

Modélisation des incertitudes

Partie 1: modèles univariés

Régis LEBRUN

Airbus Central Research & Technology
regis.lebrun@airbus.com

20 janvier 2020

Plan de l'exposé

- 1 Incertitudes probabilistes
 - Cadre probabiliste et statistique
 - Rappels de probabilité
- 2 Modèles univariés
 - Quantifier les incertitudes
 - Adéquation de lois

Probabilités et statistiques

Différents points de vue

Très schématiquement, les probabilités et les statistiques se différencient par leur rapport au réel :

- Les **statistiques** s'intéressent au recueil d'information qualitative et quantitative sur un phénomène et à la construction d'un modèle probabiliste pertinent pour analyser cette information ;
- Les **probabilités** s'intéressent à la construction d'objets mathématiques abstraits et à leur manipulation.

Les probabilités apparaissent donc comme un creuset de concepts et de règles de calcul permettant aux statistiques de définir des procédures de recueil et d'analyse de l'information efficaces et pertinentes.

Probabilités et statistiques

Différents points de vue

Les probabilités ont deux composantes principales :

- **la partie théorique**, où les questions typiques sont la construction d'objets mathématiques et l'étude de leurs propriétés ;
- **la partie calcul**, où la question est de déterminer les propriétés d'objets obtenus par transformation d'autres objets probabilistes. La conception et l'analyse d'algorithmes stochastiques relèvent essentiellement de cette partie.

Les statistiques ont également deux composantes principales :

- **la partie mathématique**, qui apparaît comme une spécialisation du calcul des probabilités : on définit des modèles probabilistes pour différentes situations concrètes, et on construit des procédures de traitement des données pour paramétrer ces modèles ;
- **la partie expérimentale**, qui établit les procédures concrètes de recueil des données ou de leur communication aux non spécialistes.

Cadre probabilistes

Etant donnés :

- Un vecteur aléatoire \mathbf{X} prenant ses valeurs dans \mathbb{R}^n (les sources d'incertitude)
- Une fonction mesurable $f : \mathbb{R}^n \rightarrow \mathbb{R}^p$ (le modèle numérique)

On souhaite obtenir de l'information sur la distribution de $\mathbf{Y} = f(\mathbf{X})$ (d'où le terme de **propagation des incertitudes**) :

- des moments $\mathbb{E}[h(\mathbf{Y})]$ pour différentes fonctions d'intérêt h
- comme cas particulier, la probabilité d'événements $\mathbb{P}(\mathbf{Y} \in B)$

En pratique, la modélisation probabiliste se résume à :

- Modèle probabiliste = fonction de répartition $F_{\mathbf{X}}$ du vecteur \mathbf{X} .
- On sépare la modélisation des lois marginales (**fonctions de répartition 1D**) et de la structure de dépendance (**copule**)
- Le modèle probabiliste est alors une conséquence du théorème de Sklar : la loi de \mathbf{X} est obtenue par **assemblage** des lois marginales et de la copule.

Variables aléatoires, vecteurs aléatoires I

Définition

Soit (Ω, \mathcal{F}) un espace mesurable et $(\mathbb{R}^n, \mathcal{B})$ l'espace mesurable numérique \mathbb{R}^n muni de la tribu des boréliens. On a :

- Un **vecteur aléatoire** \mathbf{X} **réel** de dimension n est une application de (Ω, \mathcal{F}) dans $(\mathbb{R}^n, \mathcal{B})$ telle que :

$$\forall B \in \mathcal{B}, \mathbf{X}^{-1}(B) \in \mathcal{F}$$

- Une **variable aléatoire** X **réelle** est un vecteur aléatoire de dimension $n = 1$.

Un vecteur aléatoire \mathbf{X} de dimension n **ne se résume pas** à la donnée des n variables aléatoires marginales X_k !

Fonction de répartition I

Définition

Soit X une variable aléatoire réelle. On appelle **fonction de répartition** de X la fonction $F : \mathbb{R} \rightarrow [0, 1]$, parfois notée F_X , définie par $F(x) = \mathbb{P}(X \leq x)$ pour tout $x \in \mathbb{R}$.

Propriété

- F est croissante, continue à droite et vérifie $\lim_{x \rightarrow -\infty} F(x) = \mathbb{P}(X = -\infty)$, $\lim_{x \rightarrow +\infty} F(x) = \mathbb{P}(X = +\infty)$. En particulier, si X est finie p.s. $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow +\infty} F(x) = 1$
- $\forall x \in \mathbb{R}$, la limite à gauche de F en x , $F(x^-) = \lim_{y \rightarrow x^-} F(y)$ existe et vaut $\mathbb{P}(X < x)$. On a ainsi $F(x) - F(x^-) = \mathbb{P}(X = x^-)$.

Théorème (Froda)

Soit F une fonction de répartition univariée. Alors F possède un nombre de points de discontinuité au plus dénombrable.

Fonction de répartition II

Démonstration.

Soit $n \in \mathbb{N}^*$.

- L'ensemble $E_n = \{x \in \mathbb{R} \mid F(x) - F(x-) \geq 1/n\}$ est tel que $\text{card} E_n \leq n$ par croissance de F .
- L'ensemble \mathcal{E} des points de discontinuité de F est tel que $\mathcal{E} \subset \bigcup_{n \in \mathbb{N}^*} E_n$
- \mathcal{E} est donc au plus dénombrable (contenu dans une union dénombrable d'ensembles finis).



Théorème

La fonction de répartition caractérise la loi d'une variable aléatoire : si X et Y sont deux variables aléatoires, elles ont la même loi si et seulement si $F_X = F_Y$.

Fonction quantile

Définition

Soit X une variable aléatoire de fonction de répartition $F : \mathbb{R} \rightarrow [0, 1]$. Sa **fonction quantile** F^{\leftarrow} est définie par :

$$\forall p \in]0, 1], F^{\leftarrow}(p) = \inf\{x \in \mathbb{R} \mid F(x) \geq p\}$$

Propriété

- La fonction F^{\leftarrow} est croissante et continue à gauche, et $\forall x \in \mathbb{R}, \forall p \in]0, 1]$:

$$F(x) \geq p \iff x \geq F^{\leftarrow}(p)$$

- $\forall p \in]0, 1]$, on a $F(F^{\leftarrow}(p)) \geq p$ avec égalité si $F^{\leftarrow}(p) > -\infty$ et si F est continue en $F^{\leftarrow}(p)$.
- Soit U une variable aléatoire de loi uniforme sur $[0, 1]$, la fonction F est la fonction de répartition de la variable aléatoire $F^{\leftarrow}(U)$.
- Soit X une variable aléatoire de fonction de répartition F . Si F est continue, alors la variable aléatoire Y définie par $Y = F(X)$ est de loi uniforme sur $[0, 1]$.

Densité de probabilité

Définition

Soit X une variable aléatoire réelle.

- Elle est dite **absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R}** s'il existe une fonction intégrable $f : \mathbb{R} \rightarrow \mathbb{R}^+$, parfois notée f_X , telle que

$$\forall x \in \mathbb{R}, F(x) = \int_{-\infty}^x f(t) dt$$

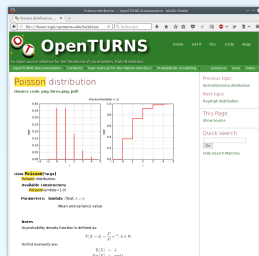
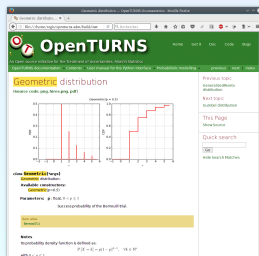
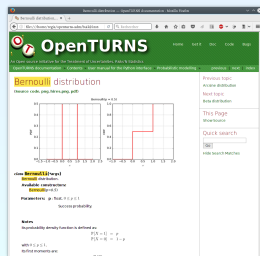
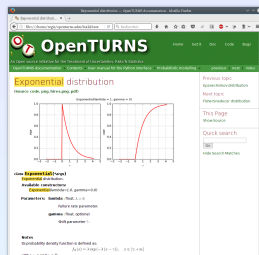
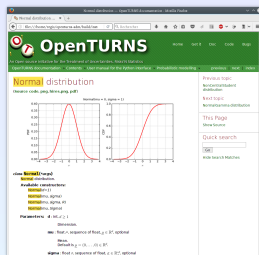
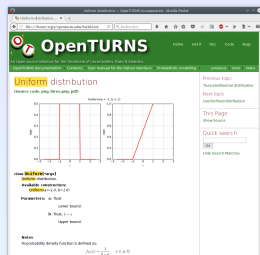
On a alors $\forall x \in \mathbb{R}, f(x) = F'(x)$ presque partout.

- Elle est dite **discrète sur \mathbb{N}** s'il existe une fonction sommable $p : \mathbb{N} \rightarrow [0, 1]$, parfois notée p_X , telle que

$$\forall x \in \mathbb{R}, F(x) = \sum_{k=0}^{\lfloor x \rfloor} p(k)$$

On a alors $\forall k \in \mathbb{N}, p(k) = F(k) - F(k-1)$.

Quelques lois classiques I



Quelques lois classiques II

Listing 1 – drawDistributions.py

```
import openturns as ot

all = [ot.Uniform(), ot.Normal(), ot.Exponential(), \
       ot.Bernoulli(), ot.Geometric(), ot.Poisson()]

for distribution in all:
    ot.Show(distribution.drawPDF())
    ot.Show(distribution.drawCDF())
```

Mixtures

Définition

Soit $F_{\theta} : \mathbb{R} \rightarrow \mathbb{R}^+$ une fonction de répartition où $\theta \in \mathcal{D} \subset \mathbb{R}^p$ est un vecteur de paramètres, et $q : \mathcal{D} \rightarrow \mathbb{R}^+$ une densité de probabilité p -dimensionnelle. On appelle **mixture continue de F_{θ} par rapport à q** la loi de probabilité de fonction de répartition F donnée par :

$$\forall x \in \mathbb{R}, F(x) = \int_{\mathcal{D}} F_{\theta}(x) q(\theta) d\theta$$

Définition

Soit $F_{\theta} : \mathbb{R} \rightarrow \mathbb{R}^+$ une fonction de répartition où $\theta \in \mathcal{D} \subset \mathbb{R}^p$ est un vecteur de paramètres, et $(\theta_n, q_n) \in (\mathcal{D} \times [0, 1])^{\mathbb{N}}$ une loi de probabilité discrète p dimensionnelle. On appelle **mixture discrète de p_{θ} par rapport à (θ_n, q_n)** la loi de probabilité de fonction de répartition F donnée par :

$$\forall x \in \mathbb{R}, F(x) = \sum_{n \in \mathbb{N}} q_n F_{\theta_n}(x)$$

Loi du temps d'attente à un feu de signalisation I

Exercice : soit T le temps d'attente d'un usager à un feu de signalisation de cycle de durée τ , en supposant qu'il se présente devant le feu selon la loi uniforme sur $[0, \tau]$. On suppose que le cycle du feu se décompose en une partie interdisant le passage de durée θ telle que $0 < \theta < \tau$ suivie d'une partie autorisant le passage d'une durée $\tau - \theta$. Quelle est la loi de T ?

Loi du temps d'attente à un feu de signalisation II

On détermine la fonction de répartition de T . Pour celà, soit $D \in [0, \tau]$ la date d'arrivée de l'usager devant le feu, supposée uniforme sur $[0, \tau]$. On a :

$$\begin{aligned}
 \forall t \in \mathbb{R}, \mathbb{P}(T \leq t) &= \mathbb{P}(T \leq t | D \leq \theta) \mathbb{P}(D \leq \theta) + \mathbb{P}(T \leq t | D > \theta) \mathbb{P}(D > \theta) \\
 &= \mathbb{P}(T \leq t, D \leq \theta) + \left(\frac{\tau - \theta}{\tau} \right) \mathbb{1}_{t \geq 0} \\
 &= \mathbb{P}(\theta - t \leq D \leq \theta) + \left(\frac{\tau - \theta}{\tau} \right) \mathbb{1}_{t \geq 0} \\
 &= \frac{\theta}{\tau} \min(t/\theta, 1) \mathbb{1}_{t \geq 0} + \left(1 - \frac{\theta}{\tau} \right) \mathbb{1}_{t \geq 0}
 \end{aligned}$$

Cette fonction de répartition correspond à une mixture entre une loi uniforme sur $[0, \theta]$ de poids $\frac{\theta}{\tau}$ et une loi discrète concentrée en 0 de poids $1 - \frac{\theta}{\tau}$.

Loi du temps d'attente à un feu de signalisation III

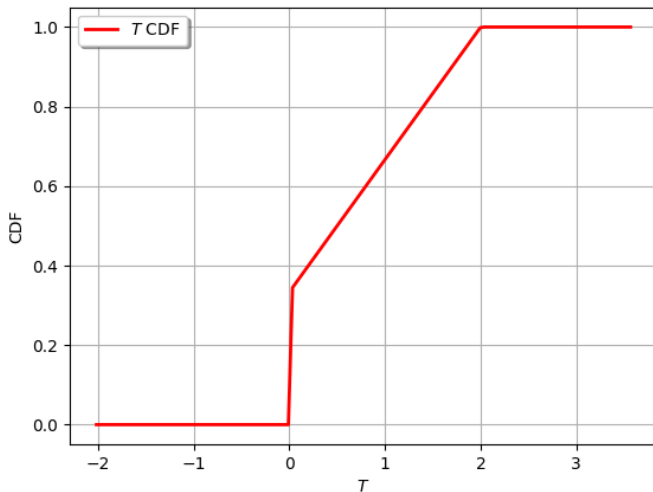
Modélisation avec OpenTURNS :

Listing 2 – feu.py

```
import openturns as ot
import openturns.viewer as otv

tau = 3.0
theta = 2.0
distribution = ot.Mixture([ot.Uniform(0.0, theta), ot.Dirac()], [theta/tau,
    1.0-theta/tau])
distribution.setDescription([r"$T$"])
otv.View(distribution.drawCDF()).save("Feu.png")
```


Loi du temps d'attente à un feu de signalisation IV



Espérance

Soit X une variable aléatoire de fonction de répartition F . L'**espérance** $\mathbb{E}[X]$ de X , si elle existe, est définie par :

$$\mathbb{E}[X] = \int_{\mathbb{R}^d} x F(dx)$$

Dans le cas d'une variable aléatoire discrète, cette définition donne :

$$\mathbb{E}[X] = \sum_{i \in \mathbb{N}} i p(i)$$

et dans le cas absolument continu :

$$\mathbb{E}[X] = \int_{\mathbb{R}^d} x f(x) dx$$

Cette notion correspond à la notion de «moyenne». Si ϕ est une fonction mesurable de \mathbb{R} dans \mathbb{R} , on définit l'espérance $\mathbb{E}[\phi(X)]$ comme étant l'espérance de la variable aléatoire $Y = \phi(X)$.

Moments I

Définition

Soit X une variable aléatoire et $n \in \mathbb{N}^*$ un entier strictement positif. On appelle **moment non centré d'ordre n** de X la quantité $m_n = \mathbb{E}[X^n]$ si cette quantité est bien définie. La quantité m_1 est l'espérance de X .

On définit de même les **moments centrés d'ordre n** de X par $\mu_n = \mathbb{E}[(X - \mathbb{E}[X])^n]$, cette dernière quantité étant bien définie dès que m_n existe. La quantité μ_2 est la **matrice de covariance** de X .

Moments II

Définition

Soit X une variable aléatoire telle que m_2 soit bien défini. On appelle **écart-type (ou standard deviation)** la quantité suivante :

$$\sigma = \sqrt{\mu_2}$$

Soit X une variable aléatoire telle que m_3 soit bien défini. On appelle **coefficient d'asymétrie (ou skewness)** la quantité normalisée suivante :

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

Soit X une variable aléatoire telle que m_4 soit bien défini. On appelle **coefficient d'aplatissement (ou kurtosis)** la quantité normalisée suivante :

$$\beta_2 = \frac{\mu_4}{\sigma^4}$$

L'interprétation usuelle de ces quantités est la suivante :

- L'écart-type décrit la dispersion de la distribution de X .

Moments III

- Si $\gamma_1 < 0$ (resp. $\gamma_1 > 0$), la distribution donne plus de masse aux faibles (resp. fortes) valeurs de X .
- Si $\beta_2 > 3$ (resp. $\beta_2 < 3$), la distribution est plus piquée (resp. plus plate) que la distribution normale.

Ces interprétations sont subjectives et ne forment qu'un résumé grossier de la distribution de X

Moments IV

Calculs de moments avec OpenTURNS :

Listing 3 – moments.py

```
import openturns as ot

distribution = ot.Normal(1.0, 3.0)
print("mean=", distribution.getMean())
print("std =", distribution.getStandardDeviation())
print("skew=", distribution.getSkewness())
print("kurt=", distribution.getKurtosis())
print("E(X^6)=", distribution.getShiftedMoment(6, [0.0]))
print("E((X-3)^6)=", distribution.getShiftedMoment(6, [3.0]))
```

```
mean= [1]
std = [3]
skew= [0]
kurt= [3]
E(X^6)= [14716]
E((X-3)^6)= [27739]
```

Moments V

Exercice : calculer l'espérance, l'écart-type, le coefficient d'asymétrie et le coefficient d'aplatissement d'une variable aléatoire X de loi de Poisson $\mathcal{P}(\lambda)$, ie telle que

$$\forall n \in \mathbb{N}, \quad \mathbb{P}(X = n) = \frac{\lambda^n e^{-\lambda}}{n!}. \text{ Application numérique : } \lambda = 3.2.$$

On utilisera les identités suivantes :

$$(n - \lambda)^2 = n(n - 1) + n(1 - 2\lambda) + \lambda^2$$

$$(n - \lambda)^3 = n(n - 1)(n - 2) + 3(1 - \lambda)n(n - 1) + (1 - 3\lambda + 3\lambda^2)n - \lambda^3$$

$$(n - \lambda)^4 = n(n - 1)(n - 2)(n - 3) + 2(3 - 2\lambda)n(n - 1)(n - 2) + \\ (7 - 12\lambda + 6\lambda^2)n(n - 1) + (1 - 4\lambda + 6\lambda^2 - 4\lambda^3)n + \lambda^4$$

Moments VI

On a, en notant $S = \sum_{n \geq 0} \frac{\lambda^n e^{-\lambda}}{n!} = \sum_{n \geq 0} \mathbb{P}(X = n) = 1$:

$$\mathbb{E}[X] = \sum_{n \geq 0} n \frac{\lambda^n e^{-\lambda}}{n!} = \lambda S = \lambda$$

$$\sigma^2(X) = \lambda^2 S + \lambda(1 - 2\lambda)S + \lambda^2 S = \lambda$$

$$\gamma_1(X) = \frac{\lambda^3 S + 3(1 - \lambda)\lambda^2 S + (1 - 3\lambda + 3\lambda^2)\lambda S - \lambda^3 S}{\lambda^{3/2}} = 1/\sqrt{\lambda}$$

$$\begin{aligned} \beta_2(X) &= \frac{\lambda^4 S + 2(3 - 2\lambda)\lambda^3 S + (7 - 12\lambda + 6\lambda^2)\lambda^2 S + (1 - 4\lambda + 6\lambda^2 - 4\lambda^3)\lambda S + \lambda^4 S}{\lambda^2} \\ &= 3 + 1/\lambda \end{aligned}$$

Moments VII

Application numérique

Listing 4 – momentsExo.py

```
import openturns as ot

d = ot.Poisson(3.2)
print([d.getMean()[0], d.getStandardDeviation()[0], \
       d.getSkewness()[0], d.getKurtosis()[0]])

[3.2, 1.7888543819998317, 0.5590169943749475, 3.3125]
```

Convergences

Définition

Une suite de vecteurs aléatoires $(\mathbf{X}_n)_{n \in \mathbb{N}}$ définis sur un espace probabilisé commun **converge presque sûrement** si l'ensemble des $\omega \in \Omega$ tels que la suite déterministe $(\mathbf{X}_n(\omega))_{n \in \mathbb{N}}$ converge (on note $\mathbf{X}_\infty(\omega)$ cette limite) est de probabilité 1 :

$$\mathbb{P}[\{\omega \mid \lim_{n \rightarrow \infty} \mathbf{X}_n(\omega) = \mathbf{X}_\infty(\omega) \text{ existe}\}] = 1$$

Concrètement, il est impossible d'observer la non convergence d'une suite presque sûrement convergente. On note $\mathbf{X}_n \xrightarrow{p.s.} \mathbf{X}_\infty$ cette convergence.

Définition

Une suite de vecteurs aléatoires $(\mathbf{X}_n)_{n \in \mathbb{N}}$ définis sur des espaces probabilisés éventuellement différents **converge en loi** vers un vecteur aléatoire \mathbf{X}_∞ si et seulement si pour toute fonction mesurable bornée g de \mathbb{R}^d dans \mathbb{R} on a :

$$\lim_{n \rightarrow \infty} \mathbb{E}[g(\mathbf{X}_n)] = \mathbb{E}[g(\mathbf{X}_\infty)]$$

Concrètement, le calcul d'une moyenne réalisé à l'aide de la loi de \mathbf{X}_n est d'autant plus proche de celui réalisé à l'aide de la loi de \mathbf{X}_∞ que n est grand.

On note $\mathbf{X}_n \xrightarrow{\mathcal{L}} \mathbf{X}_\infty$ cette convergence.

Quantifier les incertitudes

Objectif

L'**objectif de cette étape** méthodologique est la **construction du modèle probabiliste** associé aux incertitudes du vecteur de paramètres incertains qui a été identifié.

Cette étape utilise de manière importante les **techniques d'estimation et de test statistiques** dans le cas fortuné où l'on dispose de mesures acquises selon un processus expérimental formalisé. Dans le cas où ces mesures font défaut, d'autres techniques mathématiques permettent de construire un **modèle le plus raisonnable possible en présence de l'information disponible**.

L'ingénieur incertitude engage pleinement sa responsabilité lors de cette étape de quantification. La qualité de l'étude d'incertitudes dépend de la qualité du modèle probabiliste de manière aussi importante que de la qualité du modèle numérique décrivant le système étudié.

Ajustement de lois

Selon la quantité et la nature des données disponibles, la quantification se fait :

- à l'aide des **outils de la statistique** s'il y a assez de données *statistiques*,
- à l'aide de l'**avis d'expert** en l'absence de données *statistiques*,
- à l'aide du **principe du Maximum d'Entropie**.

Mais ...

- ❶ la présence de données même massives n'interdit pas les avis d'expert : le mélange des types d'informations (avis d'expert + données) mène aux techniques d'ajustement paramétriques,
- ❷ l'avis d'expert est toujours accompagné de quelques données : calculs selon des scénarios particuliers (en moyenne, dans les extrêmes, ..) et le choix de la loi se base sur l'ensemble des informations (données + avis d'expert).

Adéquation de loi sur données

Deux étapes : ajustement + validation

La présence de données permet d'utiliser les outils classiques de la statistique. Les étapes d'une adéquation de loi sur données sont :

Etape 1 : Ajustement d'une loi :

- 1a - ajustement paramétrique : suppose une forme de loi et calcule au mieux ses paramètres par la méthode du maximum de vraisemblance ou des moments : calculs d'estimateurs des paramètres,
- 1b - par la méthode de reconstruction à noyaux : suppose l'existence d'une densité pour la loi à identifier, et construit une approximation de cette densité sans supposer de forme à priori.

Etape 2 : Validation de l'ajustement :

- Test numérique d'adéquation,
- Validation graphique.

Etape 1a - Ajustement paramétrique sur données I

Estimateur : Définition

Un **estimateur** θ_n du paramètre θ de la densité p_θ de la variable aléatoire X est une fonction des réalisations (x_1, \dots, x_n) de X : $\theta_n = f(x_1, \dots, x_n)$.

Etape 1a - Ajustement paramétrique sur données II

Exemples d'estimateurs

Estimateur du Maximum de vraisemblance : la vraisemblance d'un échantillon est $\prod_{i=1}^n p_{\theta}(x_i)$. Le principe est de choisir le paramètre θ qui maximise la vraisemblance de l'échantillon, ie de choisir le θ qui rende *maximale la probabilité d'apparition de l'échantillon obtenu*.

Estimateur de la Méthode des moments : le principe est d'exprimer les premiers moments de la loi de X en fonction des paramètres de la loi, puis d'inverser ces relations et d'estimer les premiers moments à l'aide d'estimateurs convergents possédant de bonnes propriétés (moyenne empirique, ...).

Loi	Maximum de vraisemblance	Méthode des moments
$\mathcal{N}(\mu, \sigma)$	$\mu_n = 1/n \sum x_i$ $\sigma_n^2 = 1/n \sum (x_i - \mu_n)^2$	idem idem
$\mathcal{E}(\lambda)$	$\lambda_n = n / \sum x_i$	idem
$\mathcal{U}(a, b)$	$a_n = \min(x_i)$ $b_n = \max(x_i)$	$a_n = \mu_n - \sqrt{3}\sigma_n$ $b_n = \mu_n + \sqrt{3}\sigma_n$

Etape 1a - Ajustement paramétrique sur données III

Propriétés d'estimateurs

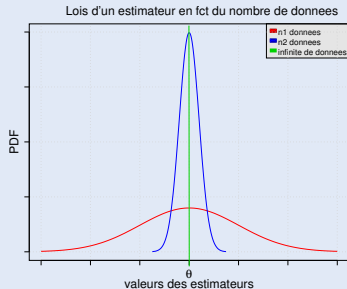
Convergence : $\lim_n \theta_n = \theta$ presque sûrement.

⇒ Plus il y a de données, plus la valeur de l'estimateur est proche de la valeur réelle.

Asymptotiquement normal : $\sqrt{n}(\theta_n - \theta) \xrightarrow{loi,n} \mathcal{N}(0, \sigma^2)$

⇒ calcul des intervalles de confiance : l'égalité $\alpha = P(|\frac{\sqrt{n}(\theta_n - \theta)}{\sigma}| < a) = P(|\mathcal{N}(0, \sigma^2)| < a)$ mène à $\theta \in [\theta_n - \frac{\sigma_n}{\sqrt{n}} a_\alpha, \theta_n + \frac{\sigma_n}{\sqrt{n}} a_\alpha]$ avec $a_\alpha = \Phi^{-1}((1 - \alpha)/2)$ (où l'on a remplacé la variance asymptotique σ^2 par la variance empirique σ_n^2 calculée à partir de l'échantillon).

- Remarque 1 : Influence du nombre de données sur la dispersion de l'estimateur (sans biais)



$$\begin{cases} n_1 \text{ grand,} & \theta_{n_1} \sim \mathcal{N}(\theta, \sigma^2/n_1) \\ n_2 \geq n_1, & \theta_{n_2} \sim \mathcal{N}(\theta, \sigma^2/n_2) \\ n_\infty, & \theta_{n_\infty} \equiv \theta \end{cases}$$

CI : Plus on a de données, plus la loi de θ_n sera resserrée autour de θ (ie plus la dispersion est faible) et plus la valeur de l'estimateur a de chances d'être proche de la valeur réelle.

Illustration avec OpenTURNS

Listing 5 – distributionFit.py

```
import openturns as ot

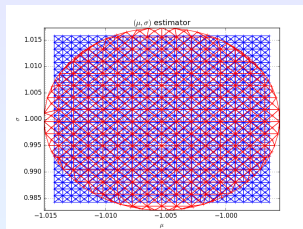
data = ot.Sample.ImportFromCSVFile('data.csv')
print(data)

# Maximum likelihood estimation
estimated_distribution = ot.UniformFactory().build(data)
print('Estimated uniform=', estimated_distribution)

# Best distribution wrt an entropy criterion
models = [ot.UniformFactory(), ot.NormalFactory(), ot.ExponentialFactory()]
best_distribution = ot.FittingTest.BestModelBIC(data, models)
print('Selected distribution=', best_distribution)

[ X0
0 : [ -0.756719 ]
1 : [ -1.50647 ]
2 : [ -1.17531 ]
...
9997 : [ -1.50181 ]
9998 : [ -0.788409 ]
9999 : [ -0.63771 ]
Estimated uniform= Uniform(a = -2.42012, b = 0.42403)
Selected distribution= Normal(mu = -1.00532, sigma = 0.400061)
```

Illustration avec OpenTURNS



Listing 6 – distributionIC.py

```
import openturns as ot

data = ot.Sample.ImportFromCSVFile('data.csv')
# Maximum likelihood estimation
estimation_result = ot.NormalFactory().buildEstimator(data)
print('Estimated distribution=', estimation_result.getDistribution())
param_dist = estimation_result.getParameterDistribution()
print('Estimator distribution=', param_dist)
RC_95 = param_dist.computeMinimumVolumeLevelSet(0.95)
print('RC 95%=\n', RC_95)
IC_95 = param_dist.computeMinimumVolumeInterval(0.95)
print('IC 95%=\n', IC_95)
ot.ResourceMap.SetAsBool('IntervalMesher-UseDiamond', True);
graph_RC_95 = ot.LevelSetMesher([50]*2).build(RC_95, param_dist.getRange())
```

Etape 1b : Ajustement non paramétrique

L'objectif est d'évaluer la densité de probabilité à partir d'un échantillon (X_1, \dots, X_n) du vecteur aléatoire \mathbf{X} .

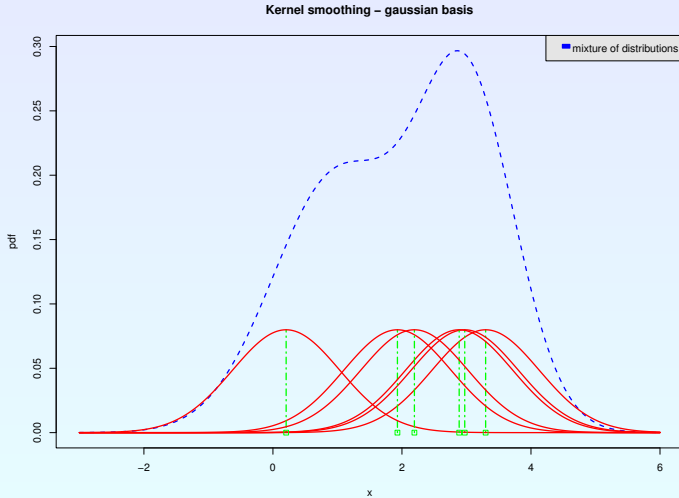
Reconstruction à noyaux : cas 1D

On approche la densité p par la densité empirique p_n :

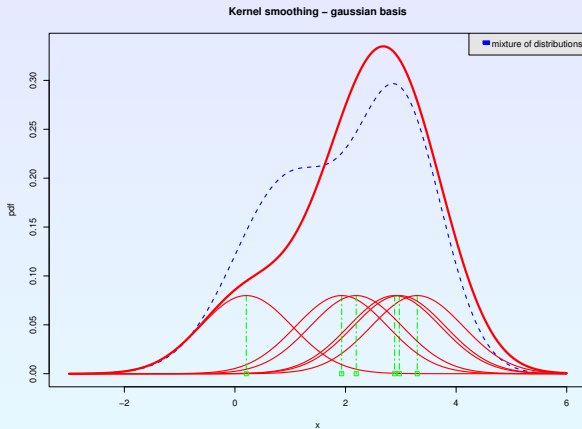
$$p_n(x) = \frac{1}{nh} \sum_{i=1}^{i=n} K\left(\frac{x - X_i}{h}\right) \quad (1)$$

- K : noyau, loi symétrique par rapport à 0 . Par exemple, gaussien $(0,1)$, uniforme $(-1,1)$, ...
- h : largeur de bande.

Etape 1b : Ajustement non paramétrique



Etape 1b : Ajustement non paramétrique

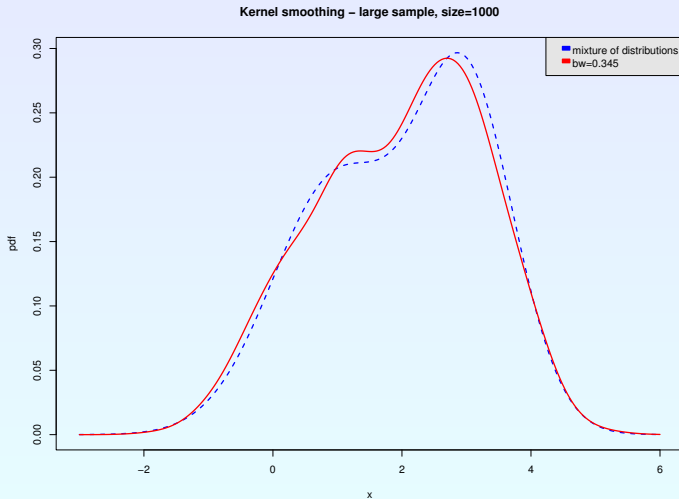


Listing 7 – kernelsmoothing1.py

```
import openturns as ot

data = [[0.2], [1.96], [2.08], [3.01], [3.1], [3.6]]
estimated_distribution = ot.KernelSmoothing().build(data)
```

Etape 1b : Ajustement non paramétrique



Etape 1b : Ajustement non paramétrique

Reconstruction à noyaux

L'erreur de modélisation est mesurée par l'**indicateur AMISE** :

$$\begin{cases} \mathbf{AMISE}(\hat{p}) &= \text{2 termes dominants ds le dll en } n \text{ de } \mathbf{MISE}(\hat{p}) \\ \mathbf{MISE}(\hat{p}) &= \mathbb{E}_X[\|\hat{p} - p\|_{L_2}^2] = \int \mathbf{MSE}(\hat{p}, x) dx \\ \mathbf{MSE}(\hat{p}, x) &= [\mathbb{E}_X[\hat{p}(x)] - p(x)]^2 + \mathbf{Var}_X(\hat{p}(x)) = \text{risque quadratique en } x \end{cases}$$

Type de noyau peu influent sur l'erreur AMISE \Rightarrow **noyau normal**.

Sélection de la largeur de bande : h grand : données noyées (*oversmoothing*), h petit : données séparées (*undersmoothing*) \Rightarrow le h optimal au sens AMISE est (en dimension 1) :

$$h_{\mathbf{AMISE}}(K) = \left[\frac{R(K)}{\mu_2(K)^2 \Phi_4} \right]^{\frac{1}{5}} n^{-\frac{1}{5}} \quad (2)$$

où $\Phi_r = \int p^{(r)} p(x) dx = \mathbb{E}_X[p^{(r)}]$ et $\mu_2(K) = \int x^2 K(x) dx = \sigma_K^2$.

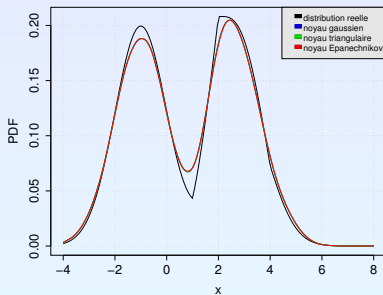
Plusieurs règles pour évaluer le terme Φ_4 et donc le h optimal : **Règle de Silverman**, **Règle de Scott**, **Méthode de Plug-in** : *Solve-the-equation*.

L'intérêt d'une reconstruction à noyaux, outre d'être non paramétrique, est de pouvoir s'ajuster à des lois multimodales, contrairement aux lois *usuelles* (*unimodales*).

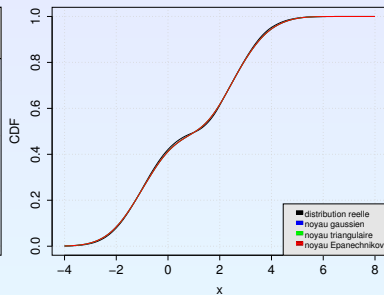
Etape 1b : Ajustement non paramétrique

Impact du choix du noyau

Reconstruction a noyaux – PDF



Reconstruction a noyaux – CDF



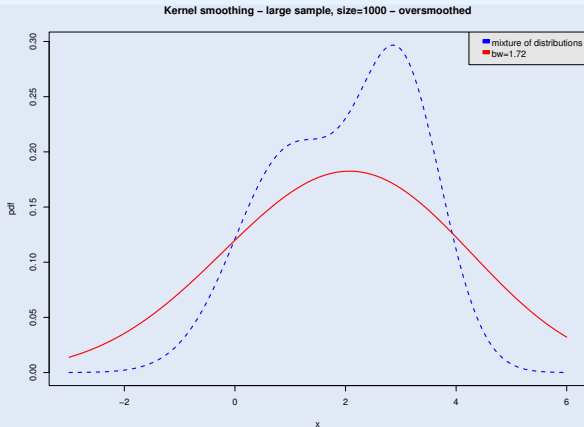
Listing 8 – kernelsmoothing2.py

```
import openturns as ot

data = [[0.2], [1.96], [2.08], [3.01], [3.1], [3.6]]
est_dist = ot.KernelSmoothing(ot.Normal()).build(data)
est_dist = ot.KernelSmoothing(ot.Triangular()).build(data)
est_dist = ot.KernelSmoothing(ot.Epanechnikov()).build(data)
```

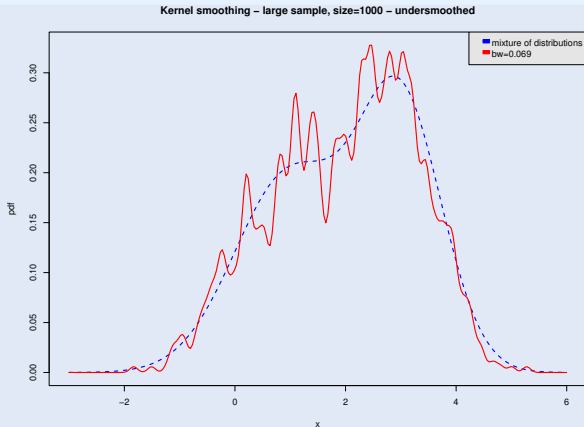

Etape 1b : Ajustement non paramétrique

Impact de la largeur de bande



Etape 1b : Ajustement non paramétrique

Impact de la largeur de bande



Etape 1b : Ajustement non paramétrique

Sélection de h : Règle de Silverman

Dans le cas où la densité p est normale de variance σ^2 , alors le terme Φ_4 peut se calculer exactement et la largeur de bande optimale s'écrit :

$$h_{AMISE}^{p=normal}(K) = \left[\frac{8\sqrt{\pi}R(K)}{3\mu_2(K)^2} \right]^{\frac{1}{5}} \sigma n^{-\frac{1}{5}} \quad (3)$$

Un estimateur de $h_{AMISE}^{p=normal}(K)$ est obtenu en remplaçant σ par un estimateur $\hat{\sigma}^n$ évalué sur l'échantillon (X_1, \dots, X_n) .

La **Règle de Silverman** consiste à prendre $\hat{h}_{AMISE}^{p=normal}(K)$ même si la densité p n'est pas normale :

$$h^{Silver}(K) = \left[\frac{8\sqrt{\pi}R(K)}{3\mu_2(K)^2} \right]^{\frac{1}{5}} \hat{\sigma}^n n^{-\frac{1}{5}} \quad (4)$$

La Règle de Silverman est *empirique* et donne de bons résultats lorsque la densité n'est pas trop *éloignée* de celle d'une normale (par exemple : cas d'une loi bimodale ou disymétrique).

Etape 1b : Ajustement non paramétrique

Sélection de h : Règle de Scott

La Règle de Scott est une simplification de la Règle de Silverman évaluée avec un noyau normal (4) :

$$h^{Silver}(N) \simeq \hat{\sigma}^n n^{-\frac{1}{5}} \quad (5)$$

De plus, en remarquant que $\sigma_K R(K) \simeq 1 \forall K$, (2) mène à :

$$\frac{h_{AMISE}(K_1)}{h_{AMISE}(K_2)} = \frac{\sigma_{K_2}}{\sigma_{K_1}} \left[\frac{\sigma_{K_1} R(K_1)}{\sigma_{K_2} R(K_2)} \right]^{\frac{1}{5}} \simeq \frac{\sigma_{K_2}}{\sigma_{K_1}} \quad (6)$$

En prenant $K_2 = N(0, 1)$ dans (6) et en approchant $h_{AMISE}(N)$ par la Règle de Silverman, on obtient :

$$h^{Silver}(K) \simeq h^{Silver}(N) \frac{1}{\sigma_K} \quad (7)$$

La **Règle de Scott** consiste à prendre le $h^{Silver}(K)$ approché des relations (7) et (5) même si p n'est pas normale :

$$h^{Scott} = \frac{\hat{\sigma}^n}{\sigma_K} n^{-\frac{1}{5}} \quad (8)$$

La Règle de Scott est *empirique* et donne de bons résultats lorsque la densité n'est pas trop éloignée de celle d'une normale (par exemple : cas d'une loi bimodale ou dissymétrique).

Etape 1b : Ajustement non paramétrique

Sélection de h : Méthode du plug-in - *Solve-the-equation*

La largeur de bande optimale $h_{AMISE}(K)$ de (2) dépend de Φ_4 : $h_{AMISE}(K) = t(\Phi_4)$.
Le terme $\Phi_4 = \int p^{(4)} p(x) dx = \mathbb{E}_X[p^{(4)}]$ est approché par son estimateur :

$$\hat{\Phi}_4 = \frac{1}{n} \sum_{i=1}^n \hat{p}^{(4)}(X_i) \quad (9)$$

En dérivant la relation (1), on obtient : $\hat{p}^{(4)}(x) = \frac{1}{nh^5} \sum_{i=1}^n K^{(4)}\left(\frac{x - X_i}{h}\right)$.

L'estimateur $\hat{\Phi}_4$ a pour expression : $\hat{\Phi}_4(h) = \frac{1}{n^2 h^5} \sum_{i=1}^n \sum_{j=1}^n K^{(4)}\left(\frac{X_i - X_j}{h}\right)$.

Le paramètre h peut être pris afin de minimiser le critère AMSE (Asymptotic Mean Square Error) entre Φ_4 and $\hat{\Phi}_4(h)$:

$$h_{AMSE}^{(4)} = \left(\frac{-2K^{(4)}(0)}{\mu_2(K)\Phi_6} \right)^{\frac{1}{7}} n^{-\frac{1}{7}} \quad (10)$$

On a donc la relation : $h_{AMISE}(K) \simeq t \circ \hat{\Phi}_4(h_{AMSE}^{(4)})$.

Etape 1b : Ajustement non paramétrique

Sélection de h : Méthode du plug-in - *Solve-the-equation*

Les relations (2) and (10) mènent à : $h_{AMSE}^{(4)} = \left(\frac{-2K^{(4)}(0)\mu_2(K)\Phi_4}{R(K)\Phi_6} \right)^{\frac{1}{7}} h_{AMISE}(K)^{\frac{5}{7}}$, qui s'écrit $h_{AMSE}^{(4)} = I(h_{AMISE}(K))$. On a donc, au final :

$$h_{AMISE}(K) = t \circ \hat{\Phi}_4 \circ I[h_{AMISE}(K)] \quad (11)$$

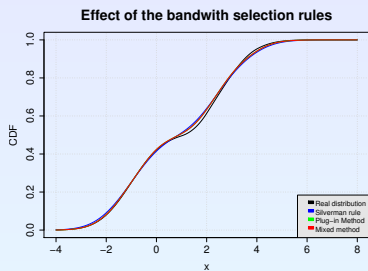
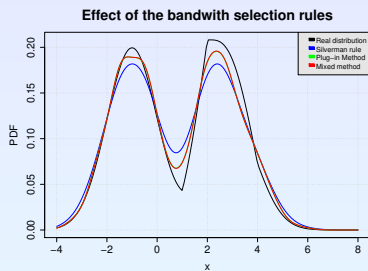
Le maillon I fait intervenir les termes Φ_4 et Φ_6 qui sont évalués via leurs estimateurs du type (9), définis à partir de leur paramètre optimal respectif (pour AMSE) g_1 et g_2 :

$$\begin{cases} g_1 &= \left(\frac{-2K^{(4)}(0)}{\mu_2(K)\Phi_6} \right)^{\frac{1}{7}} n^{-\frac{1}{7}} \\ g_2 &= \left(\frac{-2K^{(6)}(0)}{\mu_2(K)\Phi_8} \right)^{\frac{1}{7}} n^{-\frac{1}{9}} \end{cases} \quad \text{avec} \quad \begin{cases} \hat{\Phi}_6 &= \frac{-15}{16\sqrt{\pi}} \hat{\sigma}^{-7} \\ \hat{\Phi}_8 &= \frac{105}{32\sqrt{\pi}} \hat{\sigma}^{-9} \end{cases} \quad (12)$$

Pour évaluer Φ_6 et Φ_8 dans les expression de g_1 et g_2 , on suppose que la densité p est normale de variance σ^2 , approchée par la variance empirique $\hat{\sigma}^2$ de l'échantillon

Etape 1b : Ajustement non paramétrique

Impact de la largeur de bande



Listing 9 – kernelsmoothing3.py

```
import openturns as ot

data = [[0.2], [1.96], [2.08], [3.01], [3.1], [3.6]]
bw1 = ot.KernelSmoothing().computeSilvermanBandwidth(data)
bw2 = ot.KernelSmoothing().computePluginBandwidth(data)
bw3 = ot.KernelSmoothing().computeMixedBandwidth(data)
est_dist1 = ot.KernelSmoothing().build(data, bw1)
est_dist2 = ot.KernelSmoothing().build(data, bw2)
est_dist3 = ot.KernelSmoothing().build(data, bw3)
```

Etape 1b : Ajustement non paramétrique

Reconstruction à noyaux : effet de bord

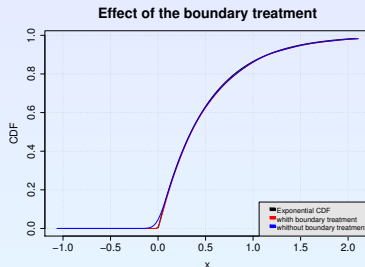
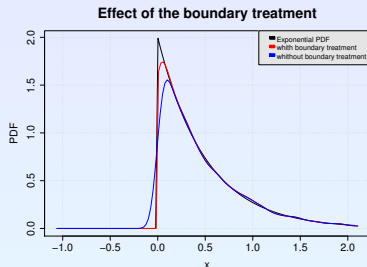
Lorsque la loi à reconstruire a un support borné avec une densité non nulle aux bornes, les lois reconstruites à noyaux peuvent avoir un support plus large (noyaux à supports finis : uniforme, triangulaire), voire infini (noyau à support infini : gaussien, ...), et si la densité à reconstruire n'est pas continue (ce qui est souvent le cas en bords de domaine) alors la reconstruction à noyaux converge vers la demi-hauteur de la discontinuité \implies les bords doivent être traités de manière spécifique.

La technique du *mirroring* permet de compenser à faible coût ce défaut (uniquement en dimension 1) :

- la largeur de bande h est évaluée à partir de l'échantillon
- on extrait de l'échantillon les points situés aux extrêmes : $[min, min + h[$ et $]max - h, max]$,
- on symétrise ces 2 échantillons extraits par rapport à leur borne propre : $]min - h, min]$ and $[max, max + h[$,
- on reconstruit la densité à partir de l'échantillon initial augmenté des 2 échantillons précédents, avec la largeur de bande initiale h ,
- on tronque la densité obtenue à l'intervalle $[min, max]$

Etape 1b : Ajustement non paramétrique

Reconstruction à noyaux : effet de bord



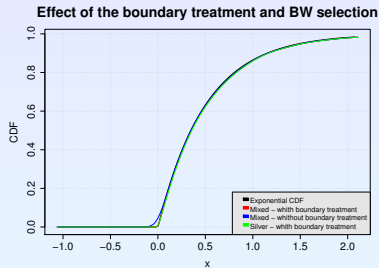
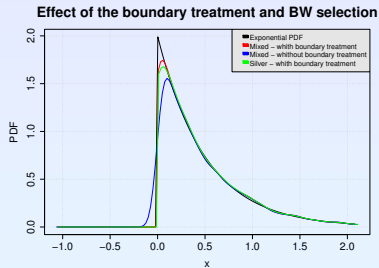
Listing 10 – kernelsmoothing4.py

```
import openturns as ot

data = [[0.2], [1.96], [2.08], [3.01], [3.1], [3.6]]
ks = ot.KernelSmoothing()
est_dist1 = ks.build(data)
ks.setBoundaryCorrection(True)
est_dist2 = ks.build(data)
```

Etape 1b : Ajustement non paramétrique

Le traitement des bords prévaut sur la règle de sélection de la largeur de bande



Listing 11 – kernelsmoothing5.py

```
import openturns as ot
data = [[0.2], [1.96], [2.08], [3.01], [3.1], [3.6]]
ks = ot.KernelSmoothing()
bw3 = ks.computeMixedBandwidth(data)
est_dist2 = ks.build(data, bw3)
ks.setBoundaryCorrection(True)
est_dist1 = ks.build(data, bw3)
bw1 = ks.computeSilvermanBandwidth(data)
est_dist3 = ks.build(data, bw1)
```

Etape 1b : Ajustement non paramétrique

Les lois multimodales ... ça existe !

Considérons le modèle suivant : $y = f(x) = \arctan(x)$ où la seule entrée incertaine est $x \in \mathbb{R}$.

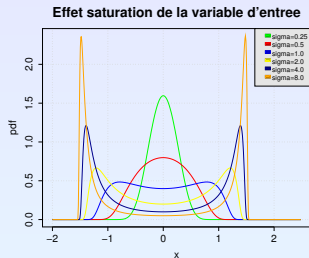
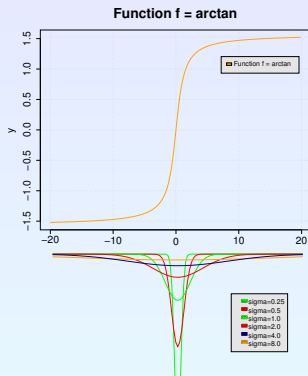
Supposons que l'incertitude de x est modélisée par une loi **Normal**(0, σ).

Selon la dispersion σ de x , différents intervalles vont être visités par la variable d'intérêt y et sa loi de probabilité va complètement changer de forme

Quelques remarques :

- selon les valeurs des entrées, la physique f passe d'un mode de fonctionnement à un autre : ici, de l'état -1.5 à l'état 1.5
- si la fonction f ne présente pas de symétrie par rapport à un point (ici (0,0)), la densité de la variable d'intérêt y n'est pas symétrique : on obtient une densité en cuvette non symétrique ... quid de l'ajustement paramétrique ?
- attention aux études de sensibilité qui font varier σ sans changer les ajustements paramétriques validés pour quelques valeurs de σ

Etape 1b : Ajustement non paramétrique



Listing 12 – atanNormal.py

```
import openturns as ot

sigma = 0.5
distribution = ot.Normal(0.0, sigma).atan()
ot.Show(distribution.drawPDF())
```

Etape 1b : Ajustement non paramétrique

Reconstruction à noyaux : cas nD

En dimension d , on peut utiliser un **noyau produit**.

Si $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ où $\mathbf{X}_i \in \mathbb{R}^d$ est l'échantillon en dimension d ,

$$K_d(\mathbf{x}) = \prod_{j=1}^{j=d} K(x^j)$$

ce qui mène à l'expression de la densité p_n :

$$p_n(\mathbf{x}) = \frac{1}{n \prod_{j=1}^{j=d} h_j} \sum_{i=1}^n K_d \left(\frac{x_1 - X_{i1}}{h_1}, \dots, \frac{x_d - X_{id}}{h_d} \right)$$

Le vecteur des largeur de bandes $\mathbf{h} = (h^1, \dots, h^d)$ peut être déterminé par la **Règle de Scott** :

$$h_{Scott}^i = \frac{\hat{\sigma}_i^n}{\sigma_K} n^{-1/(d+4)}$$

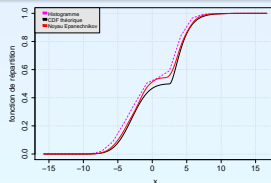
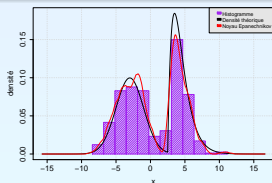
où $\hat{\sigma}_i^n$ est l'écart-type empirique de la composante i de l'échantillon.

Etape 1b : Ajustement non paramétrique

Comparaison Reconstruction à noyaux / Histogramme (cas 1D)

Méthode de reconstruction	h opt. selon AMISE (Silv.)	Erreur AMISE
Histogramme	$\propto \frac{1}{n^{1/3}}$	$\propto \frac{1}{n^{2/3}}$
Reconstruction à noyaux (K)	$\propto \frac{1}{n^{1/5}}$	$\propto \frac{1}{n^{4/5}}$

Conclusion : Une reconstruction par noyaux est préférable à un histogramme !



Listing 13 – ksHistogram.py

```
import openturns as ot
data = [[0.2], [1.96], [2.08], [3.01], [3.1], [3.6]]
histogram = ot.HistogramFactory().build(data)
kernel = ot.KernelSmoothing().build(data)
```

Ajustement paramétrique vs non paramétrique

Influence du nombre de points sur un ajustement

Attention, remarquons que :

- moins il y a de points dans l'échantillon, moins l'ajustement est robuste (trop adhérent à l'échantillon utilisé)
- les tests d'ajustement sont des tests asymptotiques qui reposent sur la convergence de la variable de décision vers une loi connue
- un ajustement paramétrique sera toujours de meilleure qualité qu'un ajustement non paramétrique **à condition d'être certain du modèle proposé !**

Etape 2 : Validation de la loi

Validation de la loi

Plusieurs méthodes sont possibles :

- Validation graphique
- Test numérique d'adéquation

Validation graphique

Plusieurs tests graphiques sont possibles :

- test du QQ-plot, notamment pour la loi normale (droite de Henry) , la loi de Weibull
- tracé de la pdf superposé à l'histogramme ou la reconstruction à noyaux
- ...

cf. la classe `VisualTest` d'OpenTURNS.

Etape 2 : Validation de la loi

Test numérique d'adéquation

On confronte 2 hypothèses H_0 et H_1 . On distingue les erreurs :

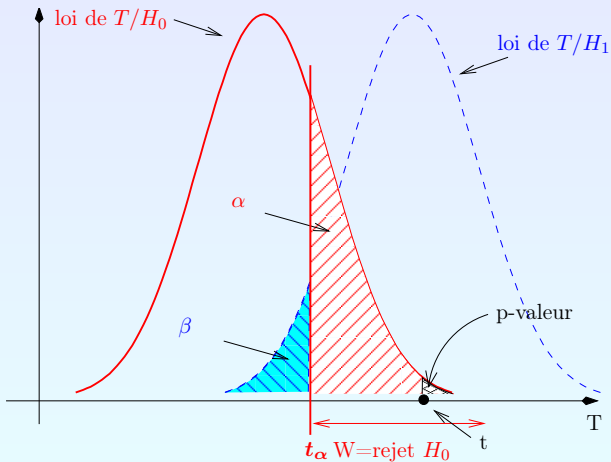
- **Erreur de 1^{ère}-espèce** : α : probabilité (rejeter H_0 à tort).
- **Erreur de 2^{ème}-espèce** : β : probabilité (rejeter H_1 à tort).

Généralement, on met dans H_0 une hypothèse forte : on redoute plus l'erreur α que l'erreur β . Ainsi, on impose α et on subit β .

Déroulement du test

- on introduit une **variable de décision** (statistique) T sur l'échantillon (x_1, \dots, x_n) , dont on connaît la loi,
- on définit une **région critique** W de rejet de H_0 par : $P(W/H_0) = \alpha$. Par exemple $W = \{(\mathbf{X}_1, \dots, \mathbf{X}_n) / T(\mathbf{X}) \geq t_\alpha\}$,
- on évalue la statistique T sur l'échantillon : $t = T(x_1, \dots, x_n)$ et on compare t au **seuil de rejet** t_α de H_0 .

Etape 2 : Test d'adéquation



Etape 2 : Test d'adéquation

Quelques tests classiques d'adéquation de loi

- en région centrale : test du χ^2 , de Kolmogorov - Smirnov, ...
- en queue de distribution : test Anderson Darling, de Cramer, ...

La variable de décision mesure l'écart entre la fonction de répartition empirique est celle proposée :

- pour Kolmogorov-Smirnov : $\sqrt{n} \sup_{y \in \mathbb{R}} |F_n(y) - F(y)|$ tend vers la loi de Kolmogorov (tabulée),
- pour le test du χ^2 , $\zeta_n^{(2)} = n \sum_{i=1}^{i=m} \frac{(\hat{p}_i - p_i^0)^2}{p_i^0}$ tend vers la loi du $\chi^2(m-1)$.

Si plusieurs tests passent ?

On peut les hiérarchiser par des critères quantitatifs :

- **p - valeur** : $P(T(X) \geq T(x_1, \dots, x_n))$ à comparer à α ,
- **BIC** (Bayesian Information Criteria) : calcul basé sur l'évaluation de la log-de vraisemblance l'échantillon.

cf. les classes FittingTest, NormalityTest d'OpenTURNS.

Ajustement sur dire d'expert

Interrogation d'expert

L'absence ou la rareté de données requiert d'**interroger un expert** et d'**interpréter** les informations qu'il donne sous forme de loi de probabilité. Se contenter de 3 questions :

- **Question 1** : Existe-t-il un historique d'utilisation d'une loi particulière ?
- **Question 2** : Existe-t-il un intervalle de variation dans lequel la variable doit rester ?
- **Question 3** : Existe-t-il des valeurs remarquables de la variable à faire ressortir ?

Choix d'une loi

Plusieurs stratégies sont possibles pour choisir une loi à partir de l'information apportée par l'expert :

- choix à partir d'un **organigramme**, reposant sur des considérations simples, de bon sens
- utilisation du **principe du maximum d'entropie**, reposant sur un calcul d'entropie

Ajustement sur dire d'expert

Organigramme de choix de loi

Rep. Q1	Rep. Q2	Rep. Q3	Loi préconisée
Non	Oui : $[a, b]$	Non	Uniforme(a,b)
		Oui : valeur prépondérante m	Triangulaire(a,m,b)
	Oui : $[a, +\infty[$	Oui : moyenne et dispersion	LogNormale
		Oui : 2 valeurs v_1, v_2	LogNormale
	Non	Oui : moyenne et dispersion	Normale
		Oui : 2 valeurs v_1, v_2	Normale
Oui	-	-	Loi historique

Principe du maximum d'entropie

Entropie statistique et Entropie

L'**entropie statistique** est une mesure du désordre du système, c'est à dire de notre ignorance de l'état du système, ou encore de l'information qui nous manque sur son état. Dans les cas où les états possibles du système ne sont pas équiprobables mais chacun associé à une probabilité p_i , l'entropie statistique du système est définie par :

$$\mathcal{S} = -k \sum_{i=1}^{i=n} p_i \log p_i$$

Ainsi, on a associé à une loi de probabilité $(p_i)_i$ un degré de désordre, ou entropie statistique, qui reflète notre niveau de connaissance sur l'état du système. Shannon a étendu la définition de l'entropie statistique du domaine de la physique statistique au cas continu en définissant l'**entropie d'une distribution de densité continue** p :

$$\mathcal{E}(p) = - \int p(x) \log p(x) dx$$

Principe du maximum d'entropie I

Démarche pour le choix de la loi

La démarche de choix est : on prend la loi de probabilité qui **maximise l'information manquante, c'est à dire l'entropie**, compte tenu des renseignements que nous avons sur le système.

Tout autre choix reviendrait à supposer implicitement que nous possédons des informations supplémentaires (ie une entropie plus petite et donc un ordre plus grand).

Quelques exemples

Information disponible	Loi du maximum d'entropie
Support : $[a, b]$	Uniforme(a,b)
Moyenne m et support : $[a, \infty[$	Exponentielle(a, $\lambda = 1/(m - a)$)
Moyenne m , écart-type σ , support : \mathbb{R}	Normale(m, σ)
Moyenne m , écart-type σ , support borné ou semi-borné	$p(x) \propto e^{ax+bx^2}$ sur le support

Validation de la loi issue du dire d'expert

Validation graphique et numérique

Le choix de la loi peut être validé :

- **graphiquement** : tracé de la densité et validation par l'expert,
- **numériquement** : calcul de quelques grandeurs caractéristiques et validation par l'expert :
 - ① médiane : valeur en dessous de laquelle le paramètre reste avec une probabilité de 0.5,
 - ② quantile à 90% : valeur en dessous de laquelle le paramètre reste avec une probabilité de 0.9,
 - ③ écart-type : *dispersion* du paramètre autour de sa moyenne (valable si loi unimodale),
 - ④ moyenne : si la loi est unimodale et symétrique, valeur prise le plus probablement,
 - ⑤ ...