École polytechnique de Montréal
Département de mathématiques et de génie industriel

MTH2302C – Probabilités et statistiques

Examen final – Hiver 2006

Vendredi, le 21 avril 2006 de 9h30 à 12h00

QUESTION 1 (5 points)

Dans une certaine région, la magnitude M des tremblements de terre suit une loi exponentielle. La magnitude moyenne est 2 (échelle de Richter).

a) Quelle est la probabilité qu'un tremblement de terre donné montre M>7 ?

L'intervalle de temps T entre deux tremblements de terre (quelle que soit leur magnitude) suit une loi exponentielle de moyenne 100 jours. T et M sont considérées indépendantes.

b) Combien de tremblements de terre (peu importe M) devrait-on observer en moyenne au cours d'une année de 365 jours ?

c) Combien de tremblements de terre ayant M>7 s'attend-on à observer au cours d'une année de 365 jours?

d) Quelle est la probabilité qu'il n'y ait aucun tremblement de terre ayant M>7 sur une période de 10 ans?

QUESTION 2 (5 points)

Un pont est dit congestionné lorsque sa traversée en véhicule excède 20 minutes. Soit les événements :

A : le pont Champlain est congestionné;
B : le pont Victoria est congestionné;
C : le pont Jacques-Cartier est congestionné.

On trouve : $P(A)=0{,}65$
$P(B)=P(C)=0{,}55$
$P(A\cap B)= P(A\cap C)= P(B\cap C)=0{,}40$
$P(A\cap B\cap \overline{C})=0{,}1$

a) Les événements A, B et C sont-ils indépendants? Justifiez.

Pour les questions b), c) et d) vous pouvez répondre en utilisant les diagrammes de Venn.

b) Quelle est la probabilité que les trois ponts soient simultanément congestionnés ?

c) Quelle est la probabilité qu'aucun pont ne soit congestionné?

d) Sachant que le pont Champlain est congestionné, quelle est la probabilité conditionnelle qu'au moins un autre pont soit congestionné?

QUESTION 3 (5 points)

Le réservoir d'un barrage a une capacité résiduelle, X_1, estimée suivre une N(1,0.1). Une pluie importante s'amorce et l'on estime que l'apport d'eau au réservoir qui en résultera pour la prochaine semaine, X_2, est distribué suivant une N(1.5,0.3). L'eau qui sera turbinée durant cette semaine, X_3, suit une N(0.8,0.04). X_1, X_2 et X_3 sont exprimés en millions de m^3.
(Précision : le volume d'eau qui devra être évacué durant cette semaine sans être turbiné est donné par : $Y=X_2-X_1-X_3$).

a) Quelle est la probabilité que l'on doive évacuer de l'eau par l'évacuateur de crue sans pouvoir la turbiner ?

b) Quelle est la probabilité que la capacité résiduelle du réservoir ait augmenté à la fin de la semaine (et donc que le niveau du réservoir ait diminué)?

QUESTION 4 (5 points)

On mesure durant 11 jours le débit moyen quotidien de deux rivières, A et B. Les statistiques suivantes sont obtenues :

$$\bar{x}_A = 350\ m^3/s; \quad \bar{x}_B = 320\ m^3/s\ ;\ s_A^2 = 900\ (m^3/s)^2\ ;\ s_B^2 = 800\ (m^3/s)^2.$$

a) Faites le test (bilatéral, niveau 10%) pour vérifier si l'on peut considérer les variances égales.

b) Peut-on considérer que les deux moyennes sont égales ? Faites le test bilatéral au niveau 5% en supposant que les variances sont égales.

c) Donner l'intervalle de confiance (bilatéral 95%) sur la différence des moyennes des débits de ces deux rivières (toujours en supposant que les variances sont égales).

QUESTION 5 (8 points)

Pour effectuer la correction géométrique d'une photo aérienne, on dispose de 26 points de contrôle. On s'intéresse à la prédiction de la coordonnée « y » sur la photo déformée à partir des coordonnées (u,v) du référentiel. On a $s_y^2 = \frac{1}{25}\sum_{i=1}^{26}(y_i - \bar{y})^2 = 100$. Le tableau suivant donne les R^2 obtenus avec différents modèles.

Modèle	Description du modèle estimé	R^2
A	$Y=b_0+b_1u+b_2v+e$	0.9
B	$Y=b_0+b_1u+b_2v+b_3u^2+b_4uv+b_5v^2+e$	0.98
C	$Y=b_0+b_1u+b_2v+b_3u^2+b_4uv+b_5v^2+b_6u^3+b_7u^2v+b_8uv^2+b_9v^3+e$	0.985

a) Le modèle A est-il significatif (i.e. rejette-t-on $H_0: \beta_1 = \beta_2 = 0$*). Faites le test requis au niveau 5%.*

b) Utilisant le test d'ajout (niveau 5%), vérifiez quel modèle semble le plus approprié.

c) Dans une régression, est-il possible que tous les coefficients soient individuellement significatifs mais conjointement non-significatifs ? Expliquez au moyen d'un croquis.

d) Quel est le nombre maximum de coefficients, incluant la constante b_0*, que l'équation de régression peut contenir dans ce problème?Que se passe-t-il pour le* R^2 *si l'on atteint ce nombre maximum ?*

QUESTION 6 (7 points)

Vous disposez de mesures de compaction sur des sols ayant été compactés par deux appareils différents. Les deux seules variables disponibles dans cette étude sont le taux de compaction et la variable indicatrice de l'appareil utilisé (1 : appareil A; 0 : appareil B)

a) Expliquez comment vous pourriez tester l'égalité des moyennes des taux de compaction par les deux appareils à l'aide d'un simple programme de régression en vous servant du test d'ajout. Indiquez clairement quels seraient les modèles réduit et complet dans ce cas. (Note : on peut montrer que le résultat de ce test est identique au test « t » classique d'égalité des moyennes)

b) Proposez une extension à la procédure décrite en a) pour le cas où l'on voudrait tester l'égalité des moyennes de compaction obtenues par trois appareils différents. À nouveau indiquez clairement ce que seraient les modèles réduit et complet. (Note : ceci permet de généraliser facilement le test « t » d'égalité de deux moyennes à un test d'égalité de « p>2 » moyennes).

QUESTION 7 (8 points)

Un site contaminé au Pb montre un modèle de variogramme comprenant 3 composantes différentes :

Composante	C (ppm^2)	Portée (m)
Effet pépite	1000	-
Sphérique 1	5000	isotrope a= 1000
Sphérique 2	7000	$a_{50} = 2000$; $a_{140} = 600$

Toutes les directions sont en azimut (i.e. 0^o vers le Nord, croissant dans le sens horaire)

a) Quel est l'écart-type de la v.a. Pb selon ce modèle?

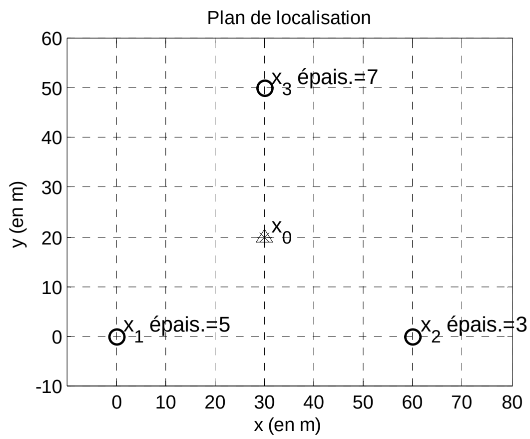
b) Si l'effet de pépite représentait uniquement une erreur de mesure, quelle serait la variance de la variable Pb observée sans erreur?

c) Selon l'azimut 30°, à quelle distance la covariance devient-elle nulle?

d) Quelle est la covariance entre les teneurs en Pb pour deux points espacés d'une distance 0+ (i.e. pas exactement 0)?

<u>QUESTION 8</u> (7 points)

L'épaisseur (en m) d'un dépôt de sable présente un variogramme sphérique isotrope avec a=100m, C_0=1m^2 et C=10m^2. La figure suivante montre les épaisseurs observées en 3 points échantillons, on désire fournir une estimation par krigeage ordinaire de l'épaisseur de sable au point x_0.



Voici quelques valeurs du terme sphérique : $(1.5h/100 - 0.5\ (h/100)^3)$ apparaissant dans la définition du variogramme sphérique et susceptibles de vous sauver quelques calculs :

h	30	$\sqrt{1300}$	$\sqrt{3400}$	60
$(1.5h/100 - 0.5\ (h/100)^3)$	0,437	0,517	0,776	0,792

a) Donnez le système de krigeage ordinaire sous forme matricielle (ne pas résoudre)

b) Que devient le système de krigeage si l'effet de pépite du modèle est 2m^2 plutôt que 1m^2?

c) Dans le cas décrit en b) doit-on s'attendre à une diminution de la variance de krigeage ?

Corrigé

Corrigé

1- a) E[M]=2=1/λ =>On a $\lambda = 1/2$. P(M>7)=exp(-7/2)=0,0302

b) Le nombre sur 100 jours suit une loi Poisson avec λ=1. Sur 365 jours ce sera Poisson(3,65). Le nombre moyen de tremblements de terre est égal à λ , donc 3,65.

c) Le nombre de tremblements de terre annuellement avec M>7 sera 3,65*P(M>7)=0,110

d) Soit X : le nombre de tremblements de terre avec M>7 sur 10 ans. X est Poisson (0,11*10). P(X=0)=exp(-1,1)=33.3%

2- a) Non car $P(A \cap B) \neq P(A)*P(B)$; $P(B \cap C) \neq P(B)*P(C)$; $P(A \cap C) \neq P(A)*P(C)$.

b) $P(A \cap B \cap C) = P(A \cap B) - P(A \cap B \cap \overline{C})$=0,4-0,1=0,3

c) 1-0,85=0,15

d) $P(B \cup C|A) = P((B \cup C) \cap A)/P(A)$=0,5/0,65=0,769

3- a) P(Y>0) Y~N(1.5-1-0.8, 0.1+0.3+0.04)=N(-0.3,0.44)
P(Y>0)=P(Z>0.3/$0.44^{0.5}$)=P(Z>0.452)=0,326
b) On cherche P(X_3>X_2) =P(X_3-X_2>0) X_3-X_2~N(-0.7,0.34)
P(X_3-X_2>0)=P(Z>0.7/$0.34^{0.5}$)= P(Z>1.2)=0,115

4- a) $F_{calculé=}$ 900/800 < $F_{10,10,5\%}$=2,98. On accepte donc que les variances sont égales.
b) On calcule s_p^2=(9000+8000)/20=850
$t_{calculé}$= $\dfrac{(350 - 320)}{s_p(1/11+1/11)^{0.5}} = 2{,}413$ comparé à une $t_{20;\ 5\%}$ =2,086. On rejette donc H_0, i.e. les moyennes sont significativement différentes.

c) L'intervalle de confiance de niveau 95% est 30 $\pm$ 2,086 $(850*2/11)^{0.5}$=[4.067, 55.932]

---

5 a) SCT_m=25*100=2500
SCR_m=0.9*2500=2250
SCE=250
$F_{calculé}$= $\dfrac{2250/2}{250/23} = 103.5$ > $F_{2,23,\ 5\%}$=3.42 On rejette H_0 donc le modèle A est très significatif.

b) B vs A : $\dfrac{(0.98 - 0.9)*2500/3}{50/20} = 26.67$ > $F_{3,20,\ 5\%}$=3.098 Le modèle B procure un ajout significatif.

C vs B : $\dfrac{(0.985 - 0.98) * 2500/4}{37.5/16} = 1.333 < F_{4,16,5\%}=3.007$ Le modèle C n'ajoute rien de significatif par rapport au modèle B.

c) Oui, l'intervalle de confiance dessine une ellipse ou un ellipsoïde. Il est possible que le point (0,0) soit dans l'ellipse (donc non-significatif) et en même temps hors de chacun des intervalles individuels sur chaque coefficient pris séparément. Un tel exemple a été vu en classe.

d) Nombre maximum de paramètres=26. Le R^2 est alors 1 pour autant que X'X soit inversible.

6- a) Modèle réduit : $Y=b_0+e$
Modèle complet : $Y=b_0+b_1I+e$

b) Modèle réduit : $Y=b_0+e$
Modèle complet : $Y=b_0+b_1I_1+b_2I_2+e$ où I_1 vaut 1 pour l'appareil A, I_2 vaut 1 pour l'appareil B. L'appareil C correspond au cas $I_1=I_2=0$.

7- a) $(1000+5000+7000)^{0.5}=114{,}02$
b) 5000+7000=12000
c) L'angle entre l'azimut 30^o et la direction de a_g est 20^o. Pour la composante anisotrope, la portée dans l'azimut 30^o est : $2000*600/(600^2 \cos^2(20)+2000^2\sin^2(20))=1354$ m. C'est à cette distance que les données ont une covariance nulle.

d) 12000

8- a)

$$\begin{array}{c|cccc|}
 & x1 & x2 & x3 & \\
x1 & 11 & 2.08 & 2.2448 & 1 \\
x2 & 2.08 & 11 & 2.2448 & 1 \\
x3 & 2.2448 & 2.2448 & 11 & 1 \\
 & 1 & 1 & 1 & 0
\end{array}
\begin{array}{c}
\lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \mu
\end{array}
=
\left|\begin{array}{c}
4.826 \\ 4.826 \\ 5.635 \\ 1
\end{array}\right|$$

b) On ajoute 1 aux 3 premières entrées sur la diagonale.
c) Non, la variance de krigeage va augmenter car on a une moins bonne précision sur les données.