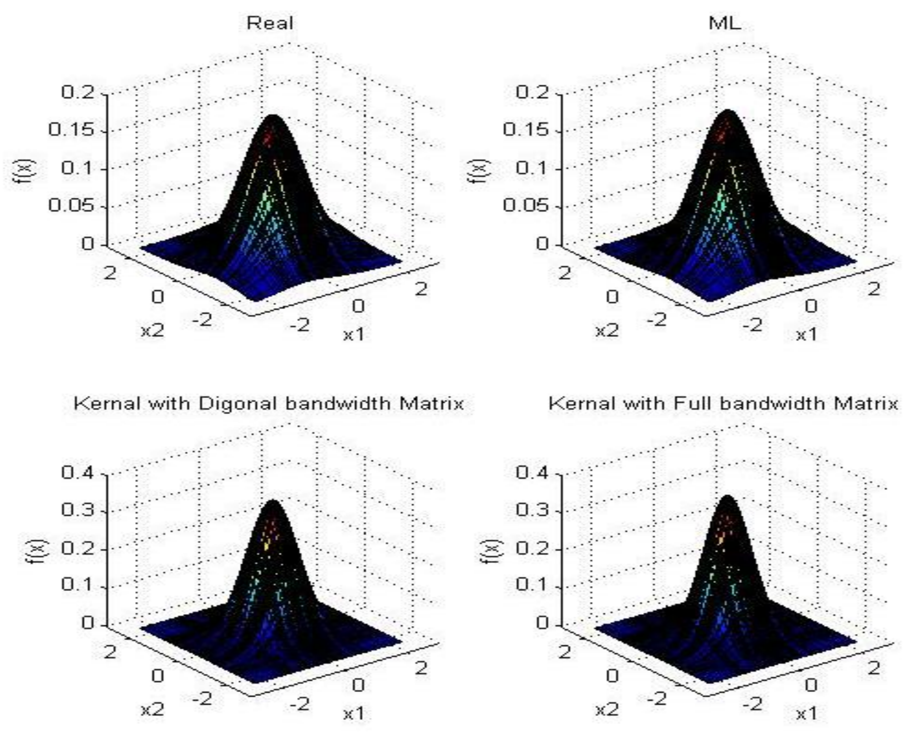
**يشير الى قيم $MSE(\hat{f}(\underline{x}))$ وفق المقدرات المستخدمة وقيم التباينات وحجوم العينات المستخدمة مع الاشارة الى ان نسبة ثنائي المنوال (التلوث) هي %30(العدد الناتج مضروب في 10000).**

| n | $\rho$ | 0.3 | | | 0.5 | | | 0.8 | | | -0.3 | | | -0.5 | | | -0.8 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\sigma_1^2$ | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 |
| | $\sigma_2^2$ | 1 | 2 | 5 | 1 | 2 | 5 | 1 | 2 | 5 | 1 | 2 | 5 | 1 | 2 | 5 | 1 | 2 | 5 |
| 50 | $\hat{f}_{ML}$ | 32 | 16 | 3.21 | 34 | 17 | 3.35 | 33 | 16 | 3.21 | 32 | 16 | 3.21 | 33 | 17 | 3.33 | 32 | 16 | 2.99 |
| | $\hat{f}_{Hd}$ | 23 | 11 | 1.66 | 30 | 14 | 2.17 | 81 | 37 | 6.15 | 23 | 11 | 1.65 | 30 | 14 | 2.18 | 81 | 37 | 6.095 |
| | $\hat{f}_{Hf}$ | 23 | 11 | 1.67 | 30 | 14 | 2.21 | 80 | 37 | 6.27 | 23 | 11 | 1.66 | 30 | 14 | 2.21 | 81 | 37 | 6.22 |
| 100 | $\hat{f}_{ML}$ | 34 | 17 | 3.40 | 35 | 18 | 3.57 | 30 | 15 | 2.86 | 34 | 17 | 3.38 | 36 | 18 | 3.55 | 29 | 15 | 2.90 |
| | $\hat{f}_{Hd}$ | 22 | 10 | 1.59 | 29 | 13 | 2.1 | 78 | 36 | 5.92 | 22 | 10 | 1.60 | 29 | 13 | 2.10 | 78 | 36 | 5.90 |
| | $\hat{f}_{Hf}$ | 22 | 10 | 1.61 | 29 | 13 | 2.14 | 78 | 36 | 6.11 | 22 | 10 | 1.61 | 29 | 13 | 2.14 | 78 | 36 | 6.097 |
| 200 | $\hat{f}_{ML}$ | 35 | 17 | 3.45 | 36 | 18 | 3.63 | 29 | 15 | 2.98 | 34 | 17 | 3.44 | 37 | 18 | 3.64 | 29 | 15 | 2.945 |
| | $\hat{f}_{Hd}$ | 21 | 9.56 | 1.52 | 28 | 13 | 1.99 | 75 | 35 | 5.64 | 21 | 9.59 | 1.51 | 28 | 13 | 2.01 | 74 | 34 | 5.64 |
| | $\hat{f}_{Hf}$ | 21 | 9.61 | 1.53 | 28 | 13 | 2.05 | 75 | 35 | 5.89 | 21 | 9.64 | 1.53 | 28 | 13 | 2.06 | 75 | 35 | 5.89 |

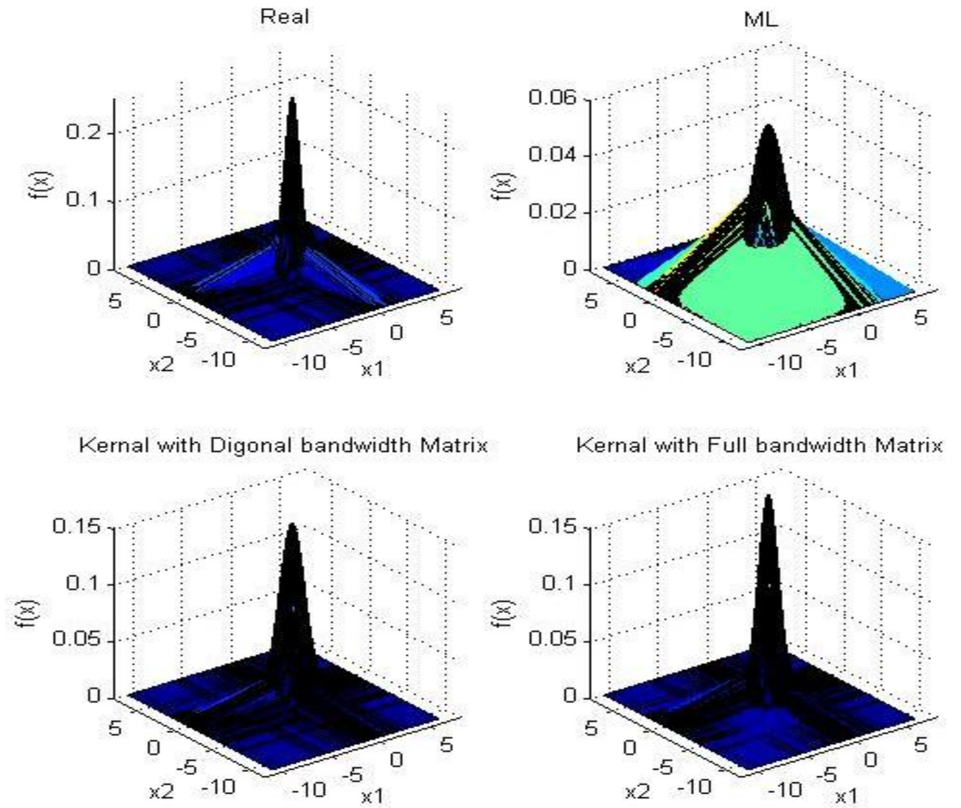
**جدول (3)**

**يشير الى قيم $MSE(\hat{f}(\underline{x}))$ وفق المقدرات المستخدمة وقيم التباينات وحجوم العينات المستخدمة مع الاشارة الى ان نسبة ثنائي المنوال (التلوث) هي40% (العدد الناتج مضروب في 10000).**

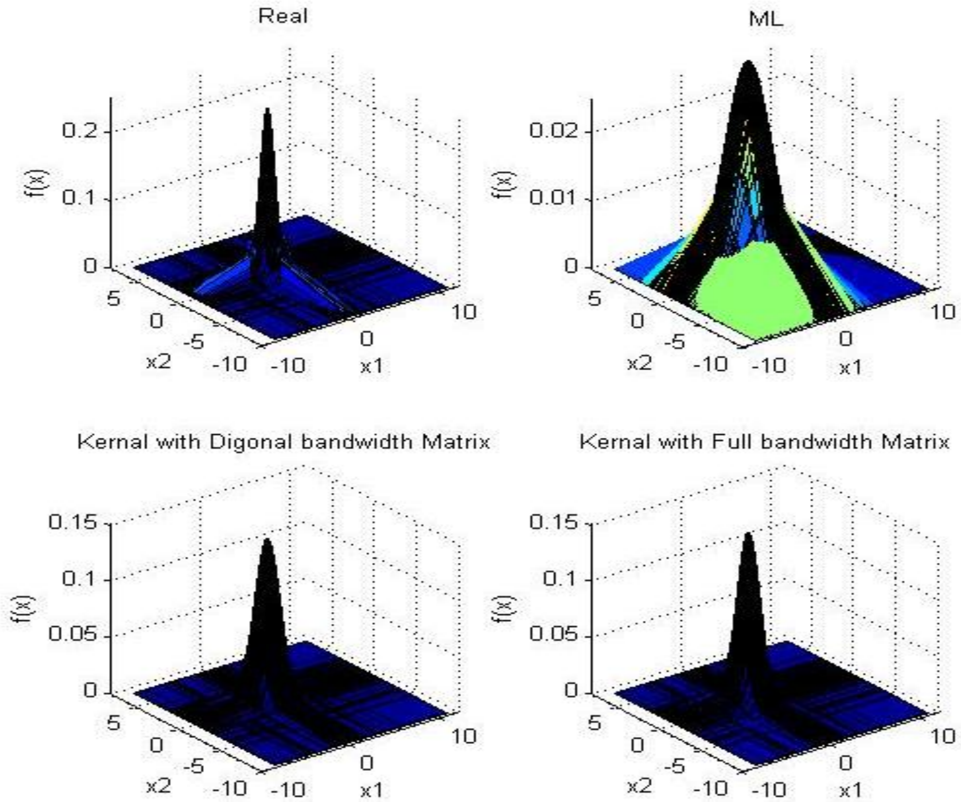
| n | $\rho$ | 0.3 | | | 0.5 | | | 0.8 | | | -0.3 | | | -0.5 | | | -0.8 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\sigma_1^2$ | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 |
| | $\sigma_2^2$ | 1 | 2 | 5 | 1 | 2 | 5 | 1 | 2 | 5 | 1 | 2 | 5 | 1 | 2 | 5 | 1 | 2 | 5 |
| 50 | $\hat{f}_{ML}$ | 33 | 16 | 3.28 | 35 | 17 | 3.48 | 38 | 19 | 3.68 | 32 | 16 | 3.21 | 35 | 18 | 3.479 | 36 | 19 | 3.676 |
| | $\hat{f}_{Hd}$ | 25 | 11 | 1.86 | 32 | 15 | 2.41 | 78 | 37 | 6.30 | 24 | 11 | 1.84 | 32 | 15 | 2.402 | 78 | 37 | 6.32 |
| | $\hat{f}_{Hf}$ | 25 | 11 | 1.87 | 32 | 15 | 2.43 | 78 | 37 | 6.39 | 24 | 11 | 1.85 | 32 | 15 | 2.426 | 78 | 37 | 6.404 |
| 100 | $\hat{f}_{ML}$ | 34 | 17 | 3.38 | 37 | 18 | 3.64 | 38 | 19 | 3.80 | 33 | 17 | 3.36 | 36 | 18 | 3.635 | 38 | 19 | 3.847 |
| | $\hat{f}_{Hd}$ | 24 | 11 | 1.81 | 30 | 14 | 2.33 | 76 | 36 | 6.15 | 23 | 11 | 1.818 | 30 | 14 | 2.337 | 77 | 36 | 6.157 |
| | $\hat{f}_{Hf}$ | 24 | 11 | 1.82 | 30 | 14 | 2.36 | 76 | 36 | 6.29 | 23 | 11 | 1.83 | 30 | 14 | 2.368 | 77 | 36 | 6.295 |
| 200 | $\hat{f}_{ML}$ | 34 | 17 | 3.40 | 37 | 19 | 3.73 | 39 | 20 | 3.93 | 34 | 17 | 3.40 | 37 | 19 | 3.70 | 39 | 20 | 3.84 |
| | $\hat{f}_{Hd}$ | 23 | 10 | 1.74 | 29 | 13 | 2.25 | 74 | 35 | 5.96 | 23 | 10 | 1.74 | 29 | 13 | 2.247 | 74 | 35 | 5.938 |
| | $\hat{f}_{Hf}$ | 23 | 10 | 1.75 | 29 | 14 | 2.29 | 75 | 35 | 6.15 | 23 | 10 | 1.75 | 29 | 14 | 2.285 | 74 | 35 | 6.127 |




**شكل رقم (1) يشير الى القيم الحقيقية والتقديرات المعلمية واللامعلمية لدالة الكثافة الإحتمالية عتد حجم عينة $n = 100$ وتباينات $\sigma_2^2 = 1$ , $\sigma_1^2 = 1$ ودرجة أرتباط $\rho = 0.5$ ونسبة تلوث 0% .**




**شكل رقم (2) يشير الى القيم الحقيقية والتقديرات المعلمية واللامعلمية لدالة الكثافة الإحتمالية عند حجم عينة n = 100 وتباينات** $\sigma_1^2 = 1$ , $\sigma_2^2 = 1$ **ودرجة أرتباط** $\rho = 0.5$ **ونسبة تلوث** $30\%$.




**شكل رقم (3) يشير الى القيم الحقيقية والتقديرات المعلمية واللامعلمية لدالة الكثافة الإحتمالية عند حجم عينة n = 100 وتباينات** $\sigma_1^2 = 1$ , $\sigma_2^2 = 1$ **ودرجة أرتباط** $\rho = 0.5$ **ونسبة تلوث** $40\%$.

## 6- تفسير النتائج

من الجدول (1) تلاحظ النتائج الاتية:

- في حالة الارتباط $\rho = \mp 0.3$ ، ولجميع حجوم العينات والتباينات المستخدمة اظهرت النتائج ان المقدر الافضل هو مقدر الامكان الاعظم يلية المقدر اللامعلمي ذو مصفوفة المعلمة التمهيدية القطرية يليه المقدر اللامعلمي ذو مصفوفة المعلمة التمهيدية الكاملة.
- في حالة الارتباط $\rho = \mp 0.5$ ، ولجميع حجوم العينات والتباينات المستخدمة اظهرت النتائج ان المقدر اللامعلمي ذو مصفوفة المعلمة التمهيدية القطرية افضل اداءا يليه المقدر اللامعلمي ذو مصفوفة المعلمة التمهيدية الكاملة واخيرا مقدر الامكان الاعظم.
- من الجدير بالاشارة الى انه عند زيادة قيم التباين للمتغيرات وحالة حجوم العينات الصغيرة والمتوسطة اظهرت النتائج افضلية مقدر الامكان ثم المقدر اللامعلمي ذو مصفوفة المعلمة التمهيدية القطرية يليه المقدر اللامعلمي ذو مصفوفة المعلمة التمهيدية الكاملة، في حين عند حجم العينة الكبير اظهرت النتائج تباينا عند قيم $\underline{\sigma_1^2} = \begin{pmatrix}1\\2\end{pmatrix}, \underline{\sigma_2^2} = \begin{pmatrix}1\\5\end{pmatrix}$ .
- عند حالة الارتباط $\rho = \mp 0.8$ و لجميع الحالات الاخرى المستعملة، اثبتت النتائج تفوق المقدر اللامعلمي ذو مصفوفة المعلمة التمهيدية القطرية يليه المقدر اللامعلمي الثاني.
- اظهرت النتائج ان قيم MSE لاتتاثر نوعا ما بقيم الارتباط لكل قيمة ارتباط مستخدم سواء اكان سالبا ام موجبا اي اذا كان الارتباط بين المتغيرين 0.3 فان النتائج كانت متشابهة نوعا ما مع حالة الارتباط هو 0.3- وهكذا لبقية القيم الاخرى للارتباط، اي ان النتائج كانت متقاربة لكل حالة ارتباط مستخدمة.
- تتصاعد قيم الخطا مع تصاعد قيم الارتباطات بين المتغيرات.
- يلاحظ ان مقدر الامكان يتحسن في حالة الارتباطات المتوسطة وقيم تباين $\sigma_1^2 = 2, \sigma_2^2 = 5$ لكن عند زيادة قيم الارتباط ينخفض اداء تلك الطريقة ويتحسن اداء المقدرات اللامعلمية.

من الجداول (2، 3) تلاحظ النتائج الاتية:

- عند زيادة نسب التلوث تزداد قيم الخطا للمتغيرات ولجميع حجوم العينات والارتباطات والتباينات المستعملة.
- في حالة الارتباطات العالية ولجميع حجوم العينات والتباينات اظهرت النتائج تحسن اداء الامكان الاعظم ويفضل استخدامه على بقية المقدرات الاخرى على الرغم من تذبذب قيمه لجميع الارتباطات الاخرى المستخدمة.
- لحالة النسبة % 30 ولحالة $\underline{\sigma_1^2} = \begin{pmatrix}1\\1\end{pmatrix}, \underline{\sigma_2^2} = \begin{pmatrix}1\\2\end{pmatrix}$ ولجميع حجوم العينات والارتباطات يتضح تساوي المقدرات اللامعلمية المستخدمة، اما في حالة $\sigma_1^2 = 2, \sigma_2^2 = 5$ اظهرت النتائج افضلية المقدر اللامعلمي ذو المصفوفة القطرية للمعلمة التمهيدية يليه المقدر اللامعلمي ذو مصفوفة المعلمة التمهيدية الكاملة واخيرا مقدر الامكان.
- تحسن قيم MSE بشكل طفيف عند زيادة حجوم العينات.
- كحالة عامة اظهرت النتائج ثبات نسبي لقيم MSE لجميع المقدرات بتغير قيم الارتباطات.
- تحسن قيم MSE لكل مقدر عند زيادة قيم التباين ولكل حالة ارتباط مستخدمة.
- اظهرت النتائج افضلية المقدرات اللامعلمية على مقدر الامكان الاعظم في حالة استخدام الارتباطات $\rho = (\mp 0.3, \mp 0.5)$ ولجميع حجوم العينات والتباينات المستخدمة.
- في حالة الارتباط $\rho = \mp 0.8$ يلاحظ افضلية مقدر الامكان على بقية المقدرات.

## 7- الاستنتاجات

- **لحالة التوزيع الطبيعي:**

1. في حالة الارتباطات الصغيرة اظهرت النتائج افضلية مقدر الامكان الاعظم على بقية المقدرات.
2. في حالة الارتباطات المتوسطة والكبيرة اظهرت النتائج افضلية المقدرات اللامعلمية خاصة المقدر اللامعلمي ذو المصفوفة القطرية للمعالم التمهيدية يليه

المقدر اللامعلمي ذو المصفوفة الكاملة للمعالم التمهيدية.

3. اظهرت النتائج عدم تاثر قيمة MSE باشارة الارتباط سواء أكان ارتباطا عكسيا ام طرديا.
4. يفضل استخدام المقدر اللامعلمي ذو المصفوفة القطرية للمعالم التمهيدية كبديل عن المقدر اللامعلمي ذو المصفوفة الكاملة للمعالم التمهيدية كون النتائج اوضحت تفوق المقدر الاول في كثير من التجارب المقامة فضلا عن سهولة استخدام المصفوفة القطرية من حيث التقديرات للمعالم التمهيدية.

- **لحالة التوزيع الملوث:**

1. اثبتت النتائج تساوي اداء المقدرين اللامعلمين المستخدمين.
2. في حالة الارتباطات الصغيرة والمتوسطة يفضل استخدام المقدرات اللامعلمية، في حين عند استخدام الارتباطات العالية يفضل استخدام مقدر الامكان على الرغم مما يعانيه هذا المقدر من سوء اداء في الاذيال وكما يلاحظ من الاشكال المعروضة انفا.

## 8- المصادر


[1] Ahmed, E, Ramani, D., David H. and Larry S. D. (2002). "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance" IEEE, Vol. 90, No. 7, July 2002.

[2] Bowman, A.W. and Azzalini, A. (1997). "Applied smoothing techniques for data analysis, the kernel approach with S-plus illustrations" Oxford science publications.

[3] Duong T. (2004). "Bandwidth selection for multivariate kernel density estimation" University of Western Australia.

[4] Duong T. (2006). "Feature significance for multivariate kernel density estimation" Statistics seminar series, University of New South Wales, Session 1.

[5] Hardle, W. and Mueller, M. (2000). "Multivariate and semi parametric kernel regression" Smoothing and regression: approaches and applications" pp357-393.

[6] Mittal, A. and Paragios, N. (2004). "Motion-Based background subtraction using adaptive kernel density estimation. Through internet.

[7] Mueller,M. "Semiparametric extensions to generalized linear models" (2000), Fakultat der Humboldt-Universitat zu Berlin.

[8] Powel, J. L. (2005). "Notes on non-parametric density estimation" Department of Economics-University of California, Berkeley.

[9] Rencher, A. (1995). "Methods of multivariate analysis" John Wiley and Sons, INC.

[10] Tavakkoli A., Nicolescu M. and Bebis G. (2005). "Automatic robust modeling using multivariate nonparametric kernel density estimation for visual surveillance" ISVC (2005), LNCS 3804, pp363-370.

[11] Wand, M.P and Jones, M.C. (1995). "Kernel Smoothing" Chapman and Hall/ CRC.



**Abstract**:

In this paper we demonstrate three methods that deal with multivariate probability density estimation. We concentrate on Bivariate normal and contamination distributions. The comparisons were by using of simulation, where we compare three estimators (Maximum likelihood, Kernel with full bandwidth matrix and Kernel with diagonal bandwidth matrix).