Approximation polynomiale de la densité de probabilité

Applications en assurance

P.O. Goffard

Axa France - Institut de Mathématiques de Marseille I2M
Aix-Marseille Université

Soutenance de thèse de doctorat

Contenu de la thèse

Chap. 1: Motivations et applications des méthodes numériques en assurance.

Chap. 2: Description de la méthode d'approximation polynomiale.

Chap. 3: Application de la méthode d'approximation polynomiale à l'évaluation des probabilités de ruine ultime dans le modèle de ruine de Poisson composé.

Chap. 4: Application de la méthode d'approximation polynomiale aux modèles collectifs multivariés: Exemple d'application en réassurance.

Chap. 5: Optimisation de l'agrégation des portefeuille de contrats d'assurance vie individuel de type épargne.

Les engagements de l'assureur: Le modèle collectif

Soit un portefeuille de contrats d'assurance non-vie.

- Sur une période d'exercice donnée,
  - $\hookrightarrow$ Le nombre de sinistres est modélisé par une variable aléatoire de comptage N,
  - $\hookrightarrow$ Les montants indemnisés forment une suite de variables aléatoires positives, indépendantes et identiquement distribuées $(U_i)_{i\in\mathbb{N}}$.
- Les engagements de l'assureur sont modélisés par

$$X = \sum_{i=1}^{N} U_i,$$

  - $\hookrightarrow$ la suite $\{U_i\}_{i\in\mathbb{N}}$ est indépendante de N.

X admet une distribution de probabilité composée.

La dépendance des risques: Le modèle collectif multivarié

Soit n portefeuilles de contrats d'assurance non vie,

- Les risques sont modélisés **conjointement** via

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{N_1} U_{1j} \\ \vdots \\ \sum_{j=1}^{N_n} U_{nj} \end{pmatrix} + \sum_{j=1}^{M} \begin{pmatrix} V_{1j} \\ \vdots \\ V_{nj} \end{pmatrix},$$

$\hookrightarrow$ $\mathbf{N} = (N_1, \dots, N_n)$ est un vecteur composé de variables aléatoire de comptage,

$\hookrightarrow$ $(U_{1j})_{j\in\mathbb{N}}, \dots, (U_{nj})_{j\in\mathbb{N}}$ sont des suites de variables aléatoires positives, et **i.i.d.**.

$\hookrightarrow$ $\{\mathbf{V}_i\}_{i\in\mathbb{N}} = \{(V_{1i}, \dots, V_{ni})\}_{i\in\mathbb{N}}$ est une suite de vecteur aléatoire **i.i.d.**,

$\hookrightarrow$ M est une variable aléatoire de comptage,

$\hookrightarrow$ $\{\mathbf{V}_i\}_{i\in\mathbb{N}}$, $\mathbf{N}$, M et $(U_{j1})_{j\in\mathbb{N}}, \dots, (U_{jn})_{j\in\mathbb{N}}$, sont indépendants.

La théorie de la ruine: Une modèlisation dynamique

Soit un portefeuille de contrats d'assurance non-vie.

- soit un instant $t \geq 0$,
  - $\hookrightarrow$ Le nombre de sinistres est modélisé par un processus stochastique de comptage $\{N_t\}_{t\geq 0}$,
  - $\hookrightarrow$ Les montants indemnisés forment une suite de variables aléatoires positives, indépendantes et identiquement distribuées $(U_i)_{i\in\mathbb{N}}$.
- Les engagements de l'assureur à l'instant t sont égaux à

$$X_t = \sum_{i=1}^{N_t} U_i,$$

  - $\hookrightarrow$ la suite $(U_i)_{i\in\mathbb{N}}$ est indépendante du processus $\{N_t\}_{t\geq 0}$.

La théorie de la ruine: Une modèlisation dynamique

Soit un portefeuille de contrats d'assurance non-vie.

- La réserve financière allouée au portefeuille de contrats est donnée par

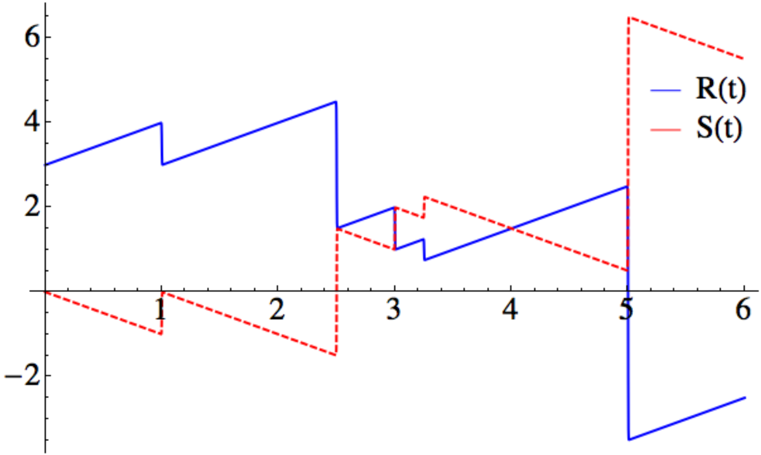
$$R_t = u + ct - \sum_{i=1}^{N_t} U_i,$$

  - $\hookrightarrow$ **u** est la réserve initiale,
  - $\hookrightarrow$ **c** est le montant des primes reçues par unité de temps.

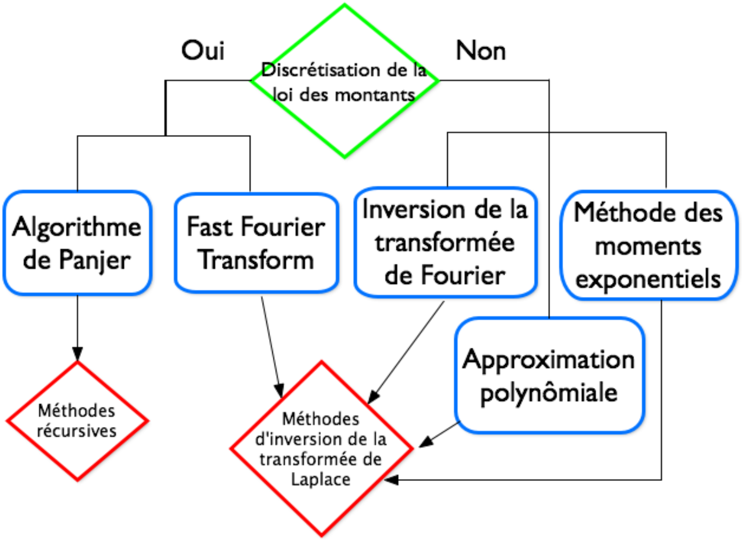
- Le processus d'excédent de sinistre est défini par

$$S_t = u - R_t.$$

La théorie de la ruine:
Une visualisation graphique



Méthodes numériques: Cartographie



EXECUTIVE SUMMARY

Application de la méthode d'approximation polynomiale à deux problèmes.

1. Évaluation de la probabilité de ruine ultime dans le modèle de ruine de Poisson composé.
   - $\hookrightarrow$ $\{N_t\}_{t\geq 0}$ est un processus de Poisson simple d'intensité λ.
2. Étude d'un modèle collectif bivarié et application dans un contexte de réassurance.

Définition de la probabilité de ruine

- Probabilité de ruine à horizon de temps infini,

$$\psi(u) = \mathbb{P}\left(\inf_{t\geq 0} R_t < 0; R_0 = u\right).$$

- Probabilité de ruine à horizon de temps fini,

$$\psi(u, T) = \mathbb{P}\left(\inf_{t\in[0,T]} R_t < 0; R_0 = u\right).$$

- Probabilité de non ruine à horizon de temps fini et infini ,

$$\phi(u) = 1 - \psi(u) \qquad \phi(u, T) = 1 - \phi(u, T).$$

Définition alternative de la probabilité de ruine

- Instant de ruine et maximum du processus de surplus,

$$\tau_u = \inf\{t \geq 0 : R_t < 0\} = \inf\{t \geq 0 : S_t > u\},$$

$$M = \sup_{t\geq 0} S_t, \qquad M_T = \sup_{t\in[0,T]} S_t.$$

- Probabilité de ruine à horizon de temps fini et infini,

$$\psi(u) = \mathbb{P}\left(\tau_u < \infty\right) = \mathbb{P}\left(M > u\right),$$

$$\psi(u, T) = \mathbb{P}\left(\tau_u < T\right) = \mathbb{P}\left(M_T > u\right).$$

Le chargement de sécurité

Le coût moyen des sinistres par unité de temps est

$$\frac{1}{t}\mathbb{E}\left(X_t\right) = \lambda\mathbb{E}(U).$$

- Le chargement de sécurité η est défini par,

$$c = (1 + \eta)\lambda\mathbb{E}(U).$$

- *Net Benefit Condition*

$$\eta > 0,$$

$\hookrightarrow$ Si $\eta < 0$ alors $\psi(u) = 1$,
$\hookrightarrow$ Si $\eta > 0$ alors $\psi(u) < 1$.

Probabilité de ruine ultime: La formule de Pollaczek-Khinchine

Dans le cadre du modèle de ruine de Poisson composé,

$$\psi(u) = \mathbb{P}(M > u) \qquad M \overset{D}{=} \sum_{i=1}^{N} V_i,$$

- N suit une loi géométrique de paramètre $p = \frac{\lambda E(U)}{c} < 1$,

$$\mathbb{P}(N = n) = (1 - p)p^n.$$

- $(V_i)_{i \in \mathbb{N}^*}$ est une suite de variables aléatoires positives, **i.i.d.** de densité,

$$f_V(x) = \frac{\mathbb{P}(U > x)}{\mathbb{E}(U)}.$$

- $(V_i)_{i \in \mathbb{N}^*}$ et N sont indépendantes.

Notations: Produit de convolution

- Soit U et V deux variables aléatoires indépendantes de densité f_U et f_V,

$$\begin{aligned} f_{U+V}(x) &= \int f_U(x-y) f_V(y) dy, \\ &= (f_U * f_V)(x). \end{aligned}$$

$\hookrightarrow$ Produit de convolution de f_U et f_V.

- Soit $S = \sum_{i=1}^{n} U_i$, la somme de n variables aléatoires **i.i.d.**,

$$\begin{aligned} f_S(x) &= \int \int \ldots \int f_U(x-y) f_U(y) dy, \\ &= f_U^{(*n)}(x). \end{aligned}$$

$\hookrightarrow$ f_U convoluée n fois avec elle même.

Notations: Transformée de Laplace

La transformée de Laplace d'une variable aléatoire est définie par,

$$\mathcal{L}_U(s) = E\left(e^{sU}\right) = \int e^{sx} f_U(x) d\lambda(x).$$

Ce qui donne pour la somme de variables aléatoires indépendantes:

$$\begin{aligned} \mathcal{L}_{U+V}(s) &= \mathcal{L}_U(s) \times \mathcal{L}_V(s), \\ \mathcal{L}_S(s) &= \left[\mathcal{L}_U(s)\right]^n. \end{aligned}$$

Distribution géométrique composée

- La variable aléatoire $M = \sum_{i=1}^{N} U_i$ admet une distribution composée,

$$d\mathbb{P}_M(x) = \mathbb{P}(N = 0)\delta_0(x) + d\mathbb{G}_M(x),$$

où

$$\begin{aligned} g_M(x) &= \sum_{n=1}^{+\infty} \mathbb{P}(N = n) f_V^{(*n)}(x) \\ &= \sum_{n=1}^{+\infty} (1-p)p^n \int\int\ldots\int f_V(x-y) f_V(y) dy, \end{aligned}$$

et

$$\begin{aligned} \psi(u) &= \mathbb{P}(M > u) \\ &= \int_u^{+\infty} \sum_{n=0}^{+\infty} (1-p)p^n \int\int\ldots\int f_V(x-y) f_V(y) dy dx. \end{aligned}$$

Transformée de Laplace de la probabilité de ruine

La transformée de Laplace de M est donnée par,

$$\mathcal{L}_M(s) = \mathcal{G}_N[\mathcal{L}_V(s)].$$

- $\mathcal{G}_N$ est la fonction génératrice des probabilités de N,

$$\mathcal{G}_N(s) = \mathbb{E}\left(s^N\right) = \sum_{k=0}^{+\infty}(1-p)p^k s^k$$

Ce qui implique que

$$\mathcal{L}_M(s) = \frac{1-p}{1-p\mathcal{L}_V(s)},$$

puis

$$\mathcal{L}_\psi(s) = \frac{1}{s}\left[1+\mathcal{L}_M(s)\right].$$

Méthode d'approximation polynomiale en dimension 1

Soit X une variable aléatoire de loi de probabilité $\mathbb{P}_X$, de densité f_X.

- ν est la mesure de probabilité de référence, de densité f_ν.
  - $\hookrightarrow$ La loi de X est absolument continue par rapport à la mesure de référence,
  $$f_{X,\nu}(x) = \frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}\nu}(x).$$
- $\{Q_k\}_{k\in\mathbb{N}}$ forme un système de polynômes orthonormaux par rapport à ν,

$$< Q_k, Q_l >= \int Q_k(x)Q_l(x)\mathrm{d}\nu(x) = \delta_{kl}, \quad k, l \in \mathbb{N}.$$

L'idée est de projeter $f_{X,\nu}$ sur la base de polynômes $\{Q_k\}_{k\in\mathbb{N}}$.

Méthode d'approximation polynomiale en dimension 1

- La mesure de probabilité $\mathbb{P}_{\mathbb{X}}$ est absolument continue par rapport à la mesure de référence ν,
  - $\hookrightarrow$ Existence de $\frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}\nu}$.
- L'ensemble de polynômes est dense dans l'ensemble $L^2(\nu)$,
  - $\hookrightarrow$ $\{Q_k\}_{k\in\mathbb{N}}$ forme une base de $L^2(\nu)$.

Si $\frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}\nu} \in L^2(\nu)$, alors

$$\frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}\nu}(x) = \sum_{k=0}^{+\infty} a_k Q_k(x), \quad x \in \mathbb{R},$$

avec

$$a_k = \mathbb{E}\left[Q_k(X)\right], \quad \forall k \in \mathbb{N}.$$

Méthode d'approximation polynomiale en dimension 1

- La densité de la variable aléatoire X admet la représentation polynomiale,

$$f_X(x) = f_{X,\nu}(x) f_\nu(x) = \sum_{k=0}^{+\infty} a_k Q_k(x) f_\nu(x).$$

- L'approximation est obtenue par troncature d'ordre K,

$$f_X^K(x) = f_{X,\nu}^K(x) f_\nu(x) = \sum_{k=0}^{K} a_k Q_k(x) f_\nu(x).$$

Les approximations de la fonction de répartition et de la fonction de survie s'obtiennent par intégration.

Approximation polynomiale de la probabilité de ruine

La mesure de probabilité associée à $M = \sum_{i=1}^{N} V_i$ s'écrit

$$d\mathbb{P}_M(x) = (1-p)\delta_0(x) + d\mathbb{G}_M(x).$$

Si $\frac{d\mathbb{G}_M}{d\nu} \in L^2(\nu)$ alors,

$$\frac{d\mathbb{G}_M}{d\nu}(x) = \sum_{k\in\mathbb{N}} < \frac{d\mathbb{G}_M}{d\nu}, Q_k > Q_k(x).$$

L'approximation de la probabilité de ruine est obtenue après troncature et intégration,

$$\psi^K(u) = \sum_{k=0}^{K} < \frac{d\mathbb{G}_M}{d\nu}, Q_k > \int_u^{+\infty} Q_k(y)d\nu(y).$$

Choix de la mesure de probabilité de référence

$\mathrm{d}\mathbb{G}_M$ est une mesure de probabilité défaillante de support $\mathbb{R}_+$.

- La loi gamma $\Gamma(m, r)$ admet un support sur $\mathbb{R}_+$, sa densité est donnée par
$$\mathrm{d}\nu(x) = \frac{x^{r-1}e^{-x/m}}{m^r\Gamma(r)}\mathrm{d}\lambda(x)$$
- Les polynômes orthogonaux par rapport à la mesure gamma sont les polynômes de Laguerre généralisés.

Le choix des paramètres m et r est très **important**.

Vérification de la condition d'intégrabilité

Théorème

Soit X variable aléatoire continue, et positive.

- **H1** Il existe $\gamma_X = \inf\{s > 0, \mathcal{L}_X(s) = +\infty\}$.
- **H2** Soit $a \geq 0$, l'application $x \mapsto f_X(x)$ est strictement décroissante pour $x \geq a$.

Alors, pour $x \geq a$,

$$f_X(x) < A(s_0)e^{-s_0 x}, \quad 0 < s_0 \leq \gamma_X.$$

Dans le cas de le probabilté de ruine, l'hypothèse **H2** est vérifiée, et $\gamma_M = \inf\{s > 0, \mathcal{L}_{g_M}(s) = +\infty\}$ est solution de l'équation

$$L_V(s) = p^{-1}.$$

Vérification de la condition d'intégrabilité

La condition d'intégrabilité s'écrit

$$\int \left[\frac{\mathrm{d}\mathbb{G}_M}{\mathrm{d}\nu}(x) \right]^2 \mathrm{d}\nu(x) < +\infty \Leftrightarrow \int_0^{+\infty} g_M^2(x) e^{x/m} x^{1-r} \mathrm{d}x < +\infty.$$

Par application de la majoration de g_M, pour $s_0 < \gamma_M$,

$$\int g_M^2(x) e^{x/m} x^{1-r} \mathrm{d}x < A(s_0) \int e^{-x\left(2s_0 - \frac{1}{m}\right)} x^{1-r} \mathrm{d}x.$$

On vérifie la condition d'intégrabilité avec

$$r \in (0,1], \qquad \frac{1}{m} \in (0, 2\gamma_M).$$

Décroissance des $\{a_k\}_{k\in\mathbb{N}}$: Premier résultat

La perte quadratique suite à l'approximation est donnée par

$$
\begin{aligned}
L\left(g_{M,\nu}, g_{M,\nu}^{K}\right) &= \int \left[g_{M,\nu}(x) - g_{M,\nu}^{K}(x)\right]^2 \mathrm{d}\nu(x) \\
&= \sum_{k=K+1}^{+\infty} a_k^2.
\end{aligned}
$$

$\Rightarrow$ Lien directe entre la décroissance et la qualité de l'approximation en fonction de l'ordre de troncature.

Proposition

H1 $x \mapsto g_{M,\nu}(x)$ continue, deux fois dérivable

H2 $g_{M,\nu}, g_{M,\nu}^{(1)}, g_{M,\nu}^{(2)} \in L^2(\nu)$

$$
a_k = o\left(\frac{1}{k}\right), \quad k \to +\infty.
$$

Décroissance des $\{a_k\}_{k\in\mathbb{N}}$: Etude de la fonction génératrice

La densité de probabilité defaillante de M admet la représentation polynomiale

$$g_M(x) = \sum_{k=0}^{+\infty} a_k Q_k(x) f_\nu(x).$$

En prenant la transformée de Laplace

$$L_{g_M}(s) = \left(\frac{1}{1-sm}\right)^r \mathcal{C}\left(\frac{sm}{1-sm}\right),$$

où $\mathcal{C}(z) = \sum_{k=0}^{+\infty} a_k c_k z^k$, avec

$$c_k = \sqrt{\binom{k+r-1}{k}}.$$

Décroissance des $\{a_k\}_{k\in\mathbb{N}}$: Etude de la fonction génératrice

La fonction génératrice des coefficients s'exprime en fonction de la transformée de Laplace, avec

$$\mathcal{C}(z) = (1+z)^{-r} L_{g_M}\left[\frac{z}{m(1+z)}\right].$$

- Obtention des coefficients par dérivation et évaluation en 0,

$$a_k = \frac{1}{c_k k!}\left[\mathcal{C}^{(k)}(z)\right]_{z=0}.$$

NB: Le choix de m et r permet d'altérer la fonction génératrice pour la rendre plus simple.

Décroissance des $\{a_k\}_{k\in\mathbb{N}}$: Cas des montants de loi $\Gamma(1,\beta)$

Si $U_i \sim \Gamma(1,\beta)$, alors

$$L_{g_M}(s) = \frac{p}{1 + \frac{\beta}{(1-p)} s},$$

et

$$\mathcal{C}(z) = \frac{pm(1+z)^{1-r}}{m + z\left(m - \frac{\delta}{1-p}\right)}.$$

- $m = \frac{\delta}{1-p}$ et $r = 1$

$$\mathcal{C}(z) = p.$$

- $a_0 = p$, et $a_k = 0$ pour $k \geq 1$.

Illustrations numériques: Cas des montants de loi $\Gamma(\alpha, \beta)$

- Intensité du processus de Poisson: $\lambda = 1$,
- Montant de sinistres de loi $\Gamma(2, 1)$,
- Primes périodiques: $c = 5$,
- Coefficient d'ajustement: $\gamma_M = \frac{1}{24}\left(19 - \sqrt{265}\right)$.

La probabilité de ruine ultime est donnée par

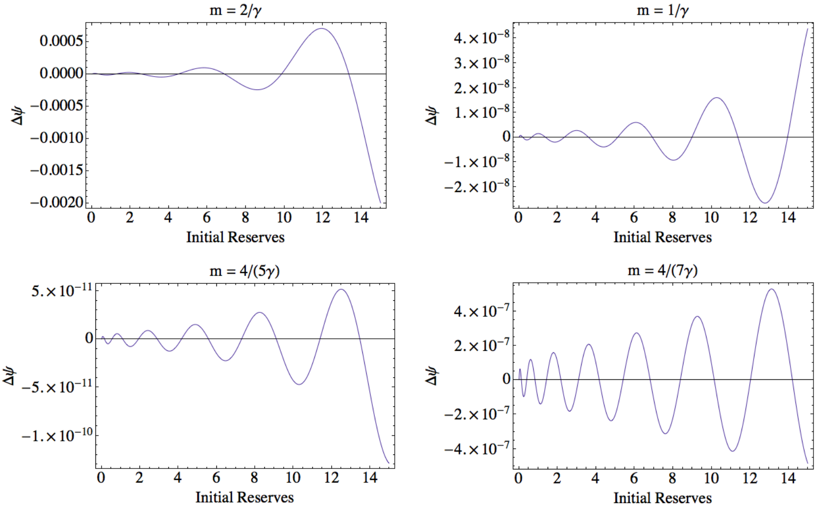
$$\psi(u) = 0.461861 e^{-0.441742u} - 0.0618615 e^{-1.35826u}.$$

Plusieurs paramétrisations sont testées avec $K = 40$.

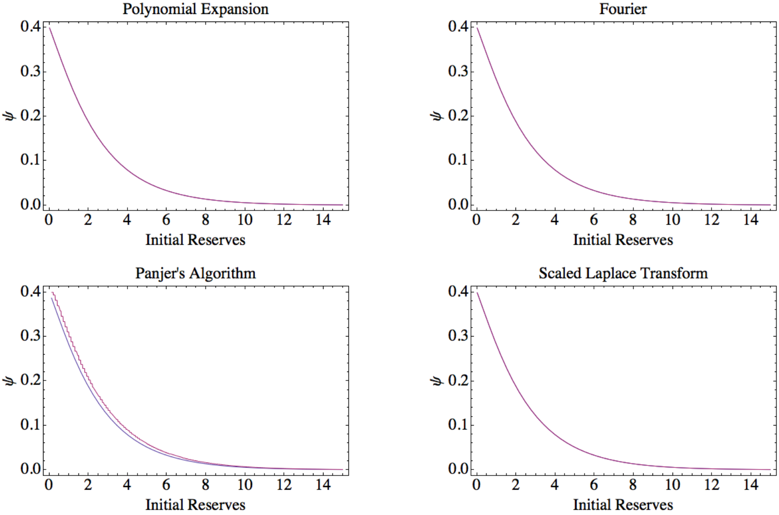
- La précision est mesurée à l'aide de l'écart relatif,

$$\Delta\psi(u) = \frac{\psi_{Approx}(u) - \psi(u)}{\psi(u)}.$$

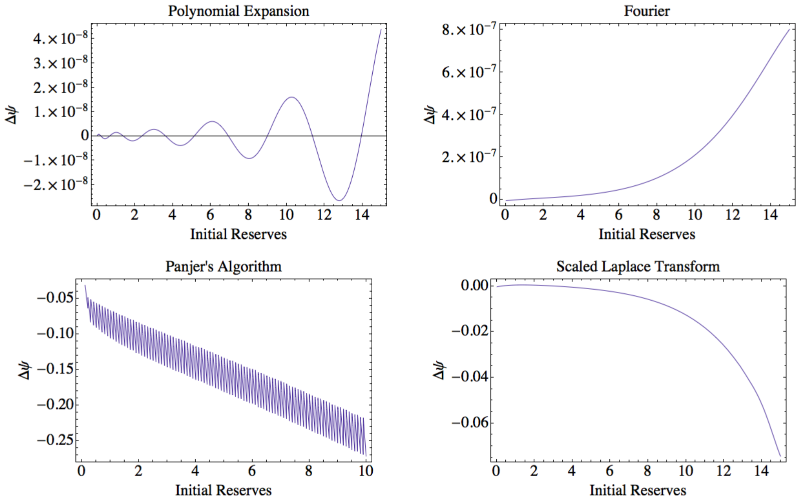
Illustrations numériques:
Cas des montants de loi $\Gamma(\alpha, \beta)$



Illustrations numériques: Cas des montants de loi $\Gamma(\alpha, \beta)$



Illustrations numériques: Cas des montants de loi $\Gamma(\alpha, \beta)$



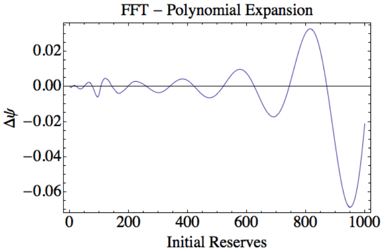
Illustrations numériques: Cas des montants de loi $\mathcal{U}(\alpha, \beta)$

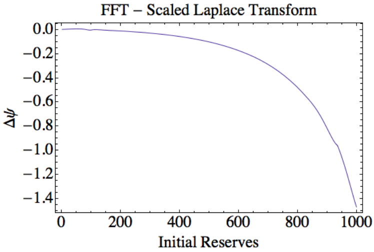
- Intensité du processus de Poisson: $\lambda = 1$,
- Montant de sinistres de loi $\mathcal{U}(0, 100)$,
- Primes périodiques: $c = 80$,
- Coefficient d'ajustement: $\gamma_M = 0.013$,
- Ordre de troncature des polynômes: K=40.

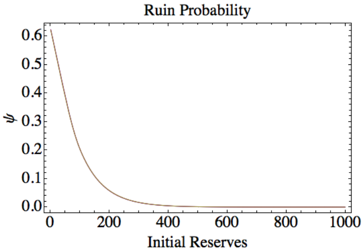
La probabilité de ruine ultime n'est pas disponible **explicitement**.

- L'approximation via l'inversion de la transformée de Fourier sert de référence.

Illustrations numériques: Cas des montants de loi $\mathcal{U}(\alpha, \beta)$







Soit 2 portefeuilles de contrats d'assurance non vie associés à la même branche d'activité et appartenant à deux assureurs.

- Les risques sont modélisés **conjointement** via

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{N_1} U_{1j} \\ \sum_{j=1}^{N_2} U_{2j} \end{pmatrix} + \sum_{j=1}^{M} \begin{pmatrix} V_{1j} \\ V_{2j} \end{pmatrix},$$

- N_1 et N_2 sont supposées indépendantes,
- Approximation polynomiale de la densité jointe du vecteur (X_1, X_2).

Un contrat de réassurance non-proportionnelle global

Le réassureur propose à l'assureur i un contrat de réassurance non proportionnelle de seuil de **rétention** b_i et de **portée** c_i, pour $i \in \{1, 2\}$.

La loi jointe de (X_1, X_2) est **utile**

- Tarifer le contrat de réassurance à seuils de rétention, et portées fixées.
- L'étude de l'exposition au risque du réassureur modélisé par

$$Z = \min\left[(X_1 - b_1)_+, c_1\right] + \min\left[(X_2 - b_2)_+, c_2\right],$$

où $(.)_+$ désigne la partie positive.

$\hookrightarrow$ **Value-at-Risk** de Z et marge de solvabilité.

Méthode d'approximation polynomiale en dimension 2

Soit (X_1, X_2) un vecteur aléatoire de loi de probabilité $\mathbb{P}_{X_1,X_2}$, de densité f_{X_1,X_2}.

- ν est la mesure de probabilité de référence, construite via le produit de deux mesures,

$$\begin{aligned} \nu(x_1, x_2) &= \nu_1(x_1) \times \nu_2(x_2), \\ f_\nu(x_1, x_2) &= f_{\nu_1}(x_1) \times f_{\nu_2}(x_2). \end{aligned}$$

- $\{Q_k^{\nu_i}\}_{k\in\mathbb{N}}$ forme un système orthonormal de polynômes par rapport à ν_i, pour $i \in \{1, 2\}$.
- $\{Q_{k,l}\}_{k,l\in\mathbb{N}}$ forme un système orthonormale de polynômes par rapport à ν, avec

$$Q_{k,l}(x_1, x_2) = Q_k^{\nu_1}(x_1) Q_l^{\nu_2}(x_2), \quad k, l \in \mathbb{N}.$$

Méthode d'approximation polynomiale en dimension 2

La méthode d'approximation polynomiale s'étend naturellement en dimension supérieure:

- Valide sous réserve de vérifier une condition d'intégrabilité.
- Borne exponentielle pour la densité jointe, et choix *ad hoc* des paramètres de la mesure de référence.
- Lien entre la transformée de Laplace multivariée, et la fonction génératrice des coefficients de la représentation polynomiale.

La distribution de probabilité du vecteur aléatoire

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{N_1} U_{1j} \\ \sum_{j=1}^{N_2} U_{2j} \end{pmatrix} + \sum_{j=1}^{M} \begin{pmatrix} V_{1j} \\ V_{2j} \end{pmatrix}$$

$$= \begin{pmatrix} W_1 \\ W_2 \end{pmatrix} + \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix},$$

admet de nombreuses singularités.

Choix de la mesure de probabilité de référence

- La distribution de $\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{N_1} U_{1j} \\ \sum_{j=1}^{N_2} U_{2j} \end{pmatrix}$, est donnée par

$$\begin{aligned} d\mathbb{P}_{W_1,W_2}(w_1,w_2) &= f_{N_1}(0)f_{N_2}(0)\delta_{0,0}(w_1,w_2) \\ &+ d\mathbb{G}_{W_1}(w_1)\times d\mathbb{G}_{W_2}(w_2) \\ &+ f_{N_1}(0)d\mathbb{G}_{W_2}(w_2)\times\delta_0(w_1) \\ &+ f_{N_2}(0)d\mathbb{G}_{W_1}(w_1)\times\delta_0(w_2). \end{aligned}$$

- Approximation polynomiale univariée de $\mathbb{G}_{W_i}$, pour $i = 1, 2$.
  $\hookrightarrow$ Mesure de référence gamma et polynômes de Laguerre généralisés.

Choix de la mesure de référence

- La distribution de $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \sum_{j=1}^{M} \begin{pmatrix} V_{1j} \\ V_{2j} \end{pmatrix}$, est donnée par

$$d\mathbb{P}_{Y_1,Y_2}(y_1, y_2) = f_M(0)\delta_{0,0}(y_1, y_2) + d\mathbb{G}_{Y_1,Y_2}(y_1, y_2).$$

$d\mathbb{G}_{Y_1,Y_2}$ est une mesure de probabilité défaillante de support $\mathbb{R}_+^2$.

- La mesure ν est définie par le produit de deux mesures gamma.
  - $\hookrightarrow$ ν_i est une mesure $\Gamma(m_i, r_i)$, pour $i = 1, 2$.
  - $\hookrightarrow$ $\{Q_k^{\nu_i}\}_{k\in\mathbb{N}}$ est une suite de polynômes de Laguerre généralisés, pour $i = 1, 2$.

Illustrations numériques: Fonction de survie de (Y_1, Y_2)

- M est de loi géométrique $\mathcal{NB}(1, 3/4)$,
- (V_1, V_2) est de loi exponentielle bivariée $DBVE(\rho, \mu_1, \mu_2)$,
  - $\hookrightarrow$ $\rho = \frac{1}{4}$,
  - $\hookrightarrow$ $\mu_1 = \mu_2 = 1$,
- L'approximation polynomiale est comparée à des approximations de Monte-Carlo.

La paramétrisation

$$m_1 = \frac{1}{(1-p)\mu_1}, \quad m_2 = \frac{1}{(1-p)\mu_2}, \quad r_1 = r_2 = 1.$$

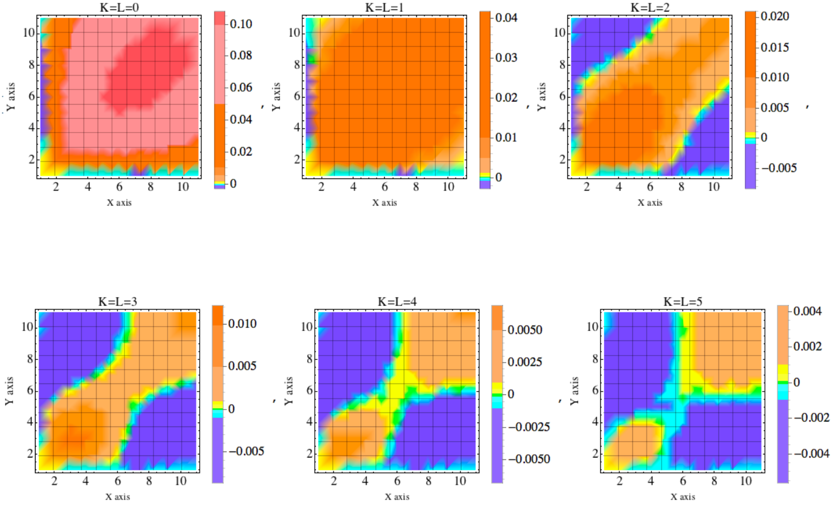
conduit à

$$\mathcal{C}(z_1, z_2) = \frac{1}{1 + z_1 z_2 (p^2 - \rho(1-p)^2 - p)},$$

et

$$a_{k,l} = \left[p^2 - \rho(1-p)^2 - p\right]^k \delta_{kl}, \quad k, l \in \mathbb{N}.$$

Illustrations numériques: Fonction de survie de (Y_1, Y_2)

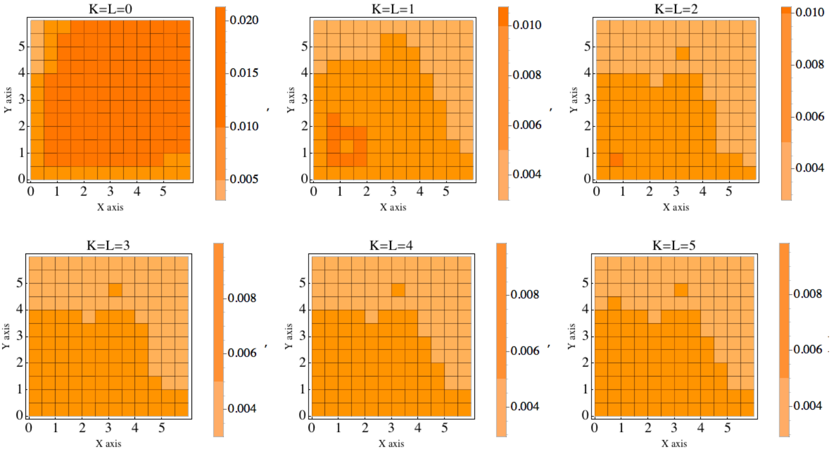


- N_1 et N_2 sont de loi géométrique $\mathcal{NB}(1, 3/4)$,
- $\{U_{1j}\}_{j\in\mathbb{N}}$ et $\{U_{2j}\}_{j\in\mathbb{N}}$ sont des suites de variables **i.i.d.** de loi $\Gamma(1,1)$,
  - $\hookrightarrow$ Densité disponible explicitement pour la distribution géométrique composée exponentielle.
- Approximation de la fonction de survie de (X_1, X_2).
- Seuils de rétention: $c_1 = c_2 = 1$,
- Portées: $b_1 = b_2 = 4$,
- Approximation de la fonction de survie de

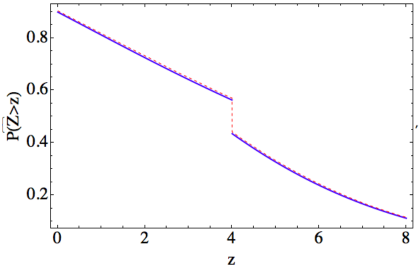
$$Z = \min\left[(X_1 - b_1)_+, c_1\right] + \min\left[(X_2 - b_2)_+, c_2\right].$$

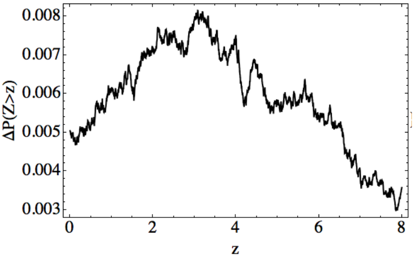
- Les approximations polynomiales sont comparées aux approximations de Monte-Carlo.

Illustrations numériques: Fonction de survie de (X_1, X_2)



Illustrations numériques: Coût de la réassurance





z	Monte Carlo approximation	Polynomial approximation
0	0.90385	0.898808
2	0.73193	0.724774
4	0.44237	0.435013
6	0.24296	0.237576

Conclusions et Perspectives

- La méthode d'approximation polynomiale est une méthode numérique efficace:
  - $\hookrightarrow$ Approximation de la probabilité de ruine ultime dans le modèle de ruine de Poisson composée,
  - $\hookrightarrow$ Approximation de la densité jointe dans un modèle collectif bivarié avec des applications intéressantes en réassurance.

Perspectives

- Applications à l'approximation de fonctions en actuariat et dans d'autres domaines des probabilités appliquées.
- Applications statistiques lorsque des données sont disponibles,

$\hookrightarrow$ Obtention d'un estimateur semi-paramétrique de la densité de probabilité.

Mes Publications

- Pierre-Olivier Goffard and Xavier Guerrault.
  Is it optimal to group policyholders by age, gender, and seniority for BEL computations based on model points?
  European Actuarial Journal, pages 1–16, 2015.
- Pierre-Olivier Goffard, S. Loisel, and D. Pommeret.
  Polynomial approximations for bivariate aggregate claim amount probability distributions.
  Working Paper, 2015.
- Pierre-Olivier Goffard, S. Loisel, and D. Pommeret.
  A polynomial expansion to approximate the ultimate ruin probability in the compound Poisson ruin model.
  Journal of Computational and Applied Mathematics, 2015.

Méthode d'approximation 1: l'algorithme de Panjer

Famille de Panjer

N admet une distribution de Panjer si

$$f_N(k+1) = \left(a + \frac{b}{k}\right) f_N(k)$$

Et son algorithme récursif

Si U admet une loi de probabilité discrète alors V et M aussi et

$$f_M(k) = \begin{cases} \mathcal{G}_N\left(f_U(0)\right) & \text{si } k = 0 \\ \frac{1}{1-af_V(0)} \sum_{j=1}^{k} \left(a + \frac{bj}{k}\right) f_V(j) f_M(k-j) & \text{si } k \geq 1 \end{cases}$$

Méthode d'approximation 2: Inversion de Fourier

Défnition de la transformée de Fourier

Soit $x \mapsto g(x)$ avec $x \in \mathbb{R}$, sa transformée de Fourier est donnée par

$$\mathcal{L}_g(is) = \int_0^{+\infty} e^{isx} g(x)\mathrm{d}x$$

Et sa formule d'inversion

Si $\int_{-\infty}^{+\infty} |\mathcal{L}_g(is)|ds < +\infty$, et g fonction continue et bornée alors

$$g(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-isx} \mathcal{L}_g(is)ds$$

Méthode d'approximation 2: Inversion de Fourier

Soit

$$g(u) = \begin{cases} e^{-ut}\psi(u) & \text{Si } u \geq 0 \\ g(-u) & \text{Si } u < 0 \end{cases}$$

alors

$$\psi(u) = \frac{2e^{ua}}{\pi} \int_0^{+\infty} \cos(uy) \Re\left[\mathcal{L}_\psi(a+iy)\right] dy$$

puis

$$\tilde{\psi}(u) = \frac{2e^{ua}}{\pi} h \left\{ \frac{1}{2} \Re\left[\mathcal{L}_\psi(a)\right] + \sum_{k=1}^{+\infty} \cos(ukh) \Re\left[\mathcal{L}_\psi(a+ikh)\right] \right\}$$

Méthode d'approximation 3: Moments exponentiels

- Approximation de la fonction de répartition par une binomiale mélange
- Y une variable aléatoire continue, à valeur sur $[0,1]$ et de loi $\mathbb{P}_Y$
  - $\hookrightarrow$ Paramètre d'une loi binomial mélange de fonction de répartition

$$
\begin{aligned}
\mathcal{B}_{n,\mathbb{P}_Y}(y) &= \sum_{k=0}^{\lfloor ny \rfloor} \int_0^1 \binom{n}{k} z^k (1-z)^{n-k} \mathrm{d}\mathbb{P}_Y(z) \\
&= \sum_{k=0}^{\lfloor ny \rfloor} \binom{n}{k} \mathbb{E}\left[Y^k (1-Y)^{n-k}\right] \\
&= \sum_{k=0}^{\lfloor ny \rfloor} \sum_{j=k}^{n} \binom{n}{j} \binom{j}{k} (-1)^{j-k} \mathbb{E}\left(Y^j\right)
\end{aligned}
$$

Méthode d'approximation 3: Moments exponentiels

- La méthode est basée sur le résultat de convregence suivant

$$\int_0^1 \sum_{k=0}^{\lfloor ny \rfloor} \binom{n}{k} z^k (1-z)^{n-k} \mathrm{d}\mathbb{P}_Y(z) \to \int_0^1 \mathbf{1}_{z<y} \mathrm{d}\mathbb{P}_Y(z), \quad n \to +\infty$$

La fonction de répartition est approchée par

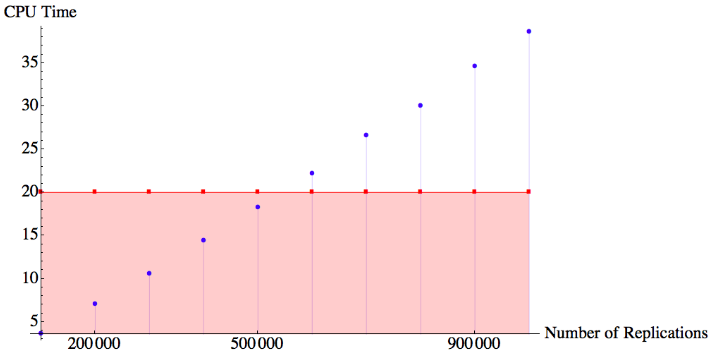
$$F_Y(y) \approx \sum_{k=0}^{\lfloor ny \rfloor} \sum_{j=k}^{n} \binom{n}{j} \binom{j}{k} (-1)^{j-k} \mathbb{E}\left(Y^j\right).$$

- X variable aléatoire sur $\mathbb{R}^+$.
  - $\hookrightarrow$ Changement de variable $Y = e^{-cX}$ avec $0 < c < 1$,

$$\overline{F}_X(x) \approx \sum_{k=0}^{\lfloor ne^{-cx} \rfloor} \sum_{j=k}^{n} \binom{n}{j} \binom{j}{k} (-1)^{j-k} \mathcal{L}_X(-jc).$$

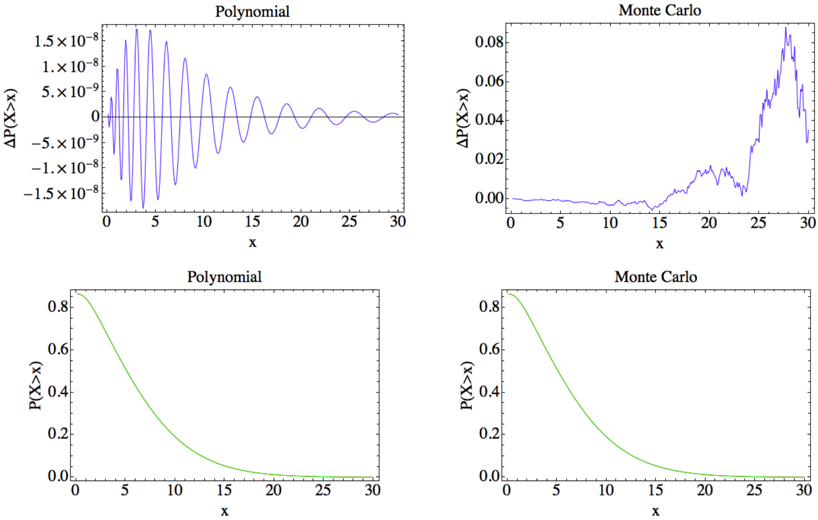
Comparaison: Monte Carlo *VS* Polynômes

Exemple: Distribution de Poisson composé $[\mathcal{P}(2), \Gamma(3,1)]$,



- Ordre de troncature: K=75 $\Rightarrow$ 20 sec,
  - $\hookrightarrow$ 600 000 simulations de Monte Carlo.

Comparaison: Monte Carlo *VS* Polynômes



Soit $X_1, \ldots, X_n$ un échantillon de taille n.

- L'approximation polynomiale de la densité admet la forme

$$f_X^K(x) = f_{X,\nu}^K(x) f_\nu(x) = \sum_{k=0}^{K} a_k Q_k(x) f_\nu(x).$$

- La formule d'approximation se transforme en estimateur de la densité, avec

$$\widehat{f_X^K}(x) = \widehat{f_{X,\nu}^K}(x) \widehat{f_\nu}(x) = \sum_{k=0}^{K} \widehat{a_k} Q_k(x) \widehat{f_\nu}(x).$$

La procédure d'estimation comprend,

1. une estimation paramétrique de la mesure de référence,
2. une estimation non paramétrique des coefficients de la représentation polynomiale.

Application Statistique: Définition de l'estimateur

L'expression des coefficients de la représentation polynômiale est

$$a_n = \mathbb{E}\left[Q_k(X)\right], \quad k \in \mathbb{N}.$$

Un estimateur sans biais de a_k est donné par

$$Z_k = \frac{1}{n}\sum_{i=1}^{n} Q_k(X_i), \quad k = 1, \dots, K,$$

sa variance est notée $\sigma^2_{k,n} = \mathbb{V}(Z_k)$.

- les coefficients de la représentation polynomiale sont estimés par

  $$\widehat{a_k} = w_k Z_k,$$

  où $\mathbf{w} = (w_1, \dots, w_K)$ est le modulateur.

  $\hookrightarrow$ Permet une optimisation de l'erreur moyenne quadratique intégrée.
- Dans la suite $K = n$.

Application Statistique: EQMI

L'Erreur Quadratique Moyenne Intégrée est donnée par

$$
\begin{aligned}
EQMI(\widehat{f^n_{X,\nu}}, f^n_{X,\nu}) &= \mathbb{E}\int \left[\widehat{f_{X,\nu}}(x) - f_{X,\nu}(x)\right]^2 dx \\
&= \sum_{k=1}^{n}(1-w_k)^2 a_k^2 + \sum_{k=n+1}^{+\infty} a_k^2 \\
&+ \sum_{k=1}^{n} w_k^2 \sigma_{k,n}^2.
\end{aligned}
$$

L'Erreur Quadratique Moyenne Intégrée modifiée est optimisée,

$$
EQMI(\widehat{f^n_{X,\nu}}, f^n_{X,\nu}) = \sum_{k=1}^{n} w_k^2 \sigma_{k,n}^2 + \sum_{k=1}^{n}(1-w_k)^2 a_k^2.
$$

Application Statistique: EQMI

Les quantités a_k^2, et $\sigma_{k,n}^2$ sont estimés sans biais via

$$\widehat{\sigma}_{k,n}^2 = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left[Q_k(X_i) - Z_k\right]^2, \quad k = 1, \ldots, n,$$

$$\widehat{a_k^2} = \max\left(Z_k^2 - \widehat{\sigma}_{k,n}^2, 0\right), \quad k = 1, \ldots, n.$$

L'Erreur Quadratique Moyenne Intégrée est estimée par

$$\widehat{EQMI}\left(\widehat{f_{X,\nu}^n}, f_{X,\nu}^n\right) \quad = \quad \sum_{k=1}^{n} w_k^2 \widehat{\sigma_{k,n}^2} + \sum_{k=1}^{n} (1 - w_k)^2 \widehat{a_k^2}.$$

Application Statistique: Modulateur SME

La classe $\mathcal{M}_{SME}$ désigne les modulateurs de Sélection de Modèle Emboités tels que $\mathbf{w} = (1, \ldots, 1, 0, \ldots, 0)$.

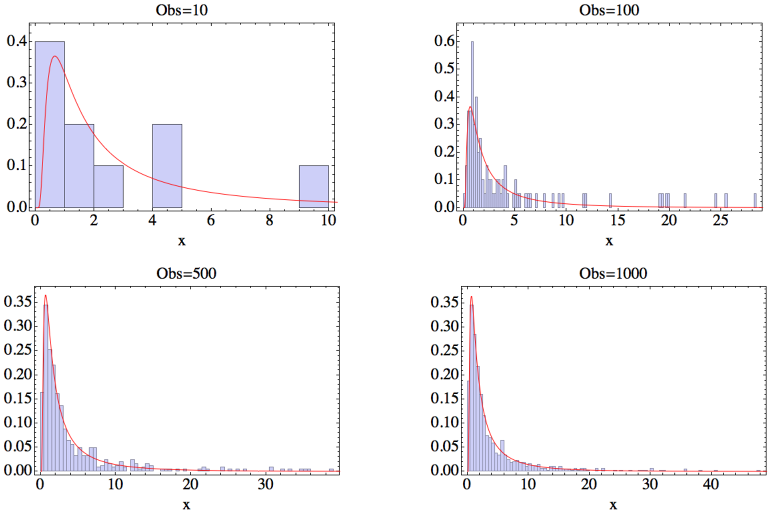
- Optimisation en calculant l'EQMI $\forall \mathbf{w} \in \mathcal{M}_{SME}$.

Illustration: Estimation de la densité d'une distribution *Inverse Gaussian*, $\mathcal{IG}(\lambda, \mu)$, de densité

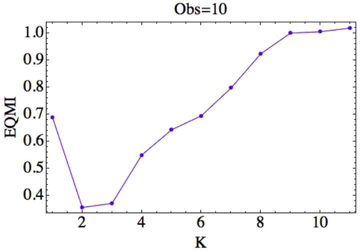
$$f_X(x) = \begin{cases} \dfrac{\sqrt{\frac{\lambda}{x^3}}\, e^{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}}}{\sqrt{2\pi}}, & x > 0, \\ 0, & \text{Sinon.} \end{cases}$$

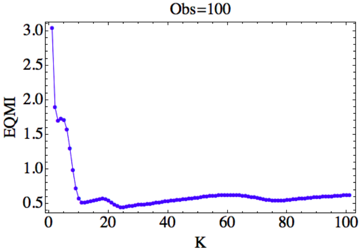
- Temps de passage à un niveau fixé d'un mouvement brownien avec drift,
- pour l'exemple $\lambda = 2$, et $\mu = 4$.

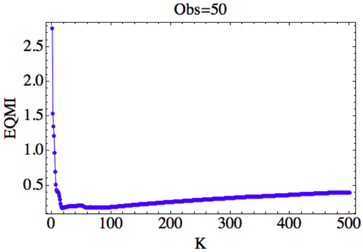
Application Statistique: Visualisation des données

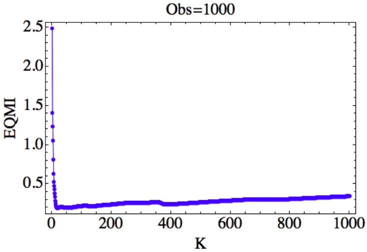


Application Statistique: EQMI

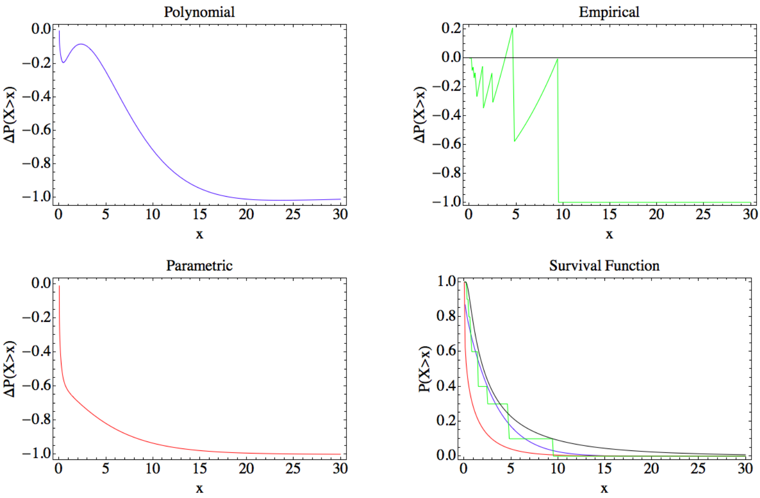




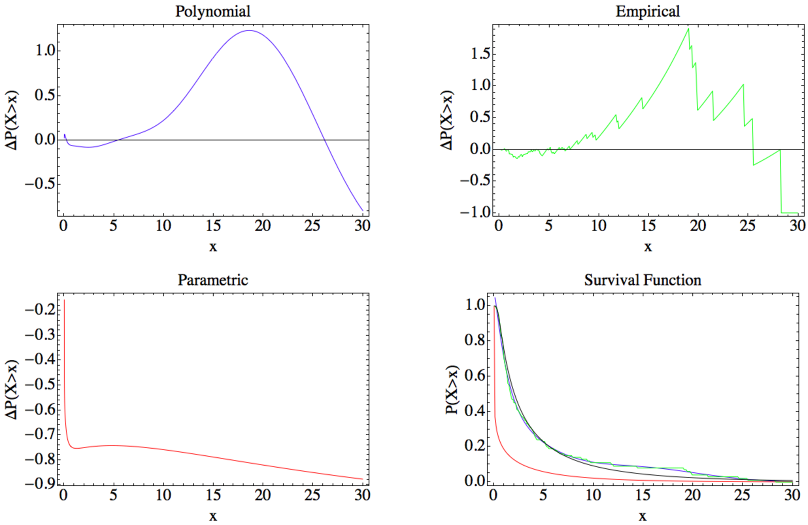




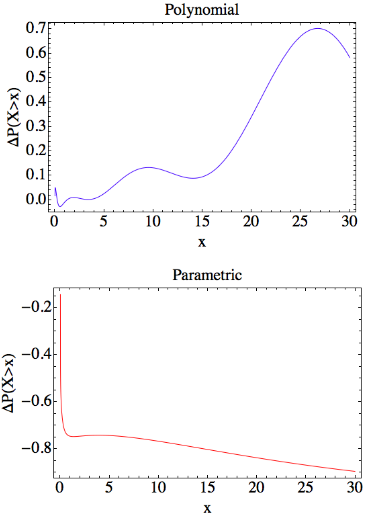
Application Statistique: Obs=10



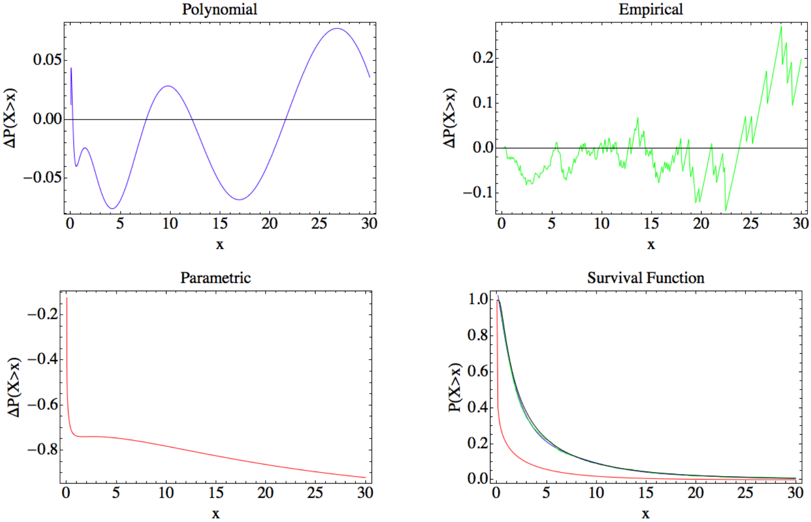
Application Statistique: Obs=100



Application Statistique: Obs=500



Application Statistique: Obs=1000



Application Statistique: Distribution Composé

- Echantillon de charges totales: $(X_1, \ldots, X_n)$,
- Echantillon de nombre de sinistres: $(N_1, \ldots, N_n)$,
- Echantillon de montants de sinistres: $(U_1, \ldots, U_{N_1}, \ldots, U_{N_1+\ldots+N_n})$.

La taille de l'échantillon pour les charges totales, et les nombres de sinistres est potentiellement faible.

- Beaucoup d'observations sont disponibles pour les montants de sinistres.

Full Parametric

- Tests d'adéquation pour calibrer la loi des montants et des fréquences,
- Inférence statistiques de paramètres,
- Méthode d'approximation.

Full NonParametric

- Inférence des paramètres de la mesure de référence,
  - $\hookrightarrow$ Sur la base de quelles observations?
- Coefficients de la représentation polynomiale estimés à l'aide de $(X_1, \ldots, X_n)$.

Semi-Parametric

- Calibration d'une loi pour la fréquence des sinistres.
- Inférence des paramètres de la loi pour la fréquence,
  - $\hookrightarrow$ Utilisation de $(N_1, \ldots, N_n)$.
- Coefficients de la représentation polynomiale estimés à l'aide de $(U_1, \ldots, U_{N_1}, \ldots, U_{N_1+\ldots+N_n})$, et d'une relation de récurrence entre les moments de X, N, et U.