Optimisation

1. Optimisation sans contraintes
2. Optimisation sous contraintes
3. Analyse convexe et dualité

Convexité

L'ensemble X est convexe ssi pour tout x, y dans X, le segment $[x, y] := \{tx + (1-t)y \mid t \in [0, 1]\}$ est inclus dans X

Une fonction $f : X \subset \mathbf{R}^n \to R$ est convexe ssi X est convexe et pour tout x, y dans X et pour tout $t \in [0, 1]$,

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

Si l'inégalité est stricte pour $t \in]0, 1[$, la fonction est dite *strictement* convexe.

Convexité

Reconnaître une fonction convexe

- les fonctions linéaires sont convexes
- les combinaisons linéaires positives de fonctions convexes sont convexes
- le maximum (point par point) d'un ensemble de fonctions convexes est convexe
- …

Si f est de classe C^2, f est convexe si et seulement si $\nabla^2 f$ est semi-définie positive sur l'intérieur de X. Si $\nabla^2 f$ est définie positive sur l'intérieur de X, f est strictement convexe. $(X = \mathrm{dom} f)$

Existence et unicité du minimum

Théorèmes

Si f est convexe et minorée, alors f admet un minimum global.

Si f est une fonction *strictement* convexe définie sur un ensemble convexe $X \subset \mathbf{R}^n$, alors f admet au plus un minimum global.

Si f est une fonction convexe définie sur un ensemble convexe $X \subset \mathbf{R}^n$, alors tout minimum local de f est aussi un minimum global de f. Si X est ouvert, alors $\nabla f(x^*) = 0$ est équivaut à x^* est un minimum global de f

Dualité

Problème considéré

$$\begin{aligned}&\text{minimize} \quad f(x)\\&\text{subject to} \quad x \in X, \qquad g_j(x) \le 0, \quad j = 1, \ldots, r,\end{aligned}$$

$$f : \mathbf{R}^n \to \mathbf{R}$$
$$g_j : \mathbf{R}^n \to \mathbf{R}$$
$$X \subset \mathbf{R}^n$$

$$f^* = \inf_{\substack{x \in X \\ g_j(x) \le 0,\, j=1,\ldots,r}} f(x).$$

$$-\infty < f^* < +\infty$$

Il existe x^*, tel que $f(x^*) = f^*$

Dualité

Lagrangien associé au problème

$$L(x, \mu) = f(x) + \sum_{j=1}^{r} \mu_j g_j(x) = f(x) + \mu' g(x).$$

Dualité

Fonction duale

$$q(\mu) = \inf_{x \in X} L(x, \mu).$$

Problème dual

$$\text{maximize} \quad q(\mu)$$
$$\text{subject to} \quad \mu \geq 0, \ \mu \in D,$$

$$D = \{\mu \mid q(\mu) > -\infty\}.$$

Résultats

Proposition 5.1.2: The domain D of the dual function q is convex and q is concave over D.

Proposition 5.1.3: (Weak Duality Theorem) We have

$$q^* \leq f^*.$$

Dualité

Dualité forte et qualification des contraintes

Example :

Si f est convexe et les contraintes g_j sont linéaires, alors il n'y a pas d'écart dual: $q^* = f^*$

Dualité

If **conditions for strong duality are met** and (x^*, μ^*) are such that $L(x^*, \mu^*) = f^* = q^*$, then (x^*, μ^*) is said to be an optimal solution Lagrange multiplier pair.

Proposition 5.1.5: (Optimality Conditions) (x^*, μ^*) is an optimal solution-Lagrange multiplier pair if and only if

$$x^* \in X, \quad g(x^*) \leq 0, \qquad \text{(Primal Feasibility)}, \tag{5.4}$$

$$\mu^* \geq 0, \qquad \text{(Dual Feasibility)}, \tag{5.5}$$

$$x^* = \arg\min_{x \in X} L(x, \mu^*), \qquad \text{(Lagrangian Optimality)}, \tag{5.6}$$

$$\mu_j^* g_j(x^*) = 0, \quad j = 1, \ldots, r, \qquad \text{(Complementary Slackness)}. \tag{5.7}$$

Un mot sur l'optimisation non lisse

Statistique et probabilités

1. Motivation : estimation problems

2. Computing probabilities

3. Computing and controlling moments and tail probabilities

Problèmes d'estimation

Cadre général

$$\text{Modèle statistique} \quad \mathcal{P}$$
$$\text{Distribution} \quad P \in \mathcal{P}$$
$$\text{Observations} \quad O \sim P$$

- Estimation

$$\text{Fonctionnelle } f : \mathcal{P} \to E$$
$$\text{Estimateur} \quad \hat{f} : \mathcal{O} \to E$$
$$\text{But} \quad \hat{f}(O) \approx f(P)$$

Deux questions fondamentales pour un problème d’estimation

A. Trouver un estimateur computationnellement efficace $\hat{f} : \mathcal{O} \to E$

Dans le cas où on estime une distribution :
Implique des calculs de probabilités (e.g. méthode du maximum de vraisemblance)

B. Evaluer sa qualité statistique $\hat{f}(O) \approx f(P)$

Implique calcul ou contrôle de moments (e.g. biais, variance, risque espéré)

Exemple A

Estimateur du maximum de vraisemblance

Modèle statistique $\mathcal{P}$ $\qquad$ $\mathcal{P} = \{\text{all distributions}\}$

Distribution $P \in \mathcal{P}$

Observations $O \sim P$

Fonctionnelle $f : \mathcal{P} \to E$ $\qquad$ $f(P) = P$

Estimateur $\hat{f} : \mathcal{O} \to E$ $\qquad$ $\hat{f}_{\mathcal{H}} : O \mapsto P_{\arg\max_{w \in W}(p_w(O))}$

$$\mathcal{H} = \{P_w \mid w \in W\} \,:\, \text{espace de recherche/d'hypothèse}$$

$$p_w \,:\, \text{densité de } P_w \text{ (par rapport à une mesure fixe)}$$

Exemple B

Qualité statistique d'un estimateur

- biais $b_P(\hat{f}) = E_{O\sim P}[\hat{f}(O)] - f(P)$ $\qquad \|b_P(\hat{f})\|_2$
- variance $\mathrm{Var}_P(\hat{f}) = \mathrm{Var}_{O\sim P}[\hat{f}(O)]$
- risque $R_P(\hat{f}) = E_{O\sim P}[\ell(f(P), \hat{f}(O))]$

ℓ Fonction de coût

Décomposition biais-variance du risque dans le cas $\ell : x, y \mapsto \|x - y\|_2^2$

$$R_P(\hat{f}) = \mathrm{Var}_P(\hat{f}) + \|b_P(\hat{f})\|_2^2$$

Statistique et probabilités

1. Motivation : estimation problems
2. Computing probabilities
3. Computing and controlling moments and tail probabilities

Univers et mesure de probabilité

Univers Ω

Evénement $A \subset \Omega$

Ensemble des événements (tribu) $\mathcal{F} \subset \mathcal{P}(\Omega)$ Stable par union dénombrable et par passage au complémentaire

Mesure de probabilité $P : \mathcal{F} \to \mathbf{R}$

$$\forall A \in \mathcal{F},\ P(A) \geq 0$$
$$P(\Omega) = 1$$

Pour tout ensemble dénombrable d'événements **disjoints** $A_1, A_2, \dots,$

$$P\left(\bigcup_{i=1}^{+\infty} A_i\right) = \sum_{i=1}^{+\infty} P(A_i)$$

Variable aléatoire

Univers Ω

Ensemble des événements (tribu) $\mathcal{F} \subset \mathcal{P}(\Omega)$ Stable par union dénombrable et par passage au complémentaire

Mesure de probabilité $P : \mathcal{F} \rightarrow \mathbf{R}$

Variable aléatoire $X : \Omega \rightarrow E$

$\mathcal{T} \subset \mathcal{P}(E)$ Tribu sur l'espace d'arrivée

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

Loi de X (mesure de probabilité induite sur E) $\forall T \in \mathcal{T},\ P_X(T) = P(X^{-1}(T))$

Conditionnement, indépendance

Probabilité conditionnelle

$$P(A \mid B) := \frac{P(A \cap B)}{P(B)}$$

Indépendance mutuelle d'événements

$$\forall S \subset \{1, 2, \ldots, n\}, \;\; P\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} P(A_i)$$

Loi de probabilité conditionnelle

$$X : \Omega \to E \quad Y : \Omega \to F \quad \mathcal{T} \subset \mathcal{P}(E) \quad \Sigma \subset P(F)$$

Origine des dépendances : Omega !

Tribus sur les espaces d'arrivée

$$T \in \mathcal{T} \quad S \in \Sigma$$

$$P_{X|Y}(T|S) = P(X^{-1}(T)|Y^{-1}(S)) = \frac{P(X^{-1}(T) \cap Y^{-1}(S))}{P(Y^{-1}(S))}$$

Variables aléatoires indépendantes

$$\forall T \in \mathcal{T}, \; \forall S \in \Sigma, \quad P_{X,Y}(T, S) = P_X(T) P_Y(S)$$

Variables aléatoires réelles

$$X : \Omega \to E \subset \mathbf{R}$$

$$p_X(x) = P_X(\{x\})$$ Fonction de masse

$$F_X(x) = P_X(\{y \mid y \leq x\})$$ Fonction de répartition

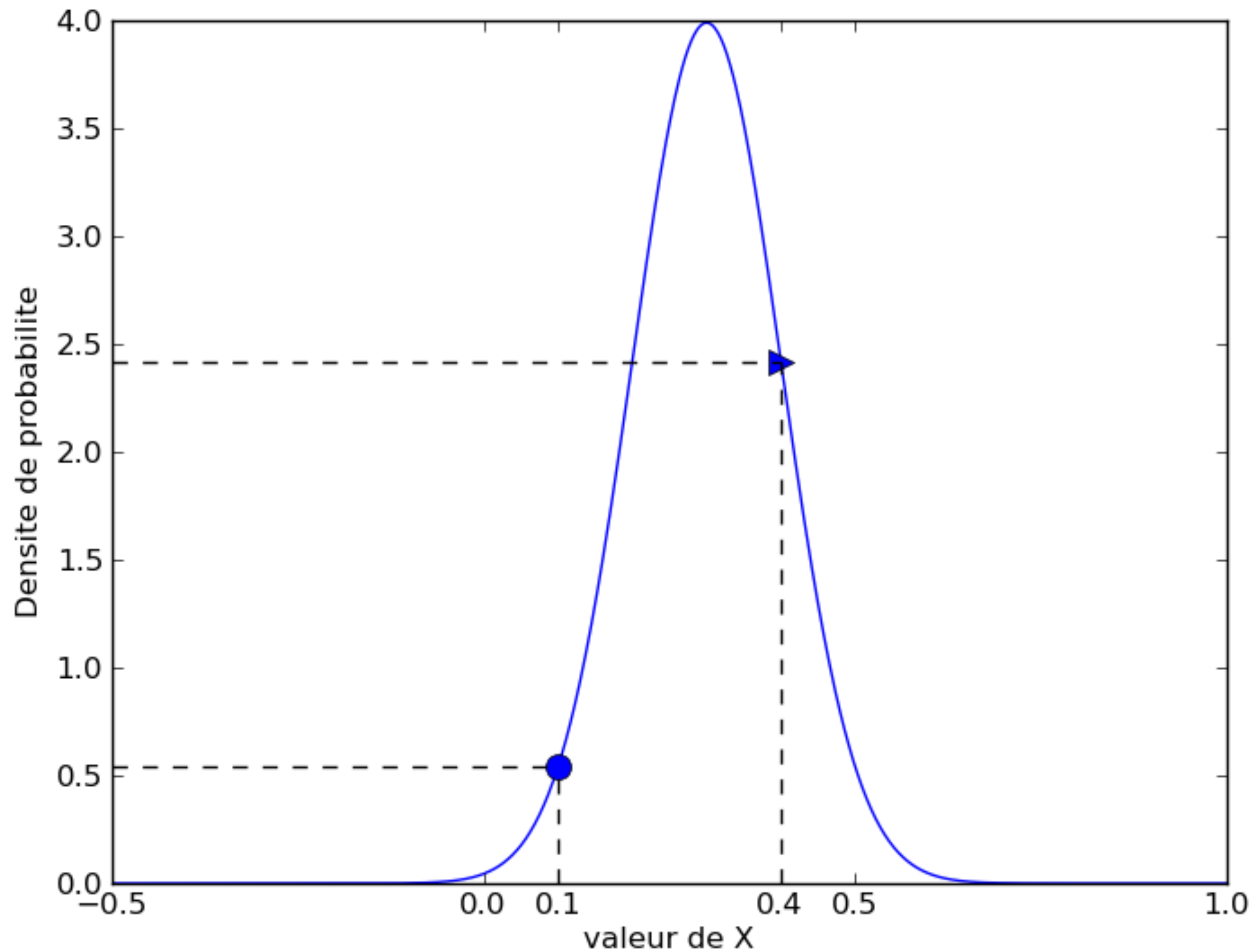
$$p_{dX} = \frac{dF_X}{dx}$$ Densité de probabilité

$$E[X] = \int x P_X(dx)$$ Espérance (intégrale de Lebesgue)

$$E[X] = \int x p_{dX}(x) dx$$ Cas continu

$$E[X] = \sum_x x p_X(x)$$ Cas discret

On a représente sur le graphe ci-dessous la densité de probabilité d'une variable aléatoire réelle X.

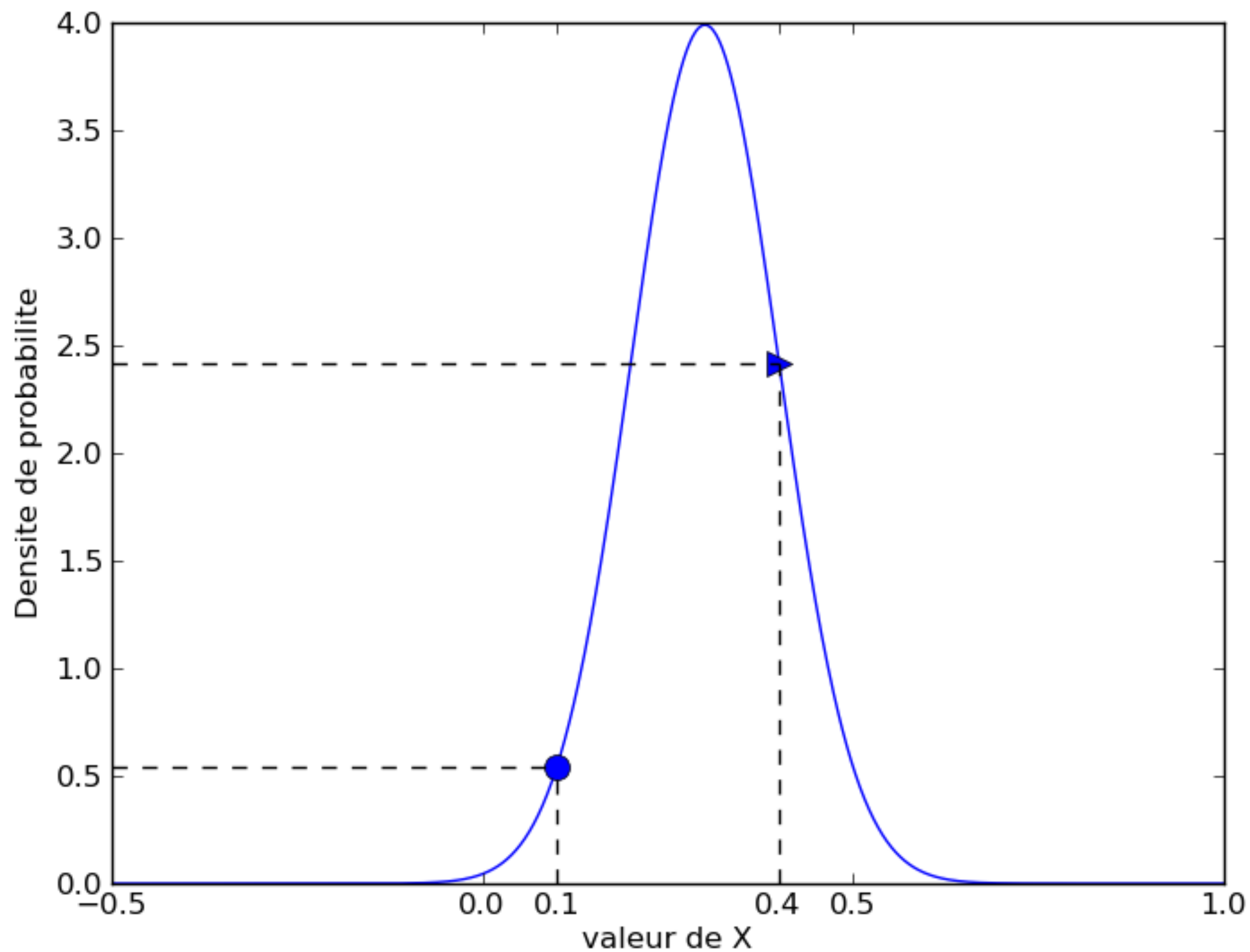


1) La densité de probabilité de X prend une valeur supérieure 1 en $X = 0.4$. Cela vous parait-il normal ? Justifiez votre réponse.

Soit x une réalisation de X.

2) Quelle est la probabilité d'avoir $x = 0.1$? Quelle est la probabilité d'avoir $x = 0.4$? Est il plus probable d'observer $x = 0.4$ ou $x = 0.1$? A quel point (approximativement) ?

On a représente sur le graphe ci-dessous la densité de probabilité d'une variable aléatoire réelle X.



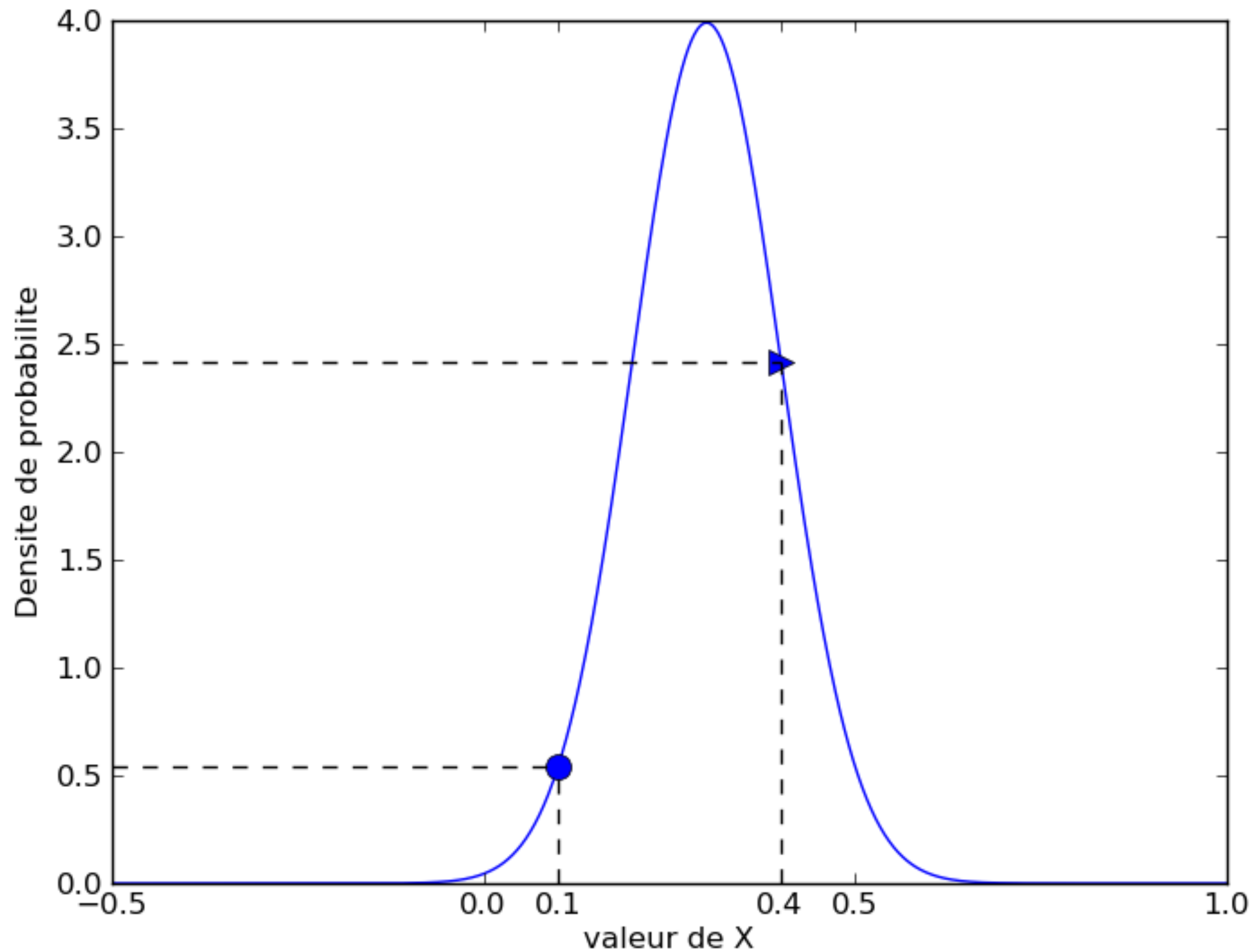
1) La densité de probabilité de X prend une valeur supérieure 1 en $X = 0.4$. Cela vous parait-il normal ? Justifiez votre réponse.

L'aire sous la courbe doit être égale à 1 mais la densité en un point particulier peut être supérieure et même arbitrairement grande.

Soit x une réalisation de X.

2) Quelle est la probabilité d'avoir $x = 0.1$? Quelle est la probabilité d'avoir $x = 0.4$? Est il plus probable d'observer $x = 0.4$ ou $x = 0.1$? A quel point (approximativement) ?

On a représente sur le graphe ci-dessous la densité de probabilité d'une variable aléatoire réelle X.



1) La densité de probabilité de X prend une valeur supérieure 1 en $X = 0.4$. Cela vous parait-il normal ? Justifiez votre réponse.

L'aire sous la courbe doit être égale à 1 mais la densité en un point particulier peut être supérieure et même arbitrairement grande.

Soit x une réalisation de X.

2) Quelle est la probabilité d'avoir $x = 0.1$? Quelle est la probabilité d'avoir $x = 0.4$? Est il plus probable d'observer $x = 0.4$ ou $x = 0.1$? A quel point (approximativement) ?

La probabilité que x soit 0.1 ou 0.4 est 0. Par contre, il est environ 5 fois plus probable d'observer $X = 0.4$ que $x = 0.1$.

Plusieurs variables aléatoires réelles

$$\mathbf{X} = (X_1, X_2, \ldots, X_n) : \Omega \to \mathbf{R}^n$$ Origine des dépendances : Omega !

$$p_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = P_{\mathbf{X}}(\{x_1, x_2, \ldots, x_n\})$$ Fonction de masse (jointe)

$$F_{\mathbf{X}}(x_1, x_2, \ldots, x_n) = P_{\mathbf{X}}(\{y_1, y_2, \ldots, y_n \mid y_1 \leq x_1, y_2 \leq x_2, \ldots, y_n \leq x_n\})$$

Fonction de répartition (jointe)

$$p_{d\mathbf{X}} : x_1, x_2, \ldots, x_n \mapsto \frac{\partial F_{\mathbf{X}}}{\partial x_1 dx_2 \ldots dx_n}(x_1, x_2, \ldots, x_n)$$

Densité de probabilité (jointe)

Espérance $f : \mathbf{R}^n \to \mathbf{R}$

Cas continu $E[f(\mathbf{X})] = \int_{x_1} \int_{x_2} \cdots \int_{x_n} f(x_1, x_2, \ldots, x_n) p_{d\mathbf{X}}(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n$

Cas discret $E[f(\mathbf{X})] = \sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} f(x_1, x_2, \ldots, x_n) p_{\mathbf{X}}(x_1, x_2, \ldots, x_n)$

Modèles graphiques

$$p(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | X_{\pi_i})$$

Statistique et probabilités

1. Motivation : estimation problems

2. Computing probabilities

3. Computing and controlling moments and tail probabilities

Moments

Variance $\mathrm{Var}[X] = E[(X - E[X])^2]$

Moment $E[X^k]$

Moment centré $E[(X - E[X])^k]$

Covariance $\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$

Linéarité de l'espérance $$E\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i E(X_i)$$

Bilinéarité de la covariance

$$\mathrm{Cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \mathrm{Cov}(X_i, Y_j)$$

Moments conditionnels

Espérance conditionnelle

$$f(y) = E[X|Y=y] = \int x P_{X|Y=y}(dx)$$

Variance conditionnelle

$$g(y) = \mathrm{Var}[X|Y=y] = E[(X - E[X])^2|Y=y]$$

Loi de l'espérance totale

$$E[X] = E[E[X|Y]]$$

Loi de la covariance totale

$$\mathrm{Cov}[X,Y] = \mathrm{Cov}[E[X|Z], E[Y|Z]] + E[\mathrm{Cov}[X,Y|Z]]$$

Statistique et probabilités

1. Motivation : estimation problems

2. Computing probabilities

3. Computing and controlling moments and tail probabilities