Probabilités et Statistiques

Année 2010/2011

laurent.carraro@telecom-st-etienne.fr

olivier.roustant@emse.fr

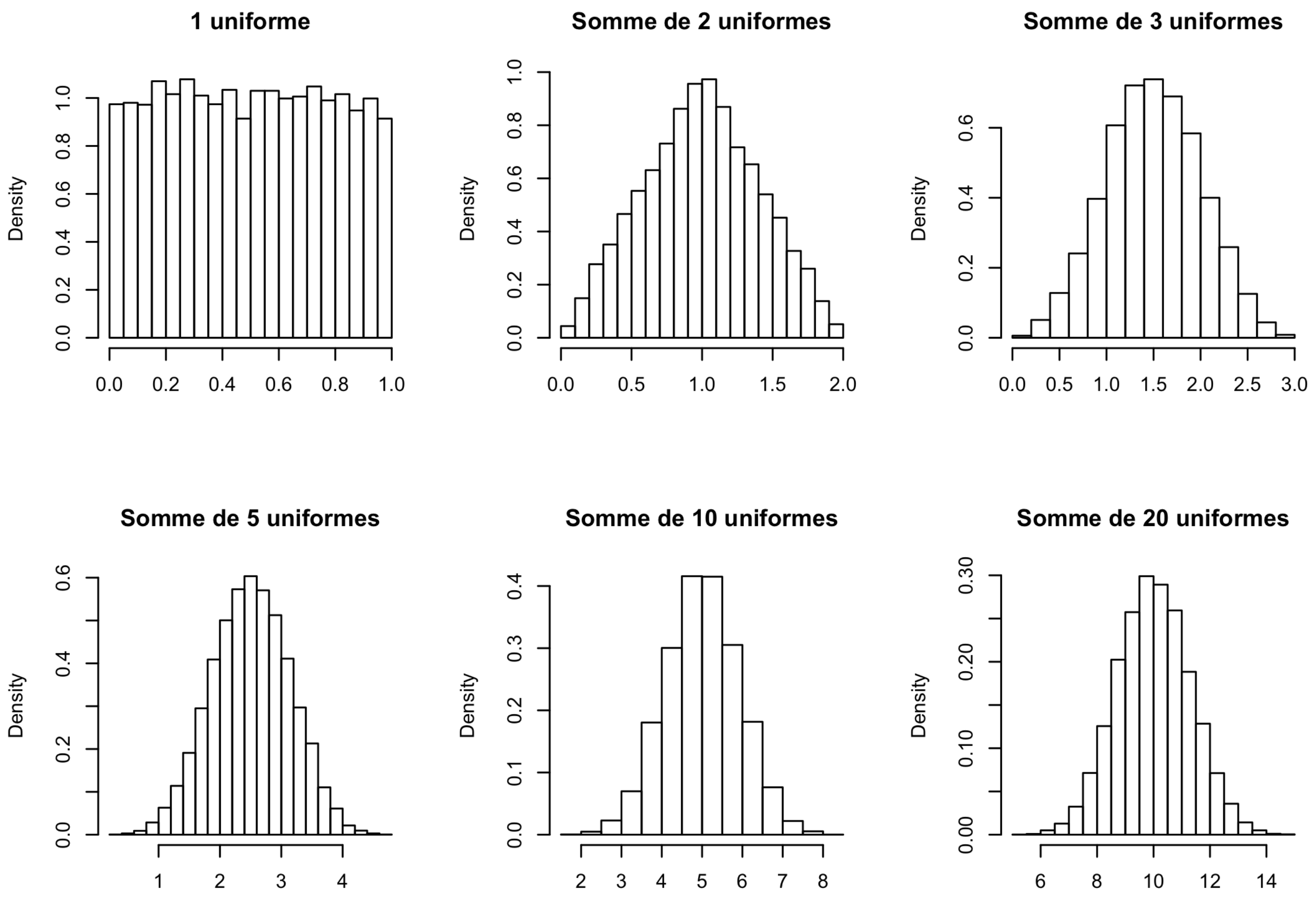
Cours n°10

Convergence vers la loi normale

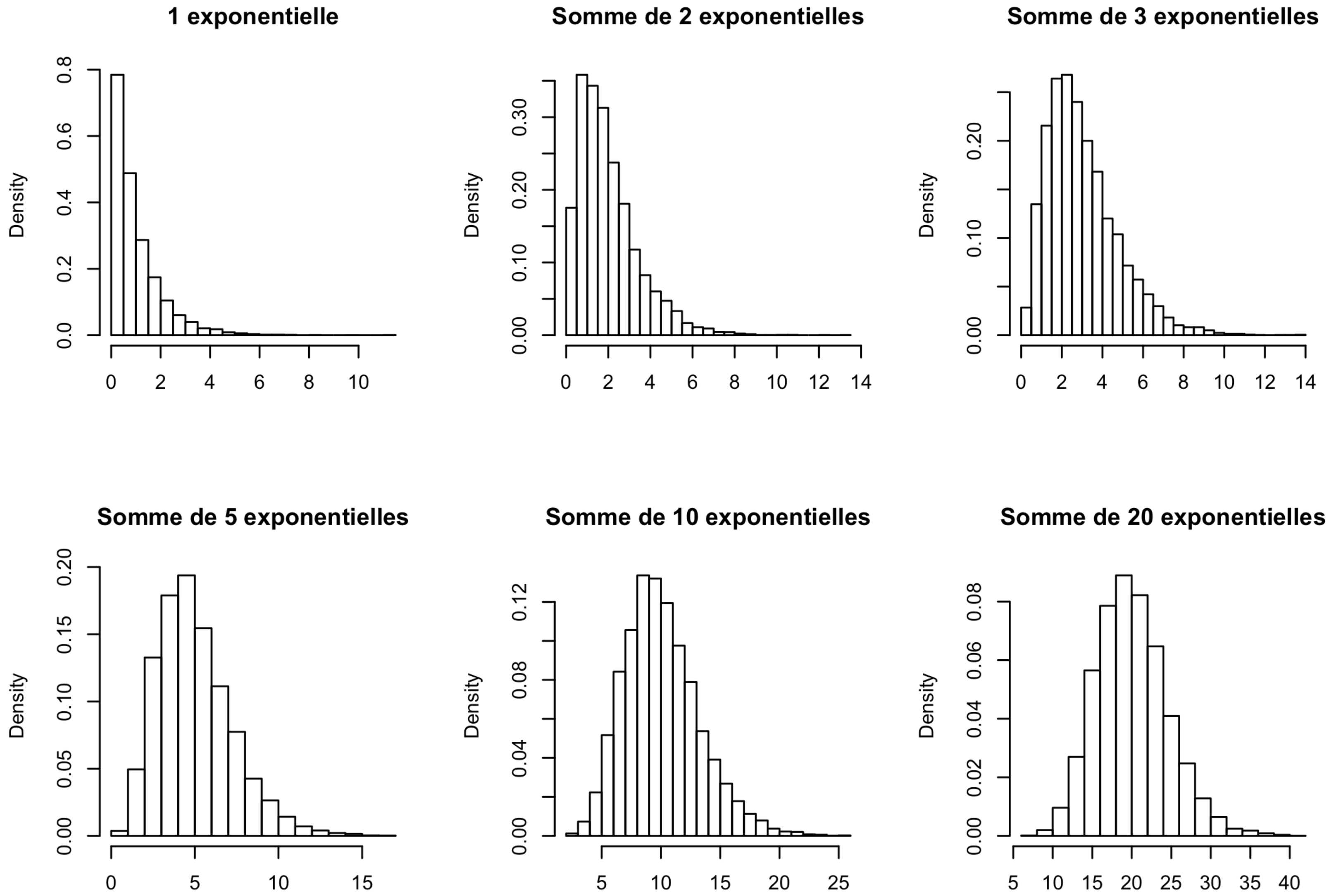
Retour sur Monte Carlo (calcul d'intervalles de confiance)

Convergence d'une somme de loi B(p/n)

Loi d’une somme de v.a. i.i.d ?



Loi d’une somme de v.a. i.i.d ?



Théorème de la limite centrée

- THEOREME. [de la limite centrée]
  Soient $X_1, \ldots, X_n$ des v.a. i.i.d. On ne précise pas la loi.
  On suppose seulement que var(X_i) existe.
  Soit $\mu := E(X_1)$, $\sigma^2 := Var(X_1)$ et $M_n := (X_1 + \ldots + X_n)/n$

  Alors pour n grand, M_n est approx. de loi $N(\mu, \sigma^2/n)$

Commentaires

- Remarquer la <u>généralité</u> des hypothèses sur la loi des X_i !
  - On peut même alléger l'hypothèse d'indépendance…

- *Formulation rigoureuse avec la convergence en loi :*

  Z_n -> Z en loi ssi $F_{Zn}(x) \rightarrow F_Z(x)$

  en tout point x pour lequel F_Z est continue.

- <u>Contre-exemple si var(X_i) n'existe pas</u>.
  - Loi de Cauchy : $f(t) = (1/\pi)\times 1/(1+t^2)$
  - *Proposition : si les X_i sont i.i.d. et de loi de Cauchy, alors pour tout n, M_n est de loi de Cauchy*.

Historique

- De Moivre (1733) : cas de la loi B(p) avec p=1/2
- Laplace (1812) : cas de la loi B(p), p quelconque
- Lévy (≈1920) : cas général

Somme de v.a. – convolution

- Rappel : soient X et Y deux v.a. supposées indépendantes et continues. Alors X+Y admet la densité :

$$f_{X+Y}(s)=\int_{-\infty}^{+\infty} f_X(s-y)\,f_Y(y)\,dy=(f_X * f_Y)(s)$$

Fonction caractéristique

- Qui dit convolution dit transformée de Fourier !
En proba., on a la fonction caractéristique

$$\varphi_X(t) = E\left(e^{itX}\right)$$

- Si X est continue, on reconnaît la forme de la transformée de Fourier :

$$\varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx} f_X(t)\, dt$$

Fonction caractéristique

Propriétés fondamentales :

- Si X et Y sont indépendantes, $\varphi_{X+Y} = \varphi_X \varphi_Y$
  -> Transforme les convolutions en produit

- X et Y ont même loi ssi $\varphi_X = \varphi_Y$
  -> La fonction caractéristique caractérise la loi

Théorème (Paul Lévy)

- Pour montrer que $X_n \to X$ en loi,
  il suffit de montrer que $\varphi_{X_n}(.) \to \varphi_X(.)$

Schéma de preuve du TLC

Variables aléatoires		Fonctions
$Z_n := \dfrac{\dfrac{X_1 + \ldots + X_n}{n} - \mu}{\sigma/\sqrt{n}}$	$\Longleftrightarrow$	$\varphi_{Z_n}(t)$
convergence en loi $P(Z_n < .) \to P(Z < .)$ $\downarrow$		Convergence simple $\varphi_{Z_n}(.) \to \varphi_Z(.)$ $\downarrow$
Z de loi N(0,1)	$\Longleftrightarrow$	$\varphi_{N(0,1)}(t)$

Démo (sur un exemple)

- Exercice : montrer que $\varphi_{N(0,1)}(t) = e^{-t^2/2}$
  - Indication : montrer que $\varphi_{N(0,1)}$ est solution de l'équa. diff.
    $z'(t) = -t\,z(t)$ avec $z(0) = 1$

- Soient $X_1, \ldots, X_n$ i.i.d de loi $U([-a,a])$, avec $a=\sqrt{3}$
  - Remarque : a est choisi pour que $var(X_i)=1$

Montrer que

$$\varphi_{Z_n}(t) \to \varphi_{N(0,1)}(t)$$

- Démo générale

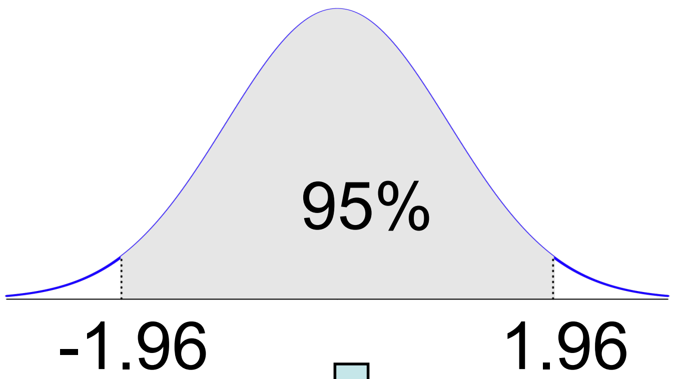
Application: intervalles de confiance

Loi normale N(μ,σ²/n)

$$\overline{X} := \frac{X_1 + \ldots + X_n}{n} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

Loi normale N(0,1)

$$Z := \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \approx N(0,1)$$



$$P(-2 < Z < 2) \approx 95\%$$

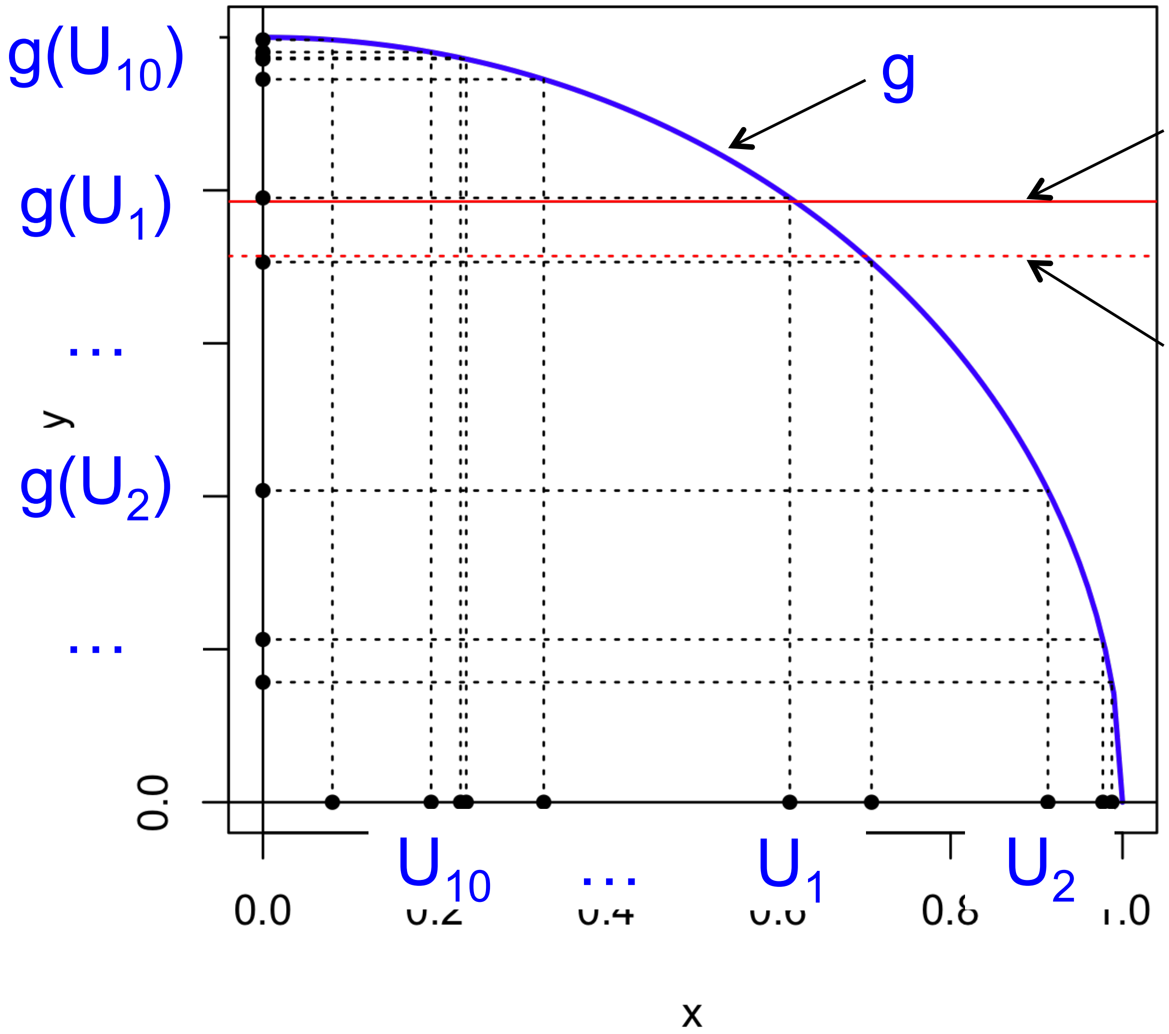
En pratique, on remplace σ par l'écart-type des X_i

$$P\left(\overline{X} - 2\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 2\frac{\sigma}{\sqrt{n}}\right) \approx 95\%$$

Méthode de Monte Carlo (MC)

- Nom donné au calcul de quantités probabilistes par simulation
  - Probabilité, espérance, variance, quantile, …
- Base théorique : loi des grands nombres
- Exemple. Calcul de pi (voir TD)
  - En utilisant un disque
  - En utilisant la fonction g : $x \to (1-x^2)^{1/2}$

Calcul d'intégrale par Monte Carlo

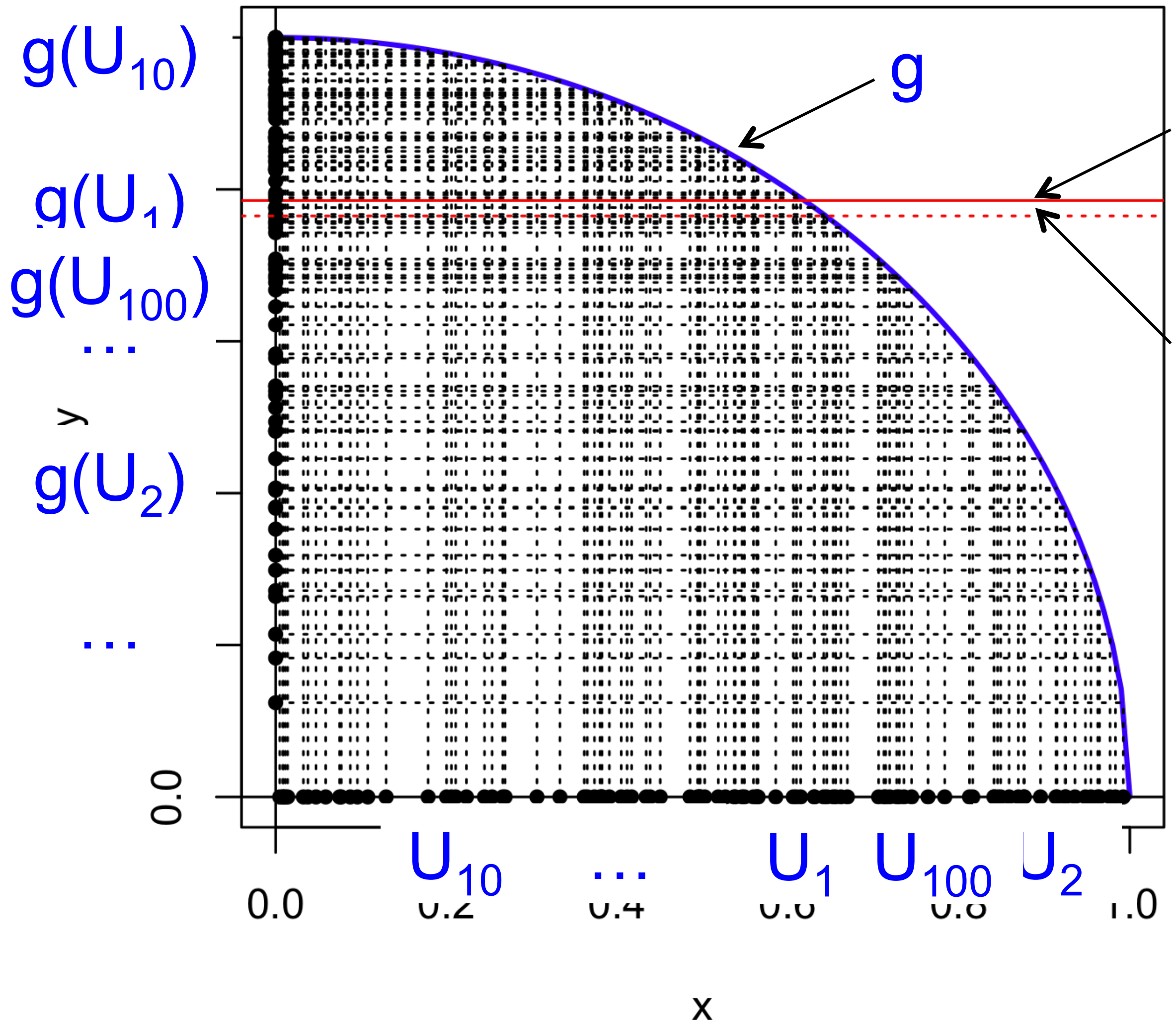


$$\int g(u)\,du$$

$$\frac{g(U_1)+\ldots+g(U_n)}{n}$$

$$n = 10$$

Calcul d'intégrale par Monte Carlo



$$\int g(u)\,du$$

$$\frac{g(U_1)+\ldots+g(U_n)}{n}$$

$$n = 100$$

Calcul d'intégrale par Monte Carlo

- Remarque de base : si U est uniforme sur [0,1],

$$\int g(u)du = E(g(U))$$

- Loi des grands nombres :

$$M_n = (\ g(U_1)+\ldots+g(U_n)\ )\ /\ n \ \ \text{->} \int g(u)du$$

avec $U_1,\ldots,\ U_n$ i.i.d. de loi unif. sur [0,1]

Calcul d'intégrale par Monte Carlo

- Variance d'une somme et vitesse de convergence

Ici $X_i = g(U_i)$

$var(M_n) = \sigma^2 / n$ avec $\sigma^2 = var(g(U_i))$, donc

$$\sigma(M_n) = \sigma(g(U)) / \sqrt{n}$$

- TLC et intervalle de confiance

$$\left[\overline{X} - 2\frac{\hat{\sigma}_X}{\sqrt{n}} ; \overline{X} + 2\frac{\hat{\sigma}_X}{\sqrt{n}}\right]$$

est un int. de conf. approché à 95% de $\int g(u)du$

Calcul d'intégrale par Monte Carlo

Commentaires

- Une partie en $1/\sqrt{n}$
- Une partie qui dépend de g -> on peut réduire la variance en changeant g (de sorte que l'intégrale reste la même)
  - Voir TD 9 et TD 10

Convergence de la loi B(p/n)

- Problème : loi du nb de spams reçus sur une période T ?

- Une solution
  - On découpe [0, T] en n intervalles
  - Sur chaque intervalle $I_k := [(k-1)T/n,\ kT/n]$, on suppose :
    - Qu'il y a au plus 1 spam
    - Que la probabilité de recevoir un spam est proportionnelle à la durée T/n
    - Que les spams sont reçus de façon indépendantes
  - Loi du nb de spams à n fixé ? Quand n -> + ∞ ?

Schéma de raisonnement

Variables aléatoires		Fonctions
$Z_n := X_1 + \ldots + X_n$ avec $X_i \sim B(p/n)$	$\Longleftrightarrow$	$\varphi_{Z_n}(t)$
$P(Z_n < .) \to P(Z < .)$ $\downarrow$		$\downarrow$ Convergence simple $\varphi_{Z_n}(.) \to \varphi_Z(.)$
Z de loi LOI (à trouver)	$\Longleftrightarrow$	$\varphi(t) = \varphi_{LOI}(t)$