250 - Loi des grands nombres. Théorème de la limite centrale. Applications.

I) <u>Préliminaires</u> [Ouv]

1) <u>Quelques inégalités</u>

Quelques inégalités (Markov, Tchebyshev, Kolmogorov) [Ouv 23, 24, 101] (ces inégalités sont de plus en plus puissantes.

Csq : CS de convergence ps [Ouv 103]

2) <u>Types de convergence</u>

Convergence ps, cv en proba, cv en loi, cv L^p, implications. Rappeler que cv en loi c'est pareil si on prend f bornée et continue, ou f continue qui tend vers 0 [Ouv 89]

3) <u>Sur la loi normale</u>

Densité, fct caractéristique

4) <u>Transformée de Fourier</u>

Th : formule d'inversion

Csq : Cc dense dans Im(TF)

II) <u>Lois des grands nombres</u> [Les] + [Ouv]

1) <u>Lois faibles</u> [Les] + [Ouv]

Th : loi faible des grands nombres version Bernoulli [Les]

Appl : polynômes de Bernstein, Weierstrass

Loi faible des grands nombres : Xn suite de va non corrélées, de carré intégrable. On suppose que la moyenne des espérances tend vers m et que 1/n²*sum(variance) tend vers 0. Alors Sn/n tend vers m en proba [Ouv 107] *(utilise Bienaymé Tchebychev)*

Théorème de Khintchine : c'est la loi faible pour des va intégrables deux à deux indépendantes de même loi de même moyenne m. Alors Sn/n cv vers m en proba [Ouv] *(on se ramène au cas précédent)*

2) <u>Lois fortes</u> [Ouv]

Th : Xn de carré intégrable et autres hypothèses [Ouv] (ici les va sont pas iid. lemme de Cesaro, théorème corollaire de l'inégalité de Kolmogorov, lemme de Kronecker)

Th : Kolmogorov-Khintchine. Celui que je connais (va iid L^1) [Ouv] *(on se sert du cas précédent)*

Appl : méthode de Monte Carlo [Ouv 128] *(on se donne une suite de va uniformes sur [0,1] (Un), on définit Xn=1_D.f o U_n. Alors la moyenne des X_n cv ps vers l'intégrale de f par la loi forte des GN, et on a une majoration de l'erreur par l'inégalité de Tchebitchev)*

III) <u>Théorème central limite</u> [Les] + [Ouv] + [Carr] + [GL]

1) <u>Enoncés</u> [Ouv] + [Les]

Moivre Laplace

Appl : lancer de pièces [Les]

Th de Lévy

TCL général

2) TCL Poissonien [Carr]

Th : Le Cam [Carr 78]

Cor : TCL Poissonien [Carr]

Interprétation et exemple : si des évènements ont une faible proba d'arrivée (ex : une faute dans une page d'un livre, proba 0,01) et que n est grand (500 pages), le nb effectif d'évènements qui arrivent suit une loi de Poisson (il y a donc P(500*0.01)=P(5) fautes dans le livre en moyenne).

3) Applications [GL] + [Carr]

Formule de Stirling [GL]

Calcul de somme [Carr 76] *(si les Xn suivent une loi de Poisson de param l, alors X1+…+xn suit une loi de Poisson de param nl. On prend l=1 et on applique le TCL, pour estimer la proba que ce soit plus petit que 0, ça nous donne une demi intégrale gaussienne dont on connait la valeur, mais nous on sait estimer cette proba exactement et c'est la série qu'il faut calculer)*

IV) Applications en statistiques [Stat] + [Ouv]

1) Estimateurs [Stat]

Déf : estimateur, consistant, fortement consistant

Ex : moyenne, fortement consistant grâce à la LFGN

Déf : biais

Exemple : moyenne, variance

2) Intervalles de confiance [Stat]

Principe : on a construit un estimateur, on arrive donc à estimer des quantités, mais avec quelle précision ? On aimerait pouvoir donner un intervalle autour de notre estimation tel qu'on soit sûr à 95% que la vraie valeur est dedans.

Exemple : sondage pour intentions de vote, candidats A, B. Xn une réalisation. On veut estimer l'espérance du candidat A, qu'on note esp, et l'estimation donne ESP. Asymptotiquement, sqrt(n)*(ESP-esp)/sigma a une loi N(0,1). Du coup, P(sqrt(n)*(X1+…+Xn)/sigma < 1,96) = 95%. Mais on a un problème parce que sigma dépend du paramètre qu'on veut estimer (et qui est inconnu). Du coup on va estimer sigma par un estimateur consistant SIGMA et le lemme de Slutsky va nous dire que asymptotiquement, sqrt(n)*(ESP-esp)/SIGMA a une loi N(0,1). Donc P(sqrt(n)*(ESP-esp)/SIGMA < 1,96) = 95%, ce qui nous donne un intervalle de confiance.

3) Test du Chi-deux [Stat]

Déf : statistique de Karl Pearson [Stat]

Prop : si notre hypothèse est juste, la distance va converger vers un Chi deux à d-1 degrés de liberté, sinon vers l'infini [Stat 54] *(par le TCL on trouve que ça suit une loi normale centrée avec une matrice de covariance qui a un rapport avec la projection orthogonale, et là on applique le th de Cochran qui dit que si on projette un vecteur gaussien de façon orthogonale sur un espace de dimension d, alors la projection suit une loi du Chi deux à d degrés de liberté et à décentrage dépendant de la moyenne du vecteur gaussien)*

Appl : on a un échantillon, on veut voir s'il colle à une loi. on le range en n classes, on calcule la distance de Karl-Pearson. Par exemple, n=6. LE chi deux est à 5 degrés de liberté. Dans le tableau, on voit que la proba que la distance de KP soit plus grande que 11,07 est égale à 0,05. Du coup si notre distance calculée est plus petite que 11,07, on peut valider l'hypothèse et on a 5% de chances de se tromper.

4) Test de Kolmogorov- Smirnov [Ouv]

Th : théorème de Glivenko Cantelli, qui dit que les fonctions de répartition empiriques convergent uniformément vers les fonctions de répartition théoriques [Ouv] *(appl de la LFGN)*

Cor : on définit Dn=sup(|Fn-F|). Alors Dn tend vers 0 de façon indépendante à la loi de X [Ouv 119]

Appl : test de Kolmogorov-Smirnov. On a un échantillon de taille n, on veut voir s'il suit une certain loi. On calcule Dn. On se fixe un niveau de confiance à atteindre, par exemple 95%. Mettons n=20. On sait alors que P(Dn<0.29)=0.95. On calcule Dn avec les données de l'échantillon. Si Dn est plus petit que 0,29, alors on dit que l'échantillon suit la loi de l'hypothèse, et sinon, on dit que non. On a 5% de chances de se tromper [Ouv]

Remarques :
- il y a d'autres moyens de construire des estimateurs (maximum de vraisemblance).
- on peut construire des intervalles de confiance non asymptotiques
- le test du chi-deux peut servir à autre chose : ajustement à lui loi donnée, ajustement à une famille paramétrée de lois, test d'indépendance, test d'homogénéité (voir si deux échantillons peuvent provenir de la même loi).
- le teste de KS sert aussi à plein de trucs : ajustement à une loi donnée, à une loi parmi les exponentielles et les normales, test d'homogénéité.

Développements :
1 - TCL + Lévy [ZQ 536] (**)
2 - Glivenko Cantelli [Nourd 109] (**)
Polynômes de Bernstein [ZQ] (* ou **)

Bibliographie :
[Ouv] Ouvrard – Probabilités 2
[Les] Lesigne – Pile ou face
[Car] Carrieu
[GL] Girardin & Limnios
[Stat] Rivoirard & Stoltz – Statistiques en action
[ZQ]