Polynomial aproximations of probability density functions

Applications to insurance

P.O. Goffard

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

Aarhus University seminar

Univariate aggregate claim amounts.

Consider a non-life insurance portfolio.

- Over a given time period,
  - $\hookrightarrow$ The number of claims is modeled through a counting random variable N,
  - $\hookrightarrow$ The claim sizes are a sequence of non-negative, **i.i.d.** random variables $(U_i)_{i\in\mathbb{N}}$.
- The aggregate claim amounts is defined as

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

  - $\hookrightarrow$ $\{U_i\}_{i\in\mathbb{N}}$ are independent of N.

X is governed by a compound distribution $(\mathbb{P}_N, \mathbb{P}_U)$.

Multivariate aggregate claim amounts.

Consider n non-life insurance portfolios,

- The risks are jointly modeled via

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

$\hookrightarrow$ $\mathbf{N} = (N_1, \ldots, N_n)$ is a vector of counting random variables,

$\hookrightarrow$ $(U_{1j})_{j\in\mathbb{N}}, \ldots, (U_{nj})_{j\in\mathbb{N}}$ are independent sequences of non-negative and **i.i.d.** random variables,

$\hookrightarrow$ $\{\mathbf{V}_i\}_{i\in\mathbb{N}} = \{(V_{1i}, \ldots, V_{ni})\}_{i\in\mathbb{N}}$ is a sequence of **i.i.d.** random vectors,

$\hookrightarrow$ M is a counting random variable,

$\hookrightarrow$ $\{\mathbf{V}_i\}_{i\in\mathbb{N}}$, $\mathbf{N}$, M and $(U_{j1})_{j\in\mathbb{N}}, \ldots, (U_{jn})_{j\in\mathbb{N}}$ are independent.

Consider a non-life insurance portfolio.

- Let $t \geq 0$ denotes the time,
  - $\hookrightarrow$ The number of claim until time t is a counting process $\{N_t\}_{t\geq 0}$,
  - $\hookrightarrow$ The claim sizes are a sequence of **i.i.d.**, non-negative random variables $(U_i)_{i\in\mathbb{N}}$.
- The liability of the insurer at time t are

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

  - $\hookrightarrow$ $(U_i)_{i\in\mathbb{N}}$ is independent of $\{N_t\}_{t\geq 0}$.

Ruin Theory: A more dynamic vision of risk management.

- The financial reserve allocated to this portfolio, **a.k.a.** *Risk Reserve Process*, is

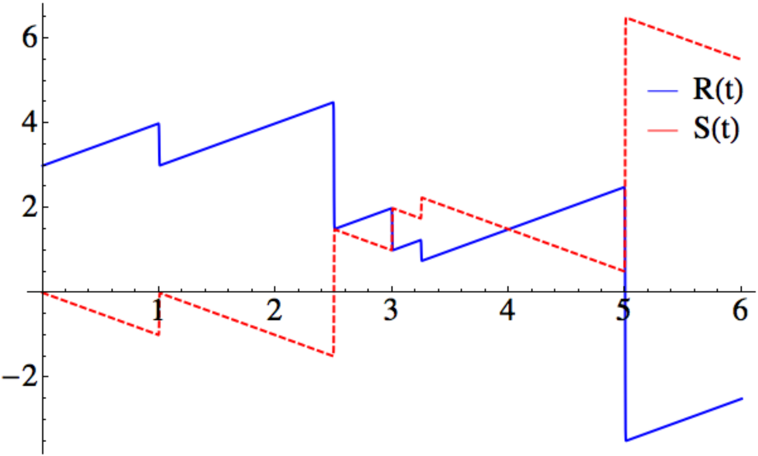
$$R_t = u + ct - \sum_{i=1}^{N_t} U_i,$$

  - $\hookrightarrow$ **u** denotes the initial reserve,
  - $\hookrightarrow$ **c** denotes the rate at which premiums are collected per unit of time.

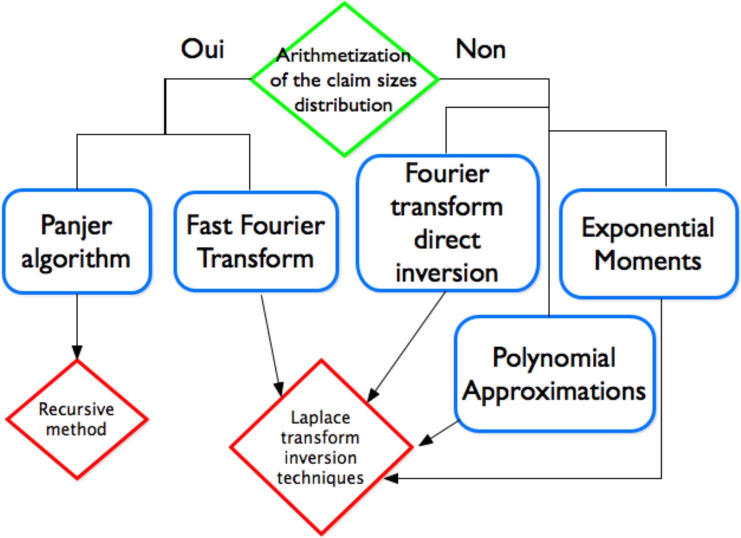
- The *Claim Surplus process* is defined as

$$S_t = u - R_t.$$

Ruin theory: A graphical visualization.



Numerical Methods: A guided tour.



EXECUTIVE SUMMARY.

Application of the polynomial approximation method to two problem.

1. Computation of the ultimate ruin probability in the compound Poisson ruin model.
   - $\hookrightarrow$ $\{N_t\}_{t\geq 0}$ is a Poisson process with intensity λ.
2. Study of a bivariate aggregate claim amounts distribution with reinsurance motivations.

Ruin probability definition

- Infinite time horizon ruin probability or ultimate ruin probability,

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

- Finite time horizon ruin probability,

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

- Infinite and finite time horizon non ruin probabilities,

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

Alternative definition for ruin probabilities.

- Time to ruin and *Claims surplus process* maximum,

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

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

- Infinite and finite time horizon ruin probabilities,

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

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

The safety loading.

The average amount of claims per unit of time is given by

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

- The safety loading η, is defined as

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

- *Net Benefit Condition*

$$\eta > 0,$$

$\hookrightarrow$ If $\eta < 0$ then $\psi(u) = 1$,
$\hookrightarrow$ If $\eta > 0$ then $\psi(u) < 1$.

Ultimate ruin probability: Pollaczek-Khinchine formula.

In the frame of the compound Poisson ruin model,

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

- N is governed by a geometric distribution $\mathcal{G}(p)$, where $p = \frac{\lambda E(U)}{c} < 1$ and

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

- $(V_i)_{i\in\mathbb{N}^*}$ is a sequence of **i.i.d.** and non-negative random variable having PDF,

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

- $(V_i)_{i\in\mathbb{N}^*}$ and N are independent.

Notations: Convolution product.

- Let U and V be two random variables with associated PDF f_U et f_V,

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

$\hookrightarrow$ Convolution product of f_U and f_V.

- Let $S = \sum_{i=1}^{n} U_i$ be the sum of n **i.i.d.** random variables,

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

$\hookrightarrow$ $f_U^{(*n)}$ referes to the n-fold convolution of f_U with itself.

Notations:
Laplace transform.

The Laplace transform of a random variable is defined as

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

Thus the Laplace transforms of sums of independent random variables is given by

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

Compound geometric distribution.

- The random variable $M = \sum_{i=1}^{N} U_i$ admits a compound distribution,

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

where

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

and

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

Laplace transform of ruin probabilities.

The Laplace transform of M is given by,

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

- $\mathcal{G}_N$ is the probability generating function of N,

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

which implies that

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

and finally

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

The polynomial approximation: Univariate case

Let X be a random variable governed by $\mathbb{P}_X$, with PDF f_X.

- ν is a reference probability measure, with PDF f_ν.
  - $\hookrightarrow$ $\mathbb{P}_X$ is absolutely continuous with respect to ν,

$$f_{X,\nu}(x) = \frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}\nu}(x).$$

- $\{Q_k\}_{k\in\mathbb{N}}$ is a system of orthogonal polynomials with respect to ν,

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

The idea is to perform an orthogonal projection of $f_{X,\nu}$ onto the orthogonal basis $\{Q_k\}_{k\in\mathbb{N}}$.

The polynomial approximation: Univariate case

- $\mathbb{P}_{\mathbb{X}}$ is absolutely continuous with respect to ν,
  $\hookrightarrow$ Existency of $\frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}\nu}$.
- The set of polynomials is a dense set of $L^2(\nu)$,
  $\hookrightarrow$ $\{Q_k\}_{k\in\mathbb{N}}$ is an orthogonal system of $L^2(\nu)$.

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

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

where

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

The polynomial approximation: Univariate case.

- The PDF of the random variable X admits a polynomial expansion,

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

- Approximations follow from truncation of order K,

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

Approximations of the cumulating distribution function or the survival function are obtained through integration.

Polynomial approximation of ruin probabilities.

$M = \sum_{i=1}^{N} V_i$ is governed by a probability measure

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

If $\frac{\mathrm{d}\mathbb{G}_M}{\mathrm{d}\nu} \in L^2(\nu)$ then,

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

The approximation of the ruin probability is derived by truncation and integration,

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

Choice of the reference probability measure.

$\mathrm{d}\mathbb{G}_M$ is a defective probability measure having support on $\mathbb{R}_+$.

- The gamma distribution $\Gamma(m, r)$ has support on $\mathbb{R}_+$, its PDF is given by

$$\mathrm{d}\nu(x) = \frac{x^{r-1} e^{-x/m}}{m^r \Gamma(r)} \mathrm{d}\lambda(x)$$

- The orthogonal polynomials associated to the gamma measure are the generalized Laguerre polynomials.

The choice of the parameters m et r is really **important**.

Satisfaction of the integrability condition.

Theorem

Let X be a continuous and non-negative random variable,

- H1 There exists $\gamma_X = \inf\{s > 0, \mathcal{L}_X(s) = +\infty\}$.
- H2 Let $a \geq 0$, $x \mapsto f_X(x)$ is monotically decreasing for $x \geq a$.

Then, for $x \geq a$,

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

In the case of ruin probabilities, **H2** is checked, and $\gamma_M = \inf\{s > 0, \mathcal{L}_{g_M}(s) = +\infty\}$ is the unique solution of

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

Satisfaction of the integrability condition.

The integrability condition may be rewritten as

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

Applying g_M, for $s_0 < \gamma_M$, leads to

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

The integrability condition is satisfied if the parameters is setted as follows

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

Study of the decay of $\{a_k\}_{k\in\mathbb{N}}$: First result.

The quadratic loss after of the polynomial approximation is given by

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

$\Rightarrow$ Direct link between the decreasing of $\{a_k\}_{k\in\mathbb{N}}$ and the accuracy of the polynomial approximation.

Proposition

H1 $x \mapsto g_{M,\nu}(x)$ is continuous and twice differentiable function

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

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

Study of the decay of $\{a_k\}_{k\in\mathbb{N}}$: Study of the generating function.

The defective PDF of M admits the polynomial representation

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

Taking the Laplace transform leads to

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

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

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

Study of the decay of $\{a_k\}_{k\in\mathbb{N}}$: Study of the generating function.

The generating function of the coefficients is expressed in term of the Laplace transform via

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

- The coefficients follow from differentiations and evaluations at 0,

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

NB: the choice of m et r alter the generating in order to make it simpler.

Study of the decay of $\{a_k\}_{k\in\mathbb{N}}$: $\Gamma(1,\beta)$ distributed claim sizes.

If $U_i \sim \Gamma(1,\beta)$, then

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

and

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

- $m = \frac{\beta}{1-p}$ and $r = 1$
$$\mathcal{C}(z) = p.$$
- $a_0 = p$, and $a_k = 0$ for $k \geq 1$.

Numerical illustrations: $\Gamma(\alpha, \beta)$ distributed claim sizes.

- Intensity of the Poisson process: $\lambda = 1$,
- Claim sizes $\Gamma(2, 1)$ distributed,
- Premium rate: $c = 5$,
- Adjustment coefficient: $\gamma_M = \frac{1}{24}\left(19 - \sqrt{265}\right)$.

The ruin probability is given by

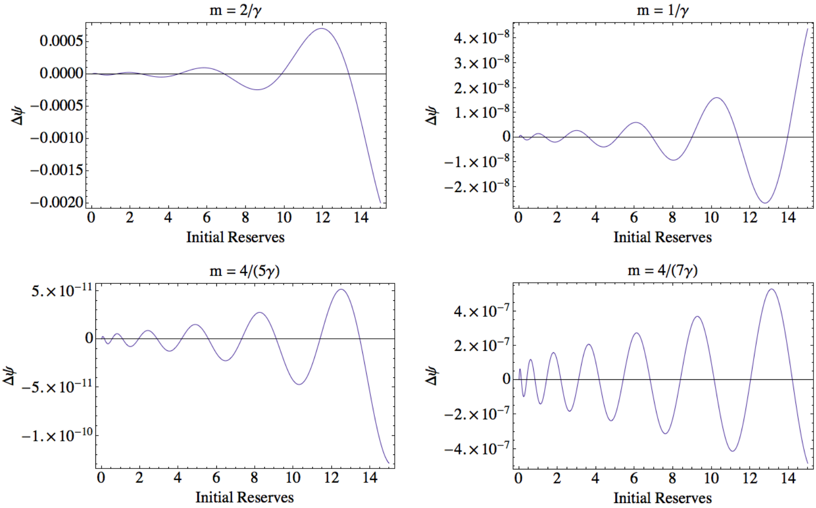
$$\psi(u) = 0.461861 e^{-0.441742u} - 0.0618615 e^{-1.35826u}.$$

Several parametrizations are tested with an order of truncation $K = 40$.

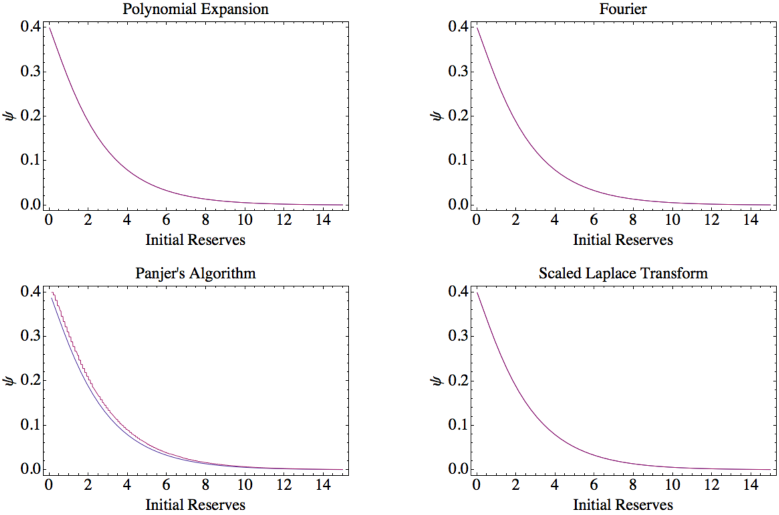
- The accuracy is given in terms of relative error,

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

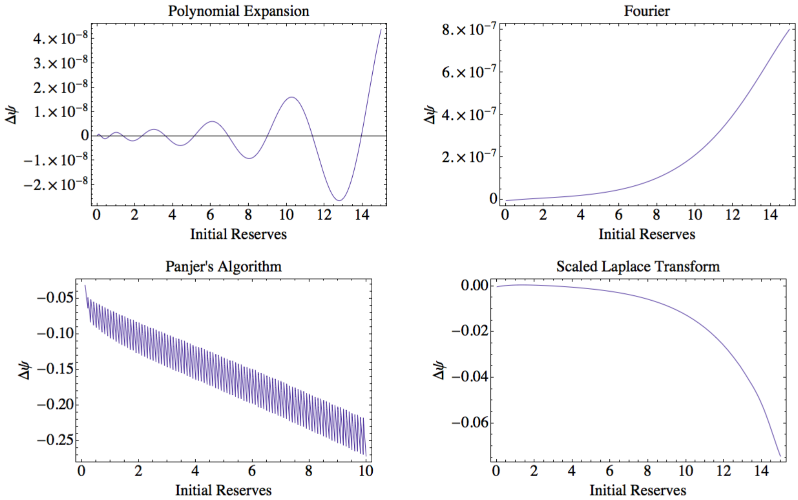
Numerical illustrations: $\Gamma(\alpha, \beta)$ distributed claim sizes.



Numerical illustrations: $\Gamma(\alpha, \beta)$ distributed claim sizes.



Numerical illustrations: $\Gamma(\alpha, \beta)$ distributed claim sizes.



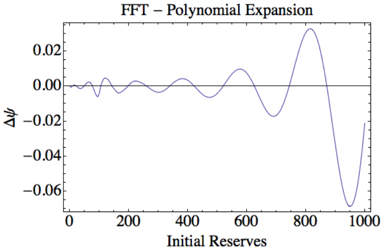
Numerical illustrations: $\mathcal{U}(\alpha, \beta)$ distributed claim sizes.

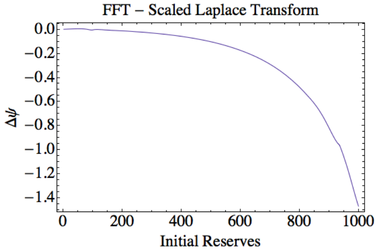
- Intensity of the Poisson process: $\lambda = 1$,
- Claim sizes $\mathcal{U}(0, 100)$ distributed,
- Premium rate: $c = 80$,
- Adjustments coefficient: $\gamma_M = 0.013$,
- Truncation order: K=40.

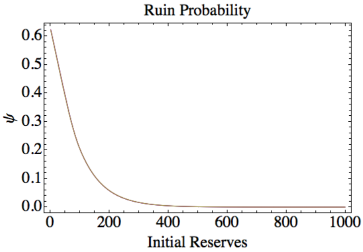
The ruin probability is not available in a closed form.

- The approximation using the direct Fourier transform inversion is used as benchmark.

Numerical illustrations: $\mathcal{U}(\alpha, \beta)$ distributed claim sizes.







Two insurers and one reinsurer: Bivariate collective model.

Consider 2 non-life insurance portfolios associated to the same line of business of two insurance company.

- The risks are modeled jointly 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 and N_2 are assumed to be independent,
- Polynomial approximation of the joint PDF of (X_1, X_2).

A global reinsurance treaty

The reinsurer propose to insurer i a non proportional reinsurance contract with priority b_i and limit c_i, for $i \in \{1, 2\}$.

The joint distribution of (X_1, X_2) is **useful**.

- To compute the premium associated with this reinsurance treaty.
- To study the risk exposure of the reinsurer

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

where $(.)_+$ denotes the positive part.

$\hookrightarrow$ **Value-at-Risk** of Z and solvability margins.

The polynomial approximation: Bivariate case.

Let (X_1, X_2) be a random vector $\mathbb{P}_{X_1,X_2}$, admiting a PDF f_{X_1,X_2}.

- ν is the reference probability measure, builded from the product of two probability measures,

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

- $\{Q_k^{\nu_i}\}_{k\in\mathbb{N}}$ is an orthonormal polynomial system with respect to ν_i, for $i \in \{1, 2\}$.
- $\{Q_{k,l}\}_{k,l\in\mathbb{N}}$ is an orthonormal polynomial system with respect to ν, where

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

The polynomial approximation: Bivariate case.

The polynomial approximation method extends naturally within a two-dimensionnal context:

- Integrability condition.
- Exponential bound for the joint density if the bivariate Laplace transforms exists for positive arguments.
- Link between the Laplace transform and the generating function of the coefficients.

Bivariate aggregate claim amounts distribution.

The probability distribution of the random vector

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{N_1} U_{1j} \\ \sum_{j=1}^{N_2} U_{2j} \end{pmatrix} + \sum_{j=1}^{M} \begin{pmatrix} V_{1j} \\ V_{2j} \end{pmatrix} = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix} + \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix},$$

admits a bunch of singularities...

Choice of the reference probability measure.

- The distribution of $\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{N_1} U_{1j} \\ \sum_{j=1}^{N_2} U_{2j} \end{pmatrix}$, is given by

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

- Univariate polynomial approximations of $\mathbb{G}_{W_i}$, for $i = 1,2$.
  $\hookrightarrow$ Gamma probability distribution and generalized Laguerre polynomials.

Choice of the reference probability measure.

- The probability distribution of $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \sum_{j=1}^{M} \begin{pmatrix} V_{1j} \\ V_{2j} \end{pmatrix}$, is given by

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

$d\mathbb{G}_{Y_1,Y_2}$ is defective probability distribution having support on $\mathbb{R}_+^2$.

- The probability measure ν is the product of two gamma measure.
  - $\hookrightarrow$ ν_i is a $\Gamma(m_i, r_i)$ probabilty measure, for $i = 1, 2$.
  - $\hookrightarrow$ $\{Q_k^{\nu_i}\}_{k\in\mathbb{N}}$ is a generalized Laguerre polynomials sequence, for $i = 1, 2$.

Numerical illustrations: Survival function of (Y_1, Y_2)

- M is governed by a negative binomial distribution $\mathcal{NB}(1, 3/4)$,
- (V_1, V_2) admits a bivariate exponential distribution of Downton type $DBVE(\rho, \mu_1, \mu_2)$,
  - $\hookrightarrow$ $\rho = \frac{1}{4}$,
  - $\hookrightarrow$ $\mu_1 = \mu_2 = 1$,
- The polynomial approximation is compared to Monte Carlo approximations.

The parametrization

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

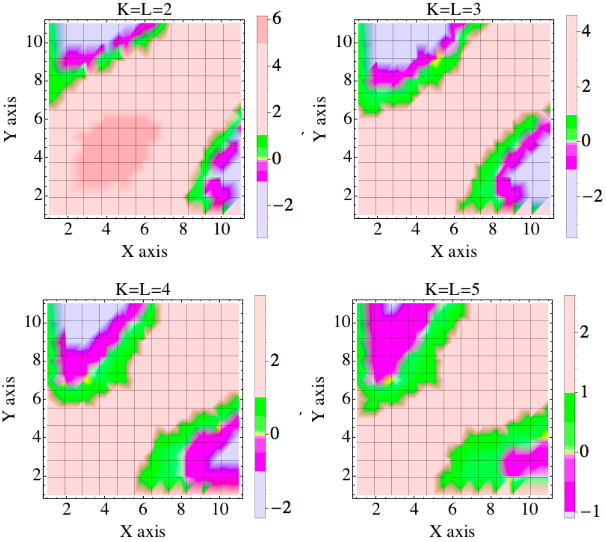
leads to

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

and therefore

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

Numerical illustrations: Survival function of (Y_1, Y_2)



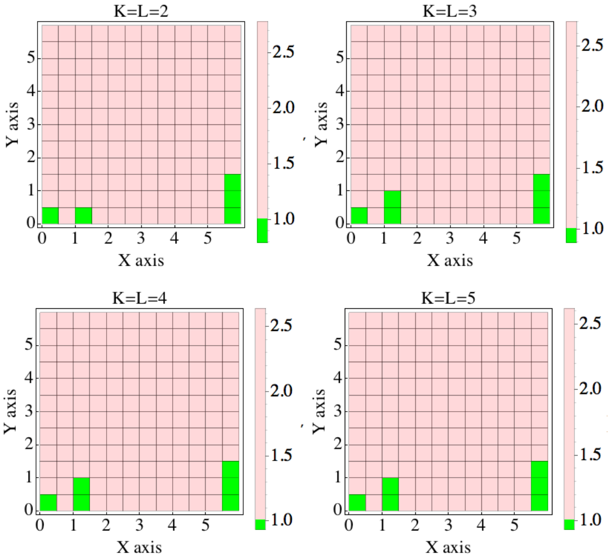
Illustrations numériques: Distribution of (X_1, X_2)

- N_1 and N_2 are $\mathcal{NB}(1, 3/4)$ distributed,
- $\{U_{1j}\}_{j\in\mathbb{N}}$ et $\{U_{2j}\}_{j\in\mathbb{N}}$ are **i.i.d.** random variables $\Gamma(1,1)$ distributed,
  - $\hookrightarrow$ PDF available in a closed form in the case of the geometric compound distribution with exponential claim sizes.
- Approximation of the survival function of (X_1, X_2).
- Priorities: $c_1 = c_2 = 1$,
- Limits: $b_1 = b_2 = 4$,
- Approximation of the survival function of

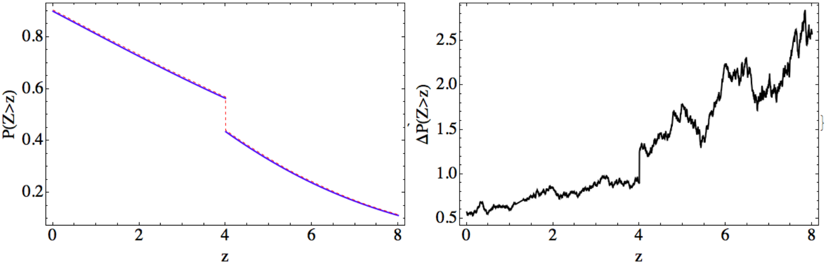
$$Z = \min\left[(X_1 - b_1)_+, c_1\right] + \min\left[(X_2 - b_2)_+, c_2\right].$$

- The polynomial approximations are compared to Monte Carlo approximations.

Numerical illustrations: Survival function of (X_1, X_2)



Numerical Illustrations: Cost of reinsurance



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

- The polynomial approximation is an efficient numerical method:
  - ↪ Approximation of the probability of ultimate ruin in the compound Poisson ruin model,
  - ↪ Approximation of probabilities associated to a bivariate aggregate claim amounts distribution with interesting reinsurance application.

Perspectives

- Application of this method to other actuarial problem or related to other fields of applied probability.
- Statistical application when data are available,

↪ The approximation formula can be turned into a semi-parametrical estimator of the PDF.

My supervisors

Approximation method 1: Panjer's algorithm.

Panjer's family

N is governed by a counting distribution that belongs to Panjer's family if

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

And its recursive algorithm

If U admits a discrete probability distribution then $M = \sum_{i=1}^{N} U_i$ too and

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

Direct Fourier transform inversion.

Fourier transform defintion

Let $x \mapsto g(x)$ be a real function, its Fourier transform is defined as

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

Fourier transform inversion formula

If $\int_{-\infty}^{+\infty} |\mathcal{L}_g(is)| ds < +\infty$, and g is continuous and bounded then

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

Apppproximation method 2: Direct Fourier trasform inversion.

Consider

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

then

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

and finally

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

Approximation method 3: Exponential moments.

- Approximation of the cumulating distribution function through a mixed binomial distribution.
- Y is a continuous random variable, with values in $[0,1]$ and governed by $\mathbb{P}_Y$.
  - $\hookrightarrow$ Mixing parameter

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

Méthode d'approximation 3: Exponential moments

- The method is based on the following convergence result

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

The cumulating distribution function is aproximated by

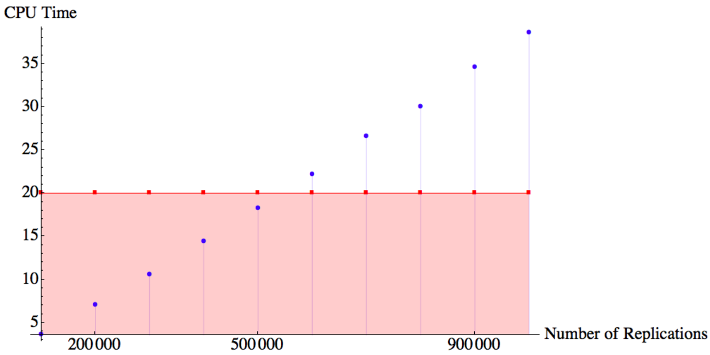
$$F_Y(y) \approx \sum_{k=0}^{\lfloor ny \rfloor} \sum_{j=k}^{n} \binom{n}{j} \binom{j}{k} (-1)^{j-k} \mathbb{E}\left(Y^j\right).$$

- X is $\mathbb{R}^+$-valued random variable.
  - $\hookrightarrow$ Change of variable $Y = e^{-cX}$ where $0 < c < 1$,

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

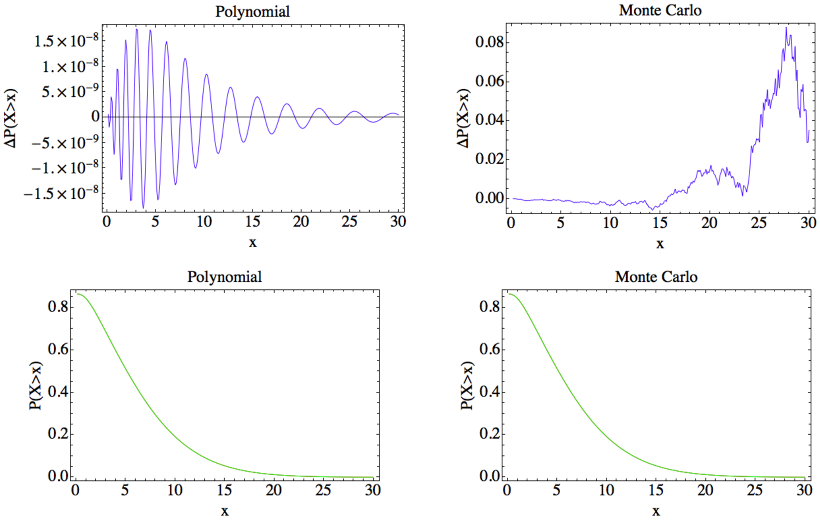
Comparison: Monte Carlo *VS* Polynomial

Example: Compound Poisson distribution $[\mathcal{P}(2), \Gamma(3,1)]$,



- Order of truncation: K=75 $\Rightarrow$ 20 sec,
  $\hookrightarrow$ 600 000 Monte Carlo simulations.

Comparison: Monte Carlo *VS* Polynomials



Statistical application: Definition of the estimator

Let $X_1, \ldots, X_n$ be a sample of size n.

- The polynomial approximation of the PDF is given by

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

- The approximation formula turns into a PDF estimator

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

The statistical inference is a two-step procedure,

1. Parametrical estimation of the parameters of the reference probability measure,
2. Nonparametric estimation of the coefficients of the polynomial expansion.

Statistical application: Definition of the estimator

The coefficients of the polynomial are

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

Nonbisasedly estimated by

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

with associated variance denoted by $\sigma_{k,n}^2 = \mathbb{V}(Z_k)$.

- The coefficients of the polynomial expansion are estimated by

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

  where $\mathbf{w} = (w_1, \dots, w_K)$ is a modulator.

  $\hookrightarrow$ Allow an optimisation of the Integrated Mean Squared Error (IMSE).
- We set $K = n$ in what comes next.

Statistical application: IMSE

The Integrated Mean Squared Error is defined as

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

A modified version of the Integrated Mean Squared Error is optimized

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

Statistical application: IMSE

The quantities a_k^2, et $\sigma_{k,n}^2$ are unbiasedly estimated through

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

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

The Integrated Mean Squared Error is estimated by

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

Statistical application: Subset Selection Modulator

$\mathcal{M}_{SSM}$ denotes the Subset Selection class of modulator such that $\mathbf{w} = (1, \ldots, 1, 0, \ldots, 0)$.

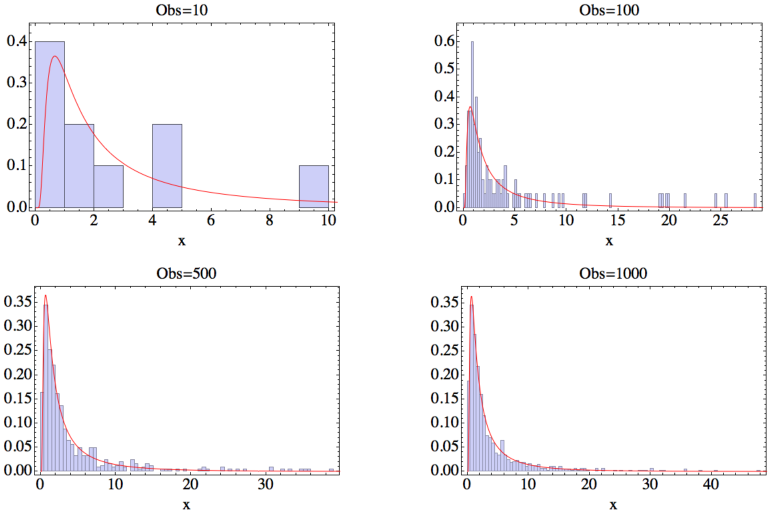
- Computation of the IMSE $\forall \mathbf{w} \in \mathcal{M}_{SME}$.

Illustration: Estimation of the PDF of the *Inverse Gaussian* distribution $\mathcal{IG}(\lambda, \mu)$. The expression of the PDF is given by

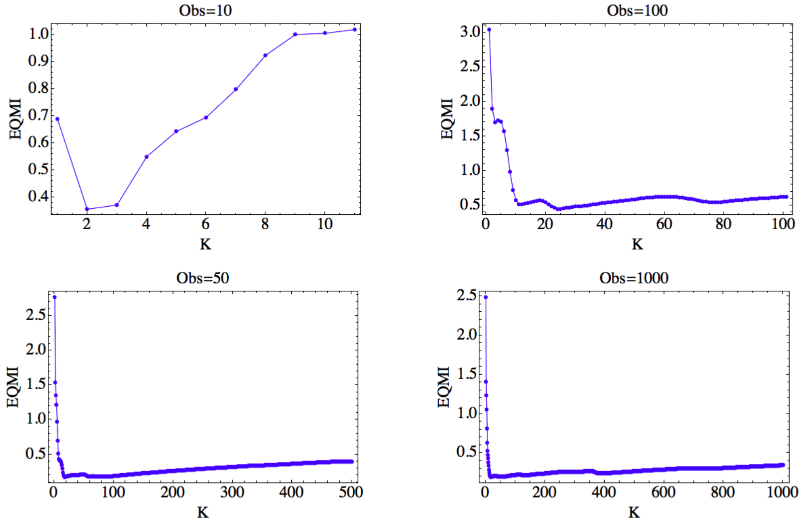
$$f_X(x) = \begin{cases} \dfrac{\sqrt{\frac{\lambda}{x^3}} e^{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}}}{\sqrt{2\pi}}, & x > 0, \\ 0, & \text{Sinon.} \end{cases}$$

- Distribution of the hitting time at a given level of a brownian motion with drift,
- We set $\lambda = 2$, and $\mu = 4$.

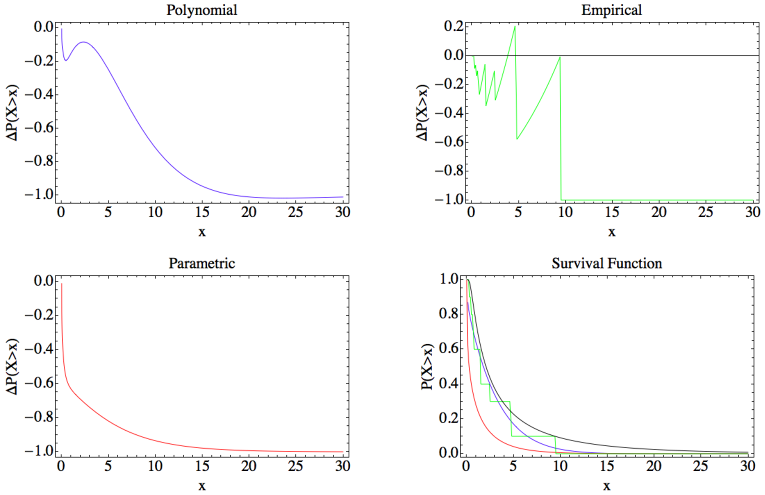
Statistical application: Visualization of the data



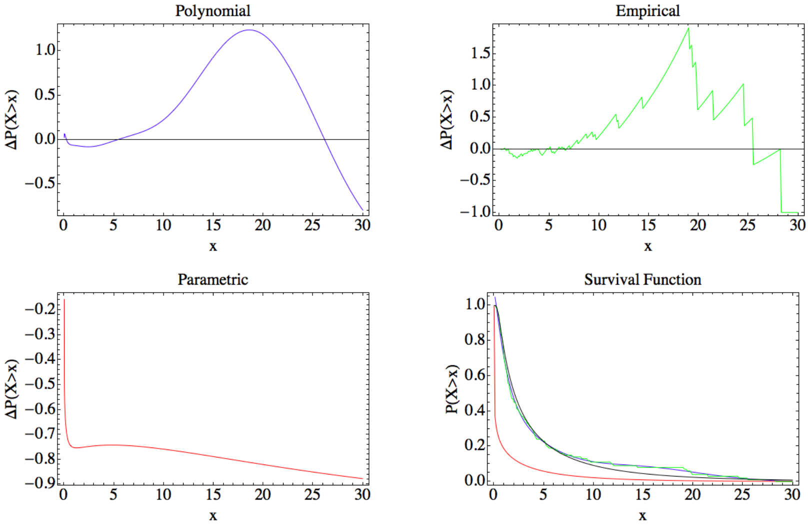
Statitistical application: IMSE



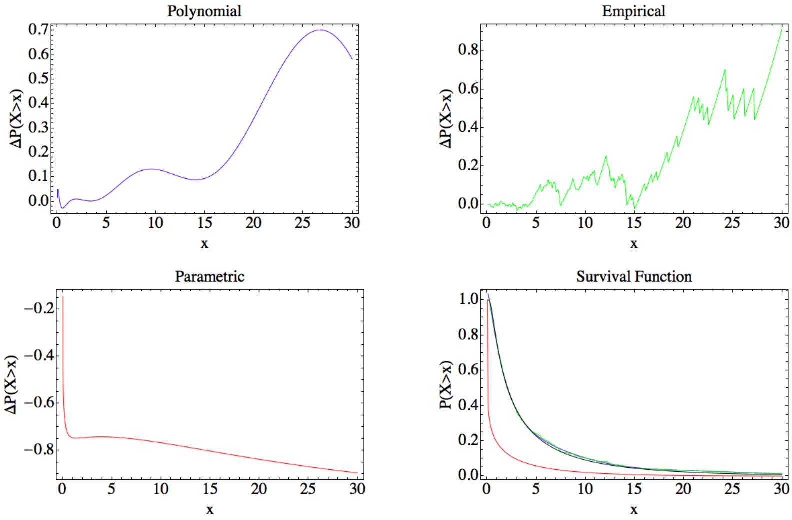
Statistical application: Sample's size=10



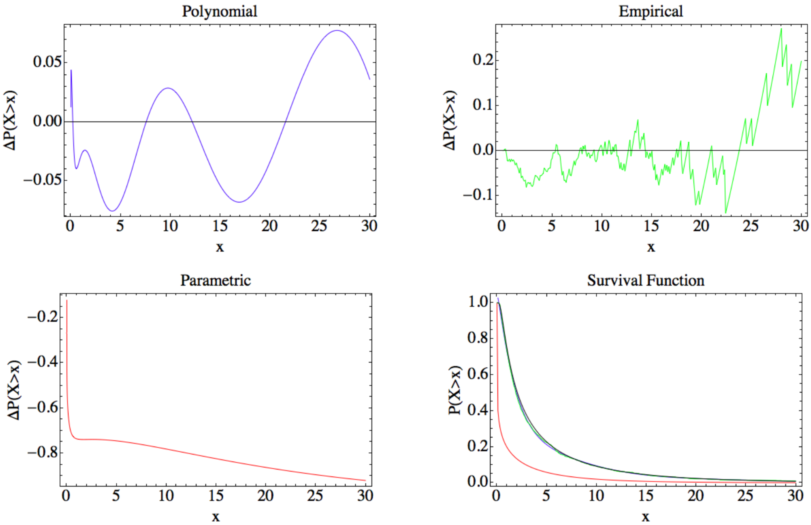
Statistical application: Sample's size=100



Statistical application: Sample's size=500



Statistical application: Sample's size=1 000



Statistical application: Compound distribution

- Aggregate claim amounts sample: $(X_1, \ldots, X_n)$,
- Claim frequency sample: $(N_1, \ldots, N_n)$,
- Claim amounts sample: $(U_1, \ldots, U_{N_1}, \ldots, U_{N_1+\ldots+N_n})$.

The size of the sample available for the aggregate claim amounts and the claim frequencies may be small.

- More observations are available for the claim sizes.

Full Parametric

- Adequacy statistical test to calibrate the model for claim sizes and claim frequency,
- Statistical inference of the parameters,
- Approximation method.

Full NonParametric

- Statistical inference of the parameters of the reference dsitribution,
  - $\hookrightarrow$ Using which sample?
- The coefficient of the polynomial expansion are estimated using $(X_1, \ldots, X_n)$.

Statistical application: Compound distribution

Semi-Parametric

- Parametric model for the claim frequency.
- Statistical inference for the claim frequency model,
  $\hookrightarrow$ We use $(N_1, \ldots, N_n)$.
- The coefficients of the polynomial expansion are estimated using $(U_1, \ldots, U_{N_1}, \ldots, U_{N_1+\ldots+N_n})$, and the recurence relationship that often exists between the moments of X, N, et U.